

# **Tool development to study ubiquitination machinery**

**A thesis submitted to Imperial College London in candidature  
for the degree of Doctor of Philosophy of Imperial College**

**By**

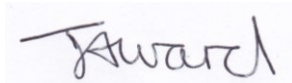
**Jennifer Ward**

**Institute of chemical Biology  
Department of Chemistry  
Imperial College London  
Exhibition Road  
London  
SW7 2AZ**

**May 2016**

## **Declaration of originality**

This Thesis is my own work and reports the results of my original research. Where information derives from the work of others or in collaboration with others, this is acknowledged in the text and references.

A handwritten signature in black ink on a light blue background. The signature reads "J Ward" in a cursive, slightly slanted script.

Jennifer Ward, May 2016.

## **Copyright declaration**

The copyright of this thesis rests with the author and is made fully available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution researchers must make clear to others the licence terms of this work.

# Abstract

---

Ubiquitination is a diverse post-translational modification, involved in a plethora of eukaryotic processes. At least three different enzymes are required for ubiquitination to occur: an ubiquitin-activating enzyme (E1), an ubiquitin-conjugating enzyme (E2) and an ubiquitin ligase (E3). Conversely, deubiquitinating enzymes (DUBs) regulate the removal of ubiquitin modifications. Together, this enzymatic machinery facilitates a wide array of ubiquitin modifications and further ubiquitin-like modifications. Such modifications play a significant role in the regulation of vital biological processes including proteasomal degradation, DNA damage response (DDR) and NF- $\kappa$ B signalling.

Ubiquitination machinery can be studied using a chemical proteomics approach. Activity-based probes (ABPs) which covalently trap enzymes can be employed. Such probes possess bioorthogonal handles, which on ligation to complementary reporter groups enable enrichment of probe-enzyme complexes prior to LC-MS/MS, western blot, or fluorescent analysis. The design, synthesis, and development of six inhibitor-inspired ABPs for ubiquitin machinery is described. Both small molecule and peptide probes were explored, and assessed for their ability to act as ubiquitin machinery probes both in intact cells and cell lysate. Spike-in SILAC methodology was employed to quantify probe targets under competitive conditions against parent inhibitors, and in a DNA damage response model.

An alternative to using literature inhibitors is to derive novel starting points for tool design through a fragment-based drug discovery approach. A high throughput screen against the minimal catalytic core of HOIP, an E3 ligase, is described together with subsequent validation and characterisation of hit fragments by waterLOGSY NMR and Micro-scale thermophoresis, and crystallography attempts. Finally, a model system for the structural analysis of transient enzyme complexes is explored, to further our understanding of these interactions and ultimately to assess the applicability of these complexes for future drug discovery. The design and synthesis of a maleimide trap is described, along with its application to covalently trap an Ub-E2-E3 complex.

# Acknowledgments

---

First I'd like to thank my supervisors Ed Tate and Eric Lam for offering me the chance to work on this interesting and varied project. I've learnt a lot over the last few years! Ed, there's certainly been highs and lows at the bench, but your support, encouragement, and guidance has been a constant for which I would like again to thank you.

Thanks to Katrin Rittinger for all your support, both on the phone and at Mill Hill. I've really enjoyed working on our collaboration, and it's enabled me to explore further techniques during my PhD. Additional thanks to Nick Brown and Steve Martin for their help with this collaboration.

I would like to thank the Chemistry department staff, particular thanks to Lisa Haigh for maintaining the Q Exactive and the Mass spectrometry service, and to Pete Haycock and Yingqi Xu for maintaining the NMR facilities.

I'd like to thank all the members of the Tate group, past and present, for putting up with me every day! Thanks to Jennie for her synthetic chemistry knowledge, Neki for all things peptide and LC-MS, Manue for teaching me cell culture, Julia for being the other part of Team LC-MS, Chris for his DSF expertise, Mostafa for his waterLOGSY knowledge, Goska and James for their guidance in proteomic analysis, and all the Tate group members who have been in charge of the Q Exactive or peptide synthesiser queue over the years.

Thanks to the all the fifth office members who have been there for me. Thanks to Julia and Lisa for keeping me sane. Monica I will miss our morning coffee! Thanks to you all! Special thanks to my proof readers: Anna, Charlotte, Julia, Kate, Kwang, Markus, Monica, and Scott.

I am grateful to AstraZeneca and the ICB for providing the funding for my PhD. In particular the ICB has provided me with a lasting support network (and friends to boot!) Thanks Ernie and Ben for our ICB groupie lunches, I'm sorry my cooking was sometimes questionable!

To Mum and Dad - thanks for your continued support over the years: on the phone, down at Ciao Bella's, over restorative Sunday roast lamb! Thanks for helping me through the hard times. Siân, thank you for dragging me down the pub! And last but by no means least, Simon, thank you for all your support, for always being there for me - for keeping me as sane as possible! You're the best. 😊

# Abbreviations

---

°C	degrees Celsius
53BP1	tumor suppressor p53-binding protein 1
ABP	activity-based probe
ABPP	activity-based protein profiling
Ac	acyl
Acetone-d <sup>6</sup>	deuterated acetone
ADRM1	proteasomal ubiquitin receptor ADRM1
AfBP	affinity-based probe
Ahx	amino hexanoic acid
AMBIC	ammonium bicarbonate
ANP32B	acidic leucine-rich nuclear phosphoprotein 32 family member B
APC/C	anaphase promoting complex/cyclosome
AQUA	absolute quantification of proteins
Arg	arginine
ARIH1	E3 ubiquitin-protein ligase ARIH1
ATG12	ubiquitin-like protein ATG12
ATG3	ubiquitin-like-conjugating enzyme ATG3
ATG8	autophagy-related protein 8
ATP	adenosine triphosphate
AzTB	azido-TAMRA-biotin capture reagent
BARD1	BRCA1-associated RING domain protein 1
BirA	bifunctional ligase/repressor BirA
Boc	tert-butyloxycarbonyl
BRCA1	breast cancer type 1 susceptibility protein
BRCA2	breast cancer type 2 susceptibility protein
BRD4	bromodomain-containing protein 4
BSA	bovine serum albumin
CAND1	cullin-associated NEDD8-dissociated protein 1
Cas-9	caspase-9
CETSA	cellular thermal shift assay
CF	cystine-free
CHIP	E3 ubiquitin-protein ligase CHIP
CID	collision induced dissociation
MeCN	acetonitrile
CPMG	Car-Purcell-Meiboom-Gill
CRBN	protein cereblon
CRISPR	clustered regularly-interspaced short palindromic repeats
CRKL	Crk-like protein
CuAAC	copper catalysed azide-alkyne cycloaddition
CUL4A	cullin-4A
CUL4B	cullin-4B
Cy3	cyanine 3 dye
CYLD	ubiquitin carboxyl-terminal hydrolase CYLD
Cys	cysteine
Da	dalton
DAPI	4',6-diamidino-2-phenylindole

DARTS	drug affinity responsive target stability
DCM	dichloromethane
DDR	DNA damage response
DDX39A	ATP-dependent RNA helicase DDX39A
DIPEA	<i>N,N</i> -diisopropylethylamine
DMAP	4-dimethylaminopyridine
DMEM	Dulbecco's modified eagle's medium
DMF	dimethylformamide
DMSO	dimethyl sulfoxide
DMSO-d <sup>6</sup>	deuterated dimethyl sulfoxide
DNA	deoxyribonucleic acid
DSB	double strand break
DSF	differential scanning fluorimetry
DTT	dithiothreitol
DUB	deubiquitinating enzyme
E1	ubiquitin-activating enzyme
E2	ubiquitin-conjugating enzyme
E3	ubiquitin ligase
EDCI	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide
EDTA	ethylenediaminetetraacetic acid
ESI	electrospray ionisation
Et	ethyl
EtOAc	ethyl acetate
FANCD2	fanconi anemia group D2 protein
FANCI	fanconi anemia group I protein
FANCL	E3 ubiquitin-protein ligase FANCL
FAT10	HLA-F adjacent transcript 10
FBDD	fragment-based drug discovery
FBS	foetal bovine serum
FDA	food and drugs administration
FDR	false discovery rate
Fmoc	9-fluorenylmethoxycarbonyl
FOXP3	forkhead box protein P3
GFP	green fluorescent protein
GGR	global genome repair
Gly	glycine
GO	gene ontology
H/L	'heavy'/'light' SILAC ratio
HA	human influenza hemagglutinin
HATU	<i>N,N,N',N'</i> -Tetramethyl-O-(7-azabenzotriazol-1-yl)uraniu <sup>m</sup> hexafluorophosphate
HBTU	<i>N,N,N',N'</i> -Tetramethyl-O-(1H-benzotriazol-1-yl)uraniu <sup>m</sup> hexafluorophosphate
HCD	higher-energy collisional dissociation
HECT	homologous to E6AP C-terminus
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HERC2	E3 ubiquitin-protein ligase HERC2
HFIP	1,1,1,3,3,3-Hexafluoro-2-propanol
His	poly-histidine
HMDS	bis(trimethylsilyl)amide
HMOX2	heme oxygenase 2
HOBT	hydroxybenzotriazole

hr	hour
HR	homologous recombination
HRMS	high resolution mass spectrometry
HRP	horseradish peroxidase
HTS	high throughput screen
HUB1	ubiquitin-like modifier hub1
HUWE1	E3 ubiquitin-protein ligase HUWE1
Hz	hertz
IA	iodoacetamide
IAP	inhibitor of apoptosis
IBR	in-between-RING
IC <sub>50</sub>	half maximal inhibitory concentration
ICL	interstrand cross link
ICPL	isotope-coded protein label
ID	identification
IFN	interferons
Ile	isoleucine
IP	isoelectric point
IPO4	importin-4
IRAK1	interleukin-1 receptor-associated kinase 1
ISG15	Interferon-simulated gene 5
ITC	isothermal titration calorimetry
ITPK1	inositol-tetrakisphosphate 1-kinase
iTRAQ	isobaric tags for relative and absolute quantitation
IκBs	inhibitors of NF-κB
<i>J</i>	coupling constant
Jak2	Tyrosine-protein kinase JAK2
JUP	junction plakoglobin
K8	<sup>13</sup> C <sup>15</sup> N labelled lysine
K <sub>d</sub>	dissociation constant
KEAP1	kelch-like ECH-associated protein 1
L/H	'light'/'heavy' SILAC ratio
LC	liquid chromatography
LC-MS	liquid chromatography coupled with mass spectrometry
LC-MS/MS	liquid chromatography coupled with tandem mass spectrometry
LFQ	label-free quantification
LUBAC	linear ubiquitin chain assembly complex
Lys	lysine
<i>M</i>	molar
[M+...]	molecular ion plus...
MALDI	matrix-assisted laser desorption/ionization
MBS	MES buffered saline
MES	2-( <i>N</i> -morpholino)ethanesulfonic acid
MDM2	E3 ubiquitin-protein ligase Mdm2
Me	methyl
MeOH	methanol
Met	methionine
min	minute
MJD	Machado-Josephin domain

MM	multiple myeloma
MNSFβ	monoclonal nonspecific suppressor factor β
mol	moles
MPN+/JAMM	JAB1/MPN/MOV34 metallo-enzymes
MS	mass spectrometry
MS/MS	tandem mass spectrometry
MS1	first mass detection in tandem mass spectrometry
MS2	second mass detection in tandem mass spectrometry
MST	microscale thermophoresis
MW	molecular weight
MWCO	molecular weight cut-off
<i>m/z</i>	mass to charge ratio
n.d	not determined
Necode	neutron encoding
NEDD4	E3 ubiquitin-protein ligase NEDD4
NEM	<i>N</i> -ethyl maleimide
NEMO	NF-κB essential modulator
NER	nucleotide excision repair
NF-κBs	NF-κB proteins
NF-κB	nuclear factor kappa-light-chain-enhancer of activated B cells
NHEJ	non-homologous end joining
NHS	<i>N</i> -Hydroxysuccinimide
NLS	nuclear localisation signal
NMM	<i>N</i> -Methylmorpholine
NMP	<i>N</i> -Methyl-2-pyrrolidone
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
NPM1	nucleoplasmin-1
NPM3	nucleoplasmin-3
NRF2	nuclear factor (erythroid-derived 2)-like 2
OTU	ovarian tumour proteases
OTUB1	ubiquitin thioesterase OTUB1
OTUB7B	OTU domain-containing protein 7B (A20)
OTUD7A	OTU domain-containing protein 7A (Cezanne)
OTULIN	ubiquitin thioesterase otulin
P53	cellular tumour antigen p53
PARK7	protein deglycase DJ-1
PARP	poly (ADP-ribose) polymerase
PBS	phosphate buffer saline
PCNA	proliferating cell nuclear antigen
PCR	polymerase chain reaction
Pg	propargyl glycine
PGP	phosphoglycolate phosphatase
polyUb	polyubiquitin
PPI	protein-protein interaction
Ppm	parts per million
Pro	proline
PROTAC	proteolysis-targeting chimera
PTM	post translational modification
PVDF	polyvinylidene difluoride



PyBop	benzotriazol-1-yl-oxytrypyrrolidinophosphonium hexafluorophosphate
Q-TOF	quadrupole time of flight
R10	<sup>13</sup> C <sup>15</sup> N labelled arginine
RAD18	E3 ubiquitin-protein ligase RAD18
RAD23B	UV excision repair protein RAD23 homolog B
RAD6B	ubiquitin-conjugating enzyme E2 B
RanGAP1	ran GTPase-activating protein 1
RBR	RING between RING
R <sub>f</sub>	retention factor
RING	really interesting new gene
RING1	E3 ubiquitin-protein ligase RING1
RNA	ribonucleic acid
RNAi	RNA interference
RNF14	E3 ubiquitin-protein ligase RNF14
RNF168	E3 ubiquitin-protein ligase RNF168
RNF169	E3 ubiquitin-protein ligase RNF169
RNF8	E3 ubiquitin-protein ligase RNF
Rpb1	DNA-directed RNA polymerase II subunit RPB1
RPL17	60S ribosomal protein L17
Rsp5	E3 ubiquitin-protein ligase RSP5
RT	room temperature
RUNX1	Runt-related transcription factor 1
s.d	standard deviation
SAR	structure activity relationship
SCF	skp, Cullin, F-box containing complex
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SENP	sentrin specific proteases
Ser	serine
SILAC	stable isotope labelling by amino acids in cell culture
SMAD2	mothers against decapentaplegic homolog 2
SMAD7	mothers against decapentaplegic homolog 7
SMURF1	E3 ubiquitin-protein ligase SMURF1
SMURF2	E3 ubiquitin-protein ligase SMURF2
SPPS	solid phase peptide synthesis
SPR	surface plasmon resonance
SRM	selected reaction monitoring
SSB	single strand break
SSRP1	FACT complex subunit SSRP1
STD	saturation Transfer Difference
STRING	search tool for the retrieval of interacting genes/proteins
SUMO	small ubiquitin like modifier
SUPT16H	FACT complex subunit SPT16
T2	spin-spin relaxation time (NMR)
TAMRA	5,6-Carboxytetramethylrhodamine
tBME	tert-butyl methyl ether
TBS	tris buffered saline
TBS-T	TBS-Tween buffer
TBTA	tris[(1-benzyl-1H-1,2,3- triazol-4-yl)methyl]amine
tBu	tert-butyl

TCEP	tris(2- carboxyethyl)phosphine
TCR	transcription coupled repair
TFA	trifluoroacetic acid
THF	tetrahydrofuran
Thr	threonine
TIS	triisopropylsilane
TLC	thin-layer chromatography
T <sub>m</sub>	melting temperature
ΔT <sub>m</sub>	change in melting temperature
TMS	trimethylsilyl
TMT	tandem mass tags
TRIM28	transcription intermediary factor 1-beta
TRIM33	E3 ubiquitin-protein ligase TRIM33
Tris	tris(hydroxymethyl)aminomethane
tRNA	transfer RNA
Ts	toluenesulfonyl
TUBEs	tandem repeat ubiquitin binding entities
Ub	ubiquitin
UB2R1	ubiquitin-conjugating enzyme E2 R1
UBA1	ubiquitin-like modifier-activating enzyme 1
UBA3	Nedd8-activating enzyme E1 catalytic subunit
UBA6	ubiquitin-like modifier-activating enzyme 6
UBC5	ubiquitin-conjugating enzyme E2-16 kDa
Ubch7	ubiquitin-conjugating enzyme E2 L3
UBD	ubiquitin binding domain
UBE2D	ubiquitin-conjugating enzyme E2 D1
UBE2N	ubiquitin-conjugating enzyme E2 N
UBE2T	ubiquitin-conjugating enzyme E2 T
UBE2V1	ubiquitin-conjugating enzyme E2 variant 1
UBE4B	ubiquitin conjugation factor E4 B
Ubl	ubiquitin-like modification
UCH	ubiquitin carboxyl-terminal hydrolases
UCHL5	ubiquitin carboxyl-terminal hydrolase isozyme L5
UFM1	ubiquitin-fold modifier 1
UHRF1	E3 ubiquitin-protein ligase UHRF1
UPS	ubiquitin proteasome system
URM1	ubiquitin-related modifier 1
USP	ubiquitin-specific proteases
UV	ultraviolet
VME	vinyl methyl ester
VS	vinyl sulphone
ZNF622	Zinc finger protein 622

# Presentations arising from the thesis

---

## Oral presentations

J. Ward *et al.* "Development of activity-based probes to target ubiquitination during DNA damage response", Imperial College London Postgraduate Symposium, London, June 2015\*

J. Ward *et al.* "Cell permeable activity-based probes for targeting ubiquitination in DNA damage response and cancer drug efficacy", Chemical biology and Molecular Medicine Symposium, Cambridge, October 2014<sup>†</sup>

J. Ward *et al.* "Cell permeable activity-based probes for targeting ubiquitination in DNA damage response and cancer drug efficacy", 2014 Joint Centre for Doctoral Training Conference, Warwick, May 2014.

## Poster presentations

J. Ward *et al.* "A chemical proteomic approach to target ubiquitination during DNA damage response", The Human Proteome Keystone Symposium, Stockholm, Sweden, April 2015

J. Ward *et al.* "Towards novel probes for ubiquitin-processing enzymes", COST Chemical Proteomics Meeting, Oxford, April 2015

J. Ward *et al.* "Cell permeable activity-based probes for targeting ubiquitination in DNA damage response and cancer drug efficacy", Chemical Biology and Molecular Medicine Symposium, Cambridge, October 2014

J. Ward *et al.* "Cell permeable activity-based probes for targeting ubiquitination in DNA damage response and cancer drug efficacy", Copenhagen Bioscience Conference, Copenhagen, Denmark, September 2014

J. Ward *et al.* "Cell permeable activity-based probes for targeting ubiquitination in DNA damage response and cancer drug efficacy", EMBO Chemical Biology 2014, Heidelberg, Germany, August 2014

J. Ward *et al.* "Activity-based probes for targeting ubiquitination in DNA damage response and cancer drug efficacy", Imperial College London Postgraduate Symposium, London, July 2014

J. Ward *et al.* "Cell permeable activity-based probes for targeting ubiquitination in DNA damage response and cancer drug efficacy", RSC Chemical Biology and Bio-Organic Postgraduate Symposium, Warwick, April 2014

J. Ward *et al.* "Novel activity-based probes for targeting ubiquitination in DNA damage response and cancer drug efficacy", The Ubiquitin System: From Basic Science to Drug Discovery Keystone Symposium, Montana, USA, January 2014

J. Ward *et al.* "activity-based probes for targeting ubiquitination", 2013 Joint Centre for Doctoral Training Conference, London, June 2013.

\* Awarded prize for best talk in Chemical Biology section.

<sup>†</sup> Flash presentation

# Contents

---

Abstract.....	1
Acknowledgments.....	2
Abbreviations.....	3
Presentations arising from the thesis .....	9
Contents.....	10
Chapter 1 Introduction .....	14
1.1 Ubiquitination as a post translational modification (PTM).....	14
1.1.1 Discovery.....	14
1.1.2 Types of modification.....	15
1.1.3 Mechanisms of ubiquitination .....	16
1.1.4 Recognition of Ub PTMs: UBDs.....	20
1.1.5 Ubiquitin like modifications (Ubls).....	21
1.1.6 Cellular roles and chemotherapeutic potential .....	24
1.2 Studying ubiquitination: a proteomics approach .....	30
1.2.1 Proteomics: the challenges of studying PTMs .....	30
1.2.2 Mass spectrometry (MS) approaches .....	30
1.2.3 Quantitative proteomic techniques.....	34
1.2.4 Profiling Ub .....	36
1.2.5 Profiling Ub machinery.....	39
1.2.6 Inhibitors of Ub machinery .....	43
1.3 Project objectives.....	47
Chapter 2 Peptide activity-based probes for studying ubiquitination machinery .....	48
2.1 Introduction .....	49
2.2 Design and synthesis.....	49
2.3 In lysate applications and initial proteomic evaluation.....	51
2.3.1 Maximising coverage: cellular fractionation.....	55
2.3.2 Evaluation of literature Ub- VME probe .....	58
2.4 Quantitative proteomics.....	60
2.4.1 SILAC analysis of competition .....	60
2.4.2 Studies of a basic DDR model .....	65
2.5 Preliminary in cell application.....	68
2.5.1 Determining peptide cellular localisation .....	69
2.5.2 Peptide development for in cell applications. ....	71
2.6 Conclusions .....	72
Chapter 3 Small molecule activity-based probes for studying ubiquitination enzymes .....	73
3.1 Introduction .....	74
3.1.1 Small molecule inhibitor leads for probe development .....	74
3.2 PYR-41 probe .....	76
3.2.1 Design and synthesis.....	76
3.2.2 In lysate application and initial proteomic evaluation.....	78

3.2.3	Optimisation of in-cell treatment .....	80
3.2.4	Summary .....	88
3.3	Tz9 probe series (T probe series) .....	89
3.3.1	Design and synthesis .....	89
3.3.2	In cell application of T- probe series .....	93
3.3.3	Cross-probe comparison .....	99
3.4	Conclusions .....	101
Chapter 4	Identifying potential scaffolds for E3 inhibitors: HOIP as a case study.....	102
4.1	Introduction .....	103
4.1.1	Fragment-based drug discovery (FBDD) .....	103
4.1.2	Biophysical and structural techniques .....	105
4.1.3	Screening E3 ligases .....	107
4.1.4	HOIP .....	108
4.2	DSF screening of HOIP catalytic core .....	108
4.2.1	Optimisation of DSF assay for P37 .....	108
4.2.2	Initial screen of P37.....	110
4.3	Concentration dependence studies by DSF .....	115
4.4	WaterLOGSY validation.....	117
4.5	Fragment characterisation by MST.....	120
4.6	X-ray crystallography .....	125
4.7	Conclusions .....	125
Chapter 5	Crystallographic models to facilitate the structural study of E3 ligases .....	127
5.1	Introduction .....	128
5.1.1	Structural study of transient complexes.....	128
5.1.2	Studying RBR ligase complexes.....	128
5.2	Trap development.....	130
5.2.1	Trap design.....	130
5.2.2	Trap synthesis .....	131
5.3	Towards a trapped Ub-Ubch7-HHARI complex.....	132
5.3.1	Mono-Ub trap formation .....	132
5.3.2	Trapping a minimal E3 domain .....	133
5.3.3	Trapping a larger HHARI fragment.....	134
5.4	Conclusions / future work.....	136
Chapter 6	Conclusions and future work .....	138
6.1	ABPs to study ubiquitination machinery.....	138
6.1.1	Conclusions .....	138
6.1.2	Future work.....	140
6.2	FBDD and structural models for RBR E3 ligases.....	142
6.2.1	Conclusions .....	142
6.2.2	Future work.....	143

Chapter 7	Materials and Methods.....	144
7.1	Chemical synthesis.....	144
7.1.1	General methods .....	144
7.1.2	<i>Tert</i> -butyl[(2 <i>E</i> )-3-(methylsulfonyl)prop-2-en-1-yl]carbamate (2) <sup>176</sup> .....	145
7.1.3	(2 <i>E</i> )-3-(methylsulfonyl)prop-2-en-1-amine trifluoroacetate (VS) <sup>176</sup> .....	145
7.1.4	<i>tert</i> -butoxy[(2 <i>E</i> )-4-methoxy-4-oxobut-2-en-1-yl]oxoammonium (3) <sup>177</sup> .....	146
7.1.5	(2 <i>E</i> )-4-methoxy-4-oxobut-2-en-1-aminium tosylate (VME) <sup>177</sup> .....	146
7.1.6	Ethyl 4-hydrazinylbenzoate hydrochloride (7) <sup>267</sup> .....	147
7.1.7	Ethyl 4-(3,5-dioxypyrazolidin-1-yl)benzoate (5) <sup>268</sup> .....	147
7.1.8	4-[(4 <i>E</i> )-4-[(5-nitrofuran-2-yl)methylidene]Ethyl4-(3,5-dioxypyrazolidin-1-yl)benzoate (PYR-41) <sup>266</sup> .....	148
7.1.9	4-(3,5-dioxypyrazolidin-1-yl)benzoic acid (9).....	148
7.1.10	4-(3,5-dioxypyrazolidin-1-yl)- <i>N</i> -(prop-2-yn-1-yl)benzamide (4).....	149
7.1.11	4-[(4 <i>E</i> )-4-[(5-nitrofuran-2-yl)methylidene]-3,5-dioxypyrazolidin-1-yl]- <i>N</i> -(prop-2-yn-1-yl)benzamide (PYR-41-P).....	149
7.1.12	<i>N</i> -(4-methylphenyl)imidodicarbonimidic diamide (11) <sup>210</sup> .....	150
7.1.13	{4-amino-6-[(4-methylphenyl)amino]-1,3,5-triazin-2-yl}methanol (10) <sup>210</sup> .....	150
7.1.14	{4-amino-6-[(4-methylphenyl)amino]-1,3,5-triazin-2-yl}methyl 4-nitrobenzoate (Tz9) <sup>210</sup> .....	151
7.1.15	<i>N</i> -(4-iodophenyl)imidodicarbonimidic diamide (20).....	152
7.1.16	{4-amino-6-[(4-iodophenyl)amino]-1,3,5-triazin-2-yl}methanol (21).....	152
7.1.17	{4-amino-6-[(4-ethynylphenyl)amino]-1,3,5-triazin-2-yl}methanol (15).....	153
7.1.18	{4-amino-6-[(4-ethynylphenyl)amino]-1,3,5-triazin-2-yl}methyl 4-nitrobenzoate (T1).....	153
7.1.19	4-nitro-3-(prop-2-yn-1-yloxy)benzoic acid (12).....	154
7.1.20	{4-amino-6-[(4-methylphenyl)amino]-1,3,5-triazin-2-yl}methyl 4-nitro-3-(prop-2-yn-1-yloxy)benzoate (T2).....	155
7.1.21	4-nitro-3-(prop-2-yn-1-ylamino)benzoic acid (13).....	156
7.1.22	{4-amino-6-[(4-methylphenyl)amino]-1,3,5-triazin-2-yl}methyl 4-nitro-3-(prop-2-yn-1-ylamino)benzoate (T3).....	156
7.1.23	4-nitro-3-(propan-1-yloxy)benzoic acid (23).....	157
7.1.24	{4-amino-6-[(4-methylphenyl)amino]-1,3,5-triazin-2-yl}methyl 4-nitro-3-(propan-1-yloxy)benzoate (T2c).....	158
7.1.25	Ethyl 2,5-dioxo-2,5-dihydro-1 <i>H</i> -pyrrole-1-carboxylate (29) <sup>309</sup> .....	159
7.1.26	Tris(3-maleimidopropyl)amine (Trap C).....	159
7.2	Peptide synthesis.....	161
7.2.1	General SPPS procedures.....	161
7.2.2	Fmoc Deprotection.....	161
7.2.3	Peptide synthesis.....	161
7.2.4	<i>N</i> -terminal manipulations.....	162
7.2.5	Resin displacement.....	162
7.2.6	In solution C-terminal electrophilic trap incorporation.....	163
7.2.7	Peptide Characterisation.....	164
7.3	Biochemical methods.....	165
7.3.1	General methods .....	165
7.3.2	Cell culture .....	165
7.3.3	Cell lysis.....	165
7.3.4	In-cell labelling assays.....	166
7.3.5	In-lysate labelling assays.....	167

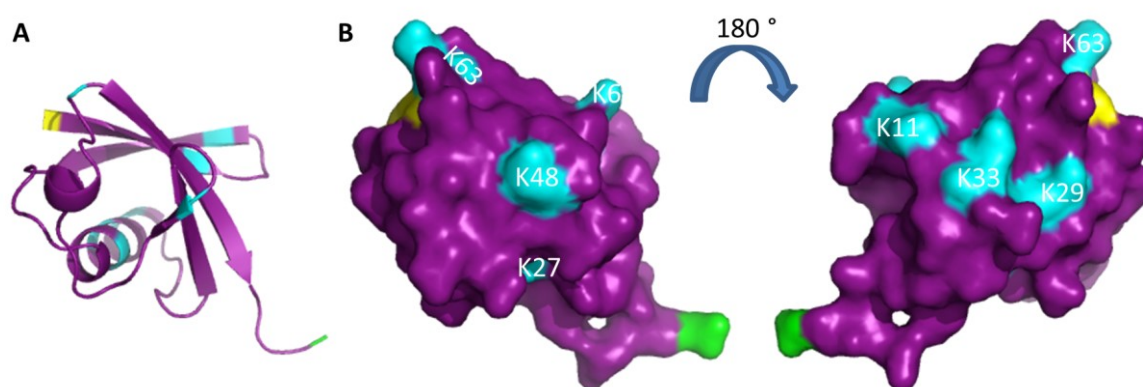
7.3.6	CuAAC and in-gel fluorescence .....	167
7.3.7	Western blot analysis.....	168
7.3.8	Live cell imaging .....	169
7.3.9	E2-E3 trap assembly.....	169
7.4	Chemical proteomics .....	170
7.4.1	General methods .....	170
7.4.2	Lysate preparation .....	170
7.4.3	Capture and enrichment of MS-based proteomics .....	171
7.4.4	On bead reduction, alkylation and digest.....	171
7.4.5	LC/MS/MS .....	172
7.5	Fragment-Based Drug Discovery (FBDD) workflow .....	174
7.5.1	Differential Scanning Fluorimetry (DSF).....	174
7.5.2	WaterLOGSY experiment .....	175
7.5.3	Micro-scale thermophoresis .....	175
	Bibliography .....	177
	Appendices and spectra.....	194

# Chapter 1 Introduction

## 1.1 Ubiquitination as a post translational modification (PTM)

### 1.1.1 Discovery

Post translational modification (PTM) is the covalent addition or removal of chemical entities on a translated protein, in order to regulate its biological properties.<sup>1, 2</sup> Several PTMs, including acetylation, methylation, phosphorylation, and ubiquitination are known to govern cellular processes. Ubiquitination, the reversible attachment of the 76 amino acid protein ubiquitin (Ub, Figure 1A) to substrate proteins, is a diverse PTM which plays a vital role in several eukaryotic processes.<sup>3</sup> Its roles include controlling cellular localisation and altering protein-protein interactions (PPIs), which in turn determine enzymatic activity and protein turnover.<sup>4</sup>



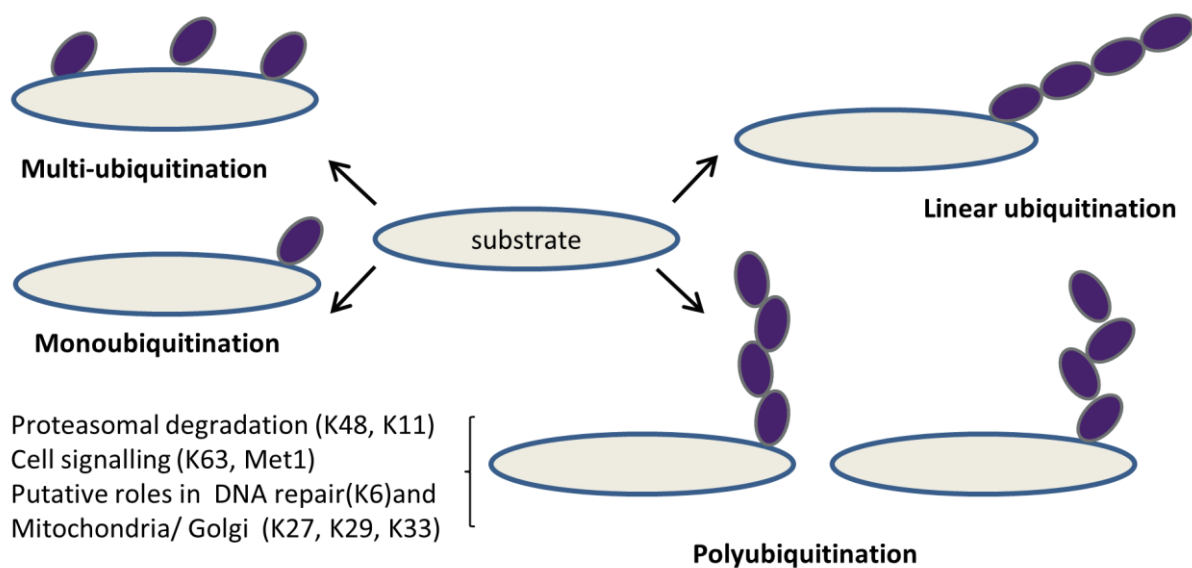
**Figure 1: (A) Cartoon and (B) space filling PyMOL representations of Ubiquitin (Ub), with the C-terminal Glycine (G76, Green), N-terminal methionine (M1, Yellow), and all Lysine residues (K, Cyan) labelled.**

Ub was first discovered in 1975 by Gideon *et al.* as a polypeptide that was highly conserved across mammalian, bacteria, yeast, and plant cells.<sup>5</sup> The first indication that Ub covalently modified other proteins came two years later, when an isopeptide bound histone-protein complex was identified by GoldKnopf and Busch, and the conjugated protein later recognised as Ub.<sup>6, 7</sup> In the 1980s, the link between the modification of proteins by Ub and their proteasomal degradation was elucidated by Ciechanover, Hershko and Rose.<sup>7-9</sup> Their work marks the birth of a now expansive field of research and later won the 2004 Nobel Prize in Chemistry.<sup>10</sup> Ub has since been associated with a plethora of other cellular processes including endocytosis, cell cycle control, transcription regulation and the repair of DNA damage.<sup>11, 12</sup> Such variety of function demonstrates the diversity of Ub as a PTM, whilst also illustrating the complexity of the ubiquitination process, which is controlled by several hundred enzymes.



### 1.1.2 Types of modification

As described in Section 1.1.1, Ub modifications can occur through isopeptide bond linkages. This isopeptide bond forms between the C-terminal glycine (Gly76, or G76) of Ub and lysine (Lys, or K) residues on substrate proteins. A substrate can be modified by one or several Ub moieties, referred to as mono- and multi-ubiquitination respectively (Figure 2). Mono-ubiquitination has been associated with several biological processes, including histone regulation, and DNA damage repair.<sup>13-16</sup> Further, both mono- and multi-ubiquitination have been linked to endocytosis.



**Figure 2: Cartoon representation of types of Ub modification, and their known physiological roles.**

As Ub possess seven Lys residues (K6, K11, K27, K29, K33, K48, and K63, Figure 1B) it can itself act as a substrate, leading to the formation of polyubiquitin (polyUb) chains. The topology of these chains is dependent on which Lys residues are ligated, providing various chain isoforms with different structures and hence biological effects.<sup>17, 18</sup> For example, K63 polyUb adopts a chain-like form, whilst K48 polyUb assumes a more compact globular structure. Some chain isoforms have been connected with their physiological functions. For example, K63 polyUb have a strong prevalence in cell signalling pathways whilst K48 and K11 polyUb modifications are predominantly associated with proteasomal degradation. Though bifurcated chains are also known to exist and it is conceivable that very complicated heterotypic chains could be formed, very little is known about such modifications due to low cellular abundance.<sup>19-21</sup>

In addition to isopeptide bond linkages, G76 can bind to the N-terminal methionine (Met1, or M1) of Ub to form head-to-tail linear Ub (Met1) chains. Ribosomal synthesis of Ub affords a Met1 polymer, which is then hydrolysed to provide monomer Ub units.<sup>22</sup> Structural studies have shown that Met1

chains have a similar topology to K63 chains, yet they play distinct cellular roles.<sup>23</sup> In 2006 Kirisako *et al.* showed that Met1 chains can be re-assembled by cells.<sup>24</sup> Since, the critical biological role of Met1 chains has been revealed, which will be considered in more detail in Section 1.1.6.3.<sup>25</sup>

### 1.1.3 Mechanisms of ubiquitination

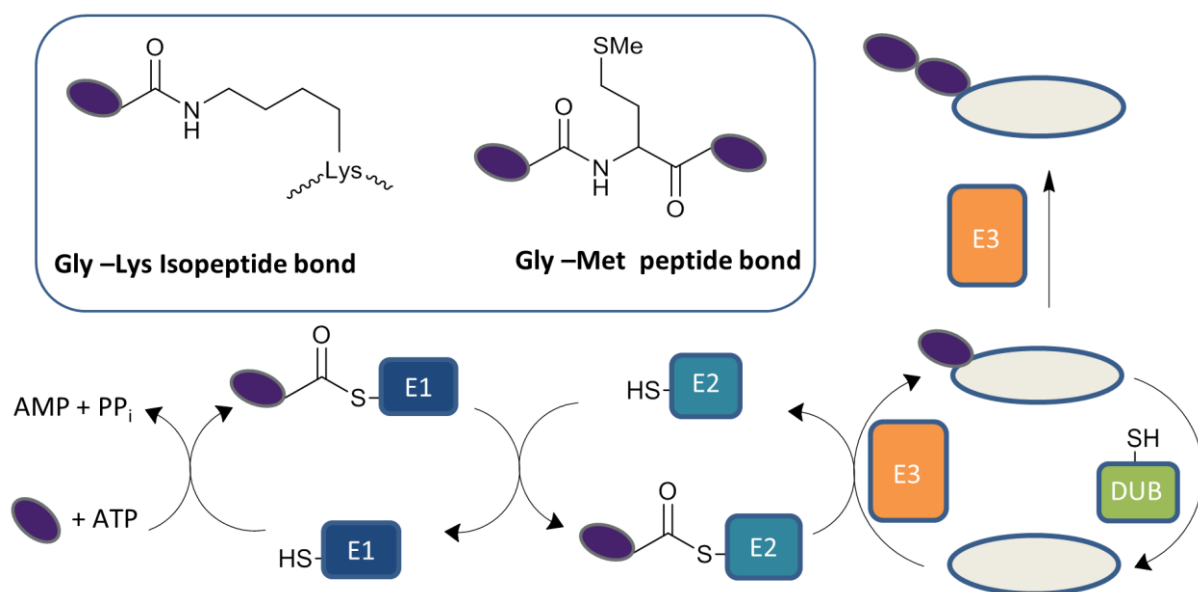
The plethora of Ub modifications available is made possible by an assortment of Ub enzymes. This Ub machinery enables precise control of several thousand target proteins through the use of different enzyme assemblies. The basic mechanism of ubiquitination was elucidated by Ciechanover and co-workers in the years following its discovery as a key player during proteasomal degradation. In 1981 the Ub activating enzyme (E1) was identified as the first piece of the puzzle, when it was shown that E1 covalently binds to Ub in an ATP dependent manner *via* a thioester bond.<sup>26, 27</sup> Ub-conjugating (E2) and Ub-ligase (E3) enzymes were later isolated and identified in 1983 by affinity purification through an Ub-sepharose column.<sup>28</sup> It was observed that E2 purification was only possible in the presence of E1 and ATP, which was not the case for E3. Further, E3 eluted from the column under salt conditions, whereas E1 and E2 did not, implying that the E3 was not covalently bound. Together this provides the current model of ubiquitination, as depicted in Figure 3. First, Ub is actively conjugated to an E1 enzyme *via* a thioester bond between Gly76 and a catalytic cysteine (Cys) residue. This Ub moiety is subsequently conjugated to an E2 enzyme *via* an analogous thioester bond to form an E2-Ub conjugate, with final transfer to a protein substrate Lys residue facilitated by an E3 ligase. The removal of such Ub PTMs is achieved by a family of deubiquitinating enzymes (DUBs).

In humans there are only two Ub E1 enzymes: UBA1 (Ubiquitin-like modifier-activating enzyme 1) and UBA6 (Ubiquitin-like modifier-activating enzyme 6).<sup>29</sup> Conversely ~40 E2s, and >600 Ub E3s have been identified to date.<sup>30, 31</sup> This hierarchy of enzymes allows high substrate specificity by means of the numerous E3s which enable selective modification of targets. Each extended family of Ub machinery is discussed in turn below.

#### 1.1.3.1 E2 enzymes

All E2 enzymes possess a ~150 amino acid catalytic domain referred to as the UBC domain.<sup>31</sup> The topology adopted by this domain is highly consistent, with additional C- and N-terminal extensions imparting enzyme specificity.<sup>31</sup> Some E2s are known to interact with multiple E3 ligase families, whilst others show specificity to a particular family. In general, E2-Ub conjugates have low reactivity and require interaction with an E3 to facilitate Ub transfer. For example, the UBE2D (Ubiquitin-conjugating enzyme E2 D) E2 family demonstrate low reactivity with Lys in isolation, but react

rapidly in the presence of their E3 partner.<sup>31, 32</sup> Multiple structural studies indicate E2-Ub conjugates adopt highly dynamic conformations, and that binding to its E3 partner promotes a reactive conformation.<sup>33-37</sup> The use of mutant E2-Ub conjugates in these studies, which are bound through a Lys rather than a Cys residue to mimic the transient thioester linkage, has allowed insight into these otherwise transient complexes. Interestingly, a small subset of E2s cannot initiate Ub chains, and instead are specific for polyUb chain elongation. These E2s do not possess intrinsic catalytic activity, and only facilitate Ub transfer when in complex with the substrate and E3 ligase partner. An example is UBE2N-UBE2V1 (Ubiquitin-conjugating enzyme E2 N- Ubiquitin-conjugating enzyme E2 variant 1) heterodimer, which is specific for K63 chain elongation.<sup>38, 39</sup>



**Figure 3: Cartoon representation of the Ub machinery assembly. Ub (purple) is actively conjugated by its Gly76 residue to an E1 enzyme *via* a thioester bond, and transferred by a cascade of E2 and E3 ligase enzymes to the substrate. Ub-enzyme conjugates are bound by a thioester bond between Gly76 and catalytic Cys residues. Ub is attached to the substrate *via* an isopeptide bond between its C-terminal G76 and Lys residues, or a Gly76–Met1 peptide bond in the case of linear chains. The removal of Ub PTMs is regulated by DUB enzymes.**

In addition to Ub conjugation, other regulatory roles of E2s are becoming increasingly more apparent. For example, UBE2D2 binds to the DUB OTUB1 (Ubiquitin thioesterase OTUB1) and in doing so enhances its protease activity. In turn, OTUB1 can bind to other E2-Ub conjugates, and in doing so inhibits their activity.<sup>31</sup>

### 1.1.3.2 E3 Ligase Families

E3 ligases fall into two major categories of catalysis; i) assisting direct Ub transfer from the E2-Ub conjugate to the substrate protein; ii) intermediary covalent binding to Ub and subsequent transfer

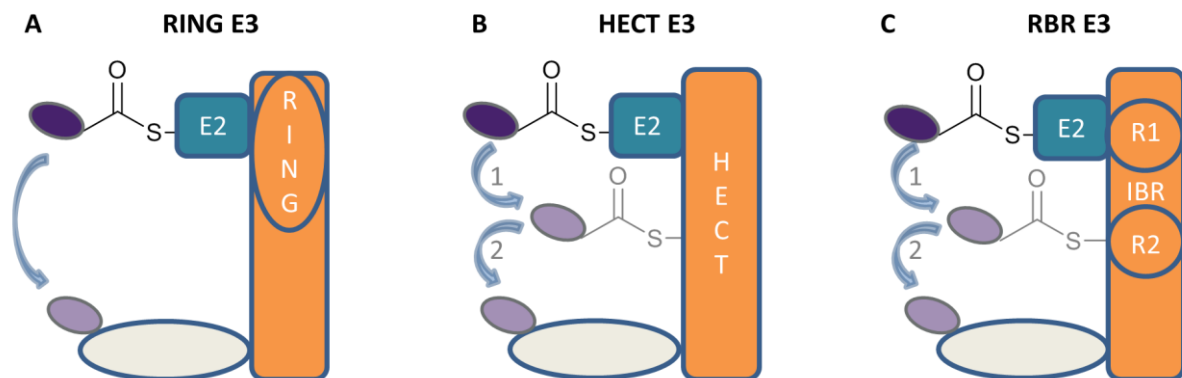
to the substrate proteins. This leads to 3 distinct E3 families: RING (Really Interesting New Gene), HECT (Homologous to E6AP C-Terminus), and RBR (RING between RING) ligases.<sup>30</sup>

Comprising ~ 600 enzymes to date, RING E3 ligases are the most abundant Ub ligase family. RINGs recruit E2-Ub conjugates and assist the direct transfer of Ub to substrate proteins (Figure 4A). RING domains characteristically consist of two zinc coordination sites which present a concave face to which E2-Ub binds. MDM2 (E3 ubiquitin-protein ligase Mdm2) is a RING E3 ligase, that mediates the proteasomal degradation of the tumour suppressor p53 (cellular tumour antigen p53).<sup>40</sup> U-box domains, which are also included in this family, are structurally analogous to RINGs but lack the coordinated zinc ions.<sup>30, 41</sup> RINGs and U-box E3s can be monomeric or homodimeric, whilst heterodimeric complexes are also known for RINGs. In the case of dimers, only one RING or U-box domain interacts with the E2 enzyme; the other domain is thought to be involved in complex stabilisation, or the recruitment of substrates. E3 ubiquitin-protein ligase CHIP (CHIP) is a homodimeric U-box E3 ligase, which has been shown to suppress apoptosis by targeting heat shock proteins.<sup>42</sup> RINGs can also be constituent parts of multi-subunit E3 complexes. This includes the cullin-E3 ligase family, which assemble on one of eight cullin scaffold proteins.<sup>43</sup> Another multi-subunit E3 ligase is the anaphase promoting complex/cyclosome (APC/C), which selectively targets cell cycle proteins for proteasomal degradation.<sup>30</sup>

In contrast, HECT E3 ligases covalently bind to Ub, prior to transfer onto the substrate (Figure 4B). HECTs represent a smaller E3 ligase family, with ~30 Human enzymes identified to date.<sup>30</sup> HECT ligases consist of two domains, the C lobe and N lobe, which are separated by a flexible tether to enabling variation of their relative orientation.<sup>30</sup> The C-lobe contains the catalytic Cys that binds to Ub, whilst the N lobe recruits E2-Ub. Variation in the characteristic domains and tether region affords differing Ub modification specificity. Substrates are largely recruited to the N-terminus, a process which is often facilitated by adaptor proteins. For example, SMURF2 (E3 ubiquitin-protein ligase SMURF2) is a HECT E3 ligase that uses at least two adaptor proteins, SMAD2 (Mothers against decapentaplegic homolog 2) and SMAD7 (Mothers against decapentaplegic homolog 7), to facilitate binding to substrates.<sup>44, 45</sup>

RBR ligases are another E3 ligase family consisting of ~12 members. RBR ligases utilise a hybrid of both RING and HECT catalytic mechanisms. RBRs consist of a RING1 and a RING2 domain separated by an IBR (In-between-RING) domain (Figure 4C). Though its historical naming is misleading, Wenzel *et al.* have since shown that RING2 behaviour is in fact akin to a HECT domain; E2-Ub conjugates bind to RING1, prior to Ub transfer to a RING2 catalytic Cys and ultimately to the substrate.<sup>32</sup> As with

RING domains RBRs can form part of multi subunit complexes. Met1 chains are assembled by an RBR E3 ligase containing complex, aptly named LUBAC (linear ubiquitin chain assembly complex).



**Figure 4: Cartoon representation of RING, HECT and RBRs E3 ligase families. RING E3s assist direct transfer of Ub to the substrate, whilst HECT and RBR E3 ligases employ catalytic Cys residues to form thioester bound E3-Ub conjugates prior to substrate transfer. RBR ligases possess a RING1 (R1) and RING2 (R2) domain. R1 mediates E2-UB binding, whilst Ub is transferred to the R2 domain.**

Analogous to E2-Ub conjugates, transfer of Ub to substrate Lys residues from E3-Ub conjugates is dependent on conformational changes. Recent studies of have provided structural insight into these changes and transfer mechanisms of HECT and RBR E3 ligases.<sup>46, 47</sup>

### 1.1.3.3 'E4' ligases

In some cases, a further 'E4' ligase, also termed a ubiquitination elongation factor is required to enable the formation of polyUb chains. E4 ligases either function enzymatically as an additional E3 ligase in the assembly, or they facilitate the progression of an existing E3 ligase. Several E4s possess a characteristic U-box domain; however, for some time it was a matter of debate as to whether such enzymes should be considered as a separate family of Ub machinery, or an E3 ligase subfamily.<sup>48</sup> Current evidence suggests the former is a more appropriate model on two counts: i) there is no evidence of a direct interaction between 'E4' and E2 enzyme; ii) not all E4s possess E3 activity *in vivo*, and as such cannot replace E3s in an enzyme assembly. UBE4B (Ubiquitin conjugation factor E4 B) is an E4 ligase, that in combination with the E3 MDM2 polyubiquitinates p53, leading to its degradation.<sup>40</sup>

### 1.1.3.4 DUBs

Approximately 100 DUBs are known to date, which can be classified into five families.<sup>49</sup> Four of these families: Ubiquitin carboxyl-terminal hydrolases (UCH), Ubiquitin-specific proteases (USPs), ovarian tumour proteases (OTU) and Josephins (MJD) are cysteine proteases. The fifth family comprises a

group of metalloproteases referred to as MPN+/JAMMs ((JAB1/MPN/MOV34 metallo-enzymes). USP represents the largest subfamily, comprising of 56 members.

Mechanistically, DUBs fall into three modes of action; i) generation of free Ub, either from polyUb chains or ribosomal fusion proteins; ii) removal of monoUb motifs; iii) polyUb 'chain editing': i.e. conversion between chain isoforms.<sup>49</sup> As with E2s and E3s, different DUBs demonstrate different specificities. For example, the UCH family are, amongst other roles, associated with hydrolysis of ribosomal Ub polymer.<sup>50</sup> Other DUBs demonstrate differing chain-linkage specificity.<sup>22</sup> For example; USP14 is proteasomal bound DUB that specifically cleaves K63 chains. In this way USP14 facilitates the recycling of Ub monomers following substrate targeting for degradation. On the other hand, USP5 and USP13 have been shown to cleave with no chain specificity. These DUBs target 'free' Ub chains, which are not attached to substrate proteins, to replenish the monomer Ub pool.<sup>51</sup> Conversely the OTU family show high linkage specificity, but the topology of this specificity varies greatly within the family.<sup>38</sup> For example, OTUB1 shows specificity for K48 chains, whilst OTULIN (Ubiquitin thioesterase otulin), OTUD7A (OTU domain-containing protein 7A, also known as Cezanne), and OTUB7B (OTU domain-containing protein 7B, also known as A20) specifically cleave Met1, K11 and bifurcated K11/K48 chains respectively.<sup>31, 38</sup>

The various Ub modifications and their associated assembly machinery is a highly complex field. For a comprehensive overview of known Ub machinery, their associated assembly systems, modification specificity, and substrate specificity, please refer to the following excellent reviews.<sup>22, 31, 52-54</sup>

For a list of all known Ub machinery in the UniProt database, please refer to Appendix Table 1.

#### **1.1.4 Recognition of Ub PTMs: UBDs**

The code generated by various Ub modifications is detected and deciphered by Ub binding domains (UBDs) present on proteins. These domains associate with Ub motifs through PPIs, in turn triggering a cellular response. Such interactions with Ub are low affinity (10-500  $\mu$ M), reflecting their transient nature.<sup>55</sup> The first UBD to be identified was a subunit of the 26S proteasome, S5a. In 1994 Deveraux *et al.* determined that this 50 kDa substrate is capable of binding Ub-conjugates and polyUb chains at least 4 Ub in length.<sup>56</sup> Now more than 20 UBDs have been identified, covering a broad landscape of structural complexity.<sup>57</sup> Structural studies indicate that UBDs interact through a variety of distinct PPIs, utilising various secondary structures, such as  $\alpha$ -helices,  $\beta$ -sheets and zinc finger domains. Furthermore, a given protein can possess multiple, homo- or hetero-, UBD domains enabling a wide range of linkage specificity.<sup>58</sup>

Conversely, the mode of Ub binding to these UBDs is highly conserved to a limited number of surface 'patches', including the isoleucine 44 (Ile44) hydrophobic 'hotspot'.<sup>55, 58</sup> However, different chain isoforms present different protein surfaces for subsequent interaction with UBDs. Accordingly different UBDs have different chain link specificities, even though they interact through mainly conserved residues. Ub-UBD interactions are often stronger with an isolated UBD domain, suggesting that these PPIs are regulated by intra, as well as inter- molecular interactions in the native environment.<sup>55</sup> Further regulation is likely to be achieved by other PTMs.

### 1.1.5 Ubiquitin like modifications (Ubls)

In addition to Ub a number of Ubiquitin-Like Modifications (Ubls) have been identified in eukaryotes and prokaryotes. Ubls are related to Ub through structural similarity and all possess a  $\beta$ -grasp fold core 3D structure, consisting of a  $\beta$ -sheet cradling a single helix (Figure 5).<sup>59</sup> Although Ub-like, these PTMs are distinct, with generally low sequence homology. Each eukaryotic Ubl is briefly discussed below.

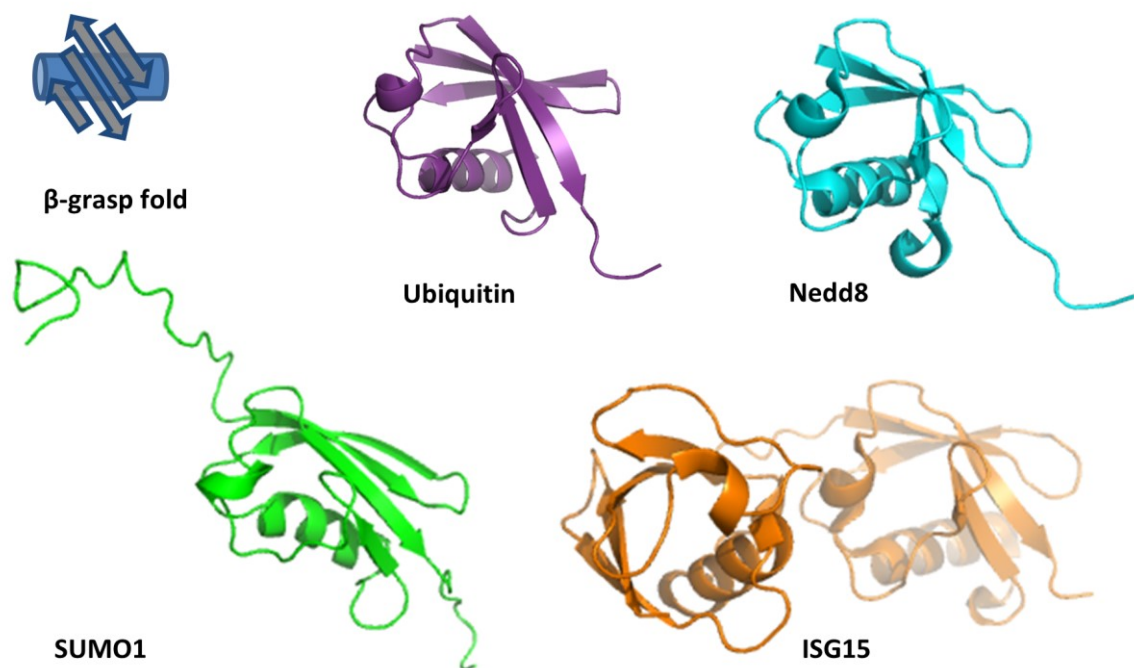
#### 1.1.5.1 SUMO

SUMO (Small Ubiquitin like modifier) was first identified in 1997 by Mahajan *et al.* as a conjugated adduct of RanGAP1 (Ran GTPase-activating protein 1) in mammalian cells, and has since been linked to several regulatory cellular events including DNA repair, apoptosis and cell cycle progression.<sup>60-62</sup> 103 Amino acids in length, SUMO is a larger protein than Ub (Figure 5). At least 4 genes encode SUMO in invertebrates: SUMO1-4. SUMO2 and SUMO3 are 97 % identical and cannot be distinguished by antibodies, therefore these isoforms are considered collectively as SUMO2/3.<sup>60</sup> Sumoylation requires an analogous yet simplified assembly to ubiquitination; there is one E2, a handful of E3s, and a small subfamily equivalent to DUBs called SENPs (senrin specific proteases) for SUMO removal. Like Ub, SUMO can form chains in yeast and mammalian cell. SUMO can also be modified by Ub leading to hybrid chains but the biological implications of both of these modifications remains elusive.<sup>63, 64</sup> Together, SUMO and Ub modifications can elicit precise control over biological processes. A well understood example of Ub and SUMO 'crosstalk' is their competitive binding to residue K164 of PCNA (proliferating cell nuclear antigen) during S phase. PCNA is a processing factor which is essential for DNA replication. K164 is mono-ubiquitinated if the DNA is damaged, resulting in stalled replication and initiation of DNA repair. Conversely, SUMO binds when there is no damage. In this way, SUMO and Ub binding act as a 'switchboard' between DNA repair and replication modes of PCNA.<sup>3, 65</sup>

### 1.1.5.2 *Nedd8*

Also discovered in 1997, Nedd8 is the closest Ubl to Ub in terms of sequence, with 58 % homology (Figure 5). Loss of function in the Nedd8 pathway was found to be lethal in plant, animal and some yeast models.<sup>60</sup> Like sumoylation, neddylation machinery is a simplified version of the ubiquitination machinery, with one E1 and just a handful of E2s and E3 ligases. Nedd8 may form chains, but their function is unknown to date.<sup>66</sup> Nedd8 is known to modify, and in doing so regulate, ubiquitination machinery. Notably Nedd8 modifies cullin-E3 ligase activity.<sup>67</sup> Cullin-E3 ligases are auto-inhibitory and only activated following neddylation at a conserved Lys residue. Indeed, mutation studies indicate an accumulation of cullin substrates if this Lys residue is not preserved.<sup>68</sup> Neddylation induces a conformational change and thus activates E3 ligase activity. Nedd8 modifications can be removed by a number of DUBs, and further regulation is achieved by CAND1 (cullin-associated NEDD8-dissociated protein 1), which prevents neddylation thereby preserving the inactive state.<sup>60, 69,</sup>

70



**Figure 5: PyMOL representations of Ub, Nedd8, SUMO1, and ISG15. A cartoon representation of the  $\beta$ -grasp fold core structure shared by Ub and Ubls is included for reference.**

### 1.1.5.3 *ISG15*

ISG15 (Interferon-simulated gene 5, Figure 5), as the name suggests, is expressed when stimulated by interferons (IFN). As for other Ubls, ISG15 machinery is simpler than Ub machinery, and interestingly the expression of this machinery is also induced by IFN.<sup>60</sup> ISG15 plays a role in IFN pathways during host antiviral response following viral infection, and its expression is limited to the



higher eukaryotes that utilise this signalling pathway. ISG15 is known to modify host and viral targets.<sup>71</sup> Akin to neddylation ISG15 modifications can regulate the activity of Ub machinery, and it is in this way that host cells impair virus 'budding' mechanisms.<sup>72, 73</sup> Poly-ISG15 chains have not been identified, whilst multi-ISG15 modifications are prevalent.

#### 1.1.5.4 Other Ubls

FAT10 (HLA-F adjacent transcript 10) and MNSF $\beta$  (monoclonal nonspecific suppressor factor  $\beta$ ) are two further Ubls. FAT10 is a reported oncogene, whilst MNSF $\beta$  may be associated with T cell function during the immune response.<sup>74-78</sup> UFM1 (Ubiquitin-fold modifier 1) is one of the newest additions to the Ubl family. The role of UFM1 remains unclear, but its specific E1 enzyme is essential for red blood cell differentiation in mice and is expected to play other developmental roles, either later stage or in alternative cell types.<sup>60, 79</sup> Two Ubls: ATG8 (Autophagy-related protein 8) and ATG12 (Ubiquitin-like protein ATG12), represent key components of the autophagy machinery. ATG12 is a canonical Ubl that modifies protein, whilst ATG8 is a lipid modifier existing as eight known isoforms in mammalian cells. These Ubls have been reviewed extensively by Van der Veen *et al.*<sup>60</sup>

Other eukaryotic Ubls function by a non-canonical mechanism; that is, they do not modify substrate through a Gly-Lys isopeptide bond. For instance, HUB1 (Ubiquitin-like modifier hub1) acts as a non-covalent modifier in yeast, through a C-terminal di-tyrosine motif, and is essential for RNA splicing.<sup>60</sup> URM1 (Ubiquitin-related modifier 1) which is conserved in yeast and human cells, is implicated in tRNA thiolation. URM1 possesses a C-terminal thiocarboxylate and may act as a sulphur carrier.<sup>80</sup>

## 1.1.6 Cellular roles and chemotherapeutic potential

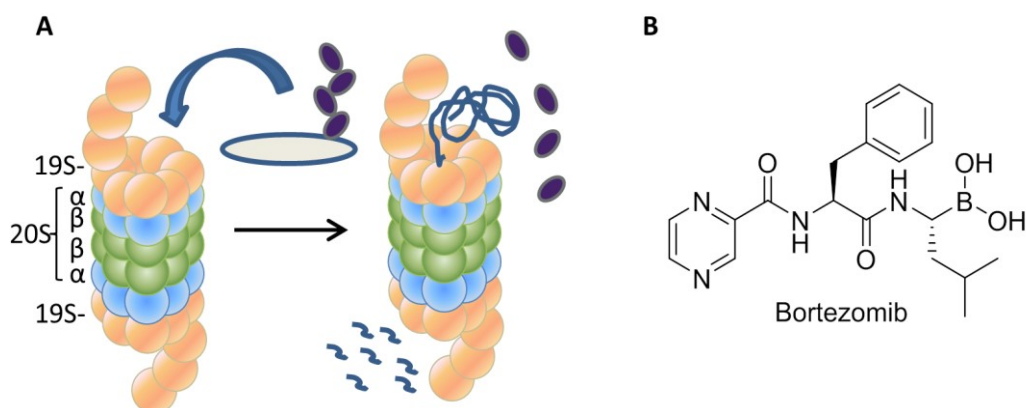
As mentioned in sections 1.1.1 and 1.1.5, Ub and Ubl modifications play critical roles in several biological processes. Here three such processes which offer chemotherapeutic potential are introduced: the ubiquitin proteasome system (UPS), DNA damage response (DDR), and NF- $\kappa$ B (nuclear factor kappa-light-chain-enhancer of activated B cells) signalling, along with their associated Ub/Ubl modifications.

### 1.1.6.1 Ubiquitin Proteasome system (UPS)

The 26S proteasome, generally referred to as the 'proteasome' is a 2000 kDa multi-protein complex conserved across all eukaryotes.<sup>29</sup> Found in the cytoplasm and nucleus, the proteasome is composed of three subunits; the core 20S particle, which is sandwiched by two 19S regulatory particles (Figure 6A). The 20S can be further classified into two outer  $\alpha$  rings, which facilitate binding to 19S, and an inner  $\beta$  ring that enables proteolytic activity. Both  $\alpha$  and  $\beta$  rings are formed from distinct subunits. In the case of the  $\beta$  ring, three of these subunits each provide a specific catalytic activity to facilitate proteolysis: trypsin-like, chymotrypsin-like, and peptidylglutamyl-peptide hydrolysing activity.<sup>81</sup> Proteins which are targeted for degradation by a K48 polyUb modification bind to the 19S particle *via* their polyUb motif. The 19S particle then ensures cleavage and recycling of Ub, whilst the 20S particle ultimately degrades the substrate protein. In this way, the UPS regulates the cellular environment by controlling levels of redundant or defective proteins.

Dysregulation of the proteasome leads to cellular stress. This makes the proteasome a prime drug target for certain cancers, which are highly dependent on UPS regulation.<sup>29</sup> One such example is Multiple myeloma (MM). MM is cancer of plasma cells, which are responsible for antibody production in bone marrow. MM cells are proliferative and produce defective proteins in excess, making them highly dependent of UPS function.<sup>82</sup> Proteasome-targeting drugs have since been used with success for MM patients, and the first UPS inhibitor, Bortezomib (Velcade, Millennium Pharmaceuticals, Figure 6B) was approved in 2003 by the FDA (Food and Drug Association).<sup>83</sup> Since then, additional proteasomal drugs have undergone clinical trials in both haematological and solid tumours, with further inhibitors under clinical investigation to improve drug efficacy and limit toxicity.<sup>29</sup>

Proteasomal inhibition is by no means a 'magic bullet' and its efficacy varies between different types of cancer. Proteasomal inhibition impacts on other cellular pathways including DNA repair, NK- $\kappa$ B signalling, and cell cycle arrest.<sup>84</sup> Accordingly, the responsiveness of a malignancy to proteasomal inhibition can be related to its dependency on these pathways.



**Figure 6: (A) Cartoon representation of UPS. PolyUb chains target substrate proteins for degradation by the proteasome, and the Ub chain is cleaved and recycled by DUBs. The major subunits of the proteasome are labelled. (B) Structure of Bortezomib, the first FDA approved proteasome inhibitor.**

#### 1.1.6.2 DNA Damage Response (DDR)

Ub was first associated with DDR in 1987 when Jentsch *et al.* identified that RAD6 (Ubiquitin-conjugating enzyme E2 2), a known DNA repair protein in yeast, demonstrated E2 activity.<sup>85</sup> RAD6, together with RAD18 (E3 ubiquitin-protein ligase RAD18), mono-ubiquitinates PCNA on sensing DNA damage during replication.<sup>65</sup> Since, Ub and Ubl modification have been linked to all the DDR pathways triggered in response to different forms of DNA damage.

DNA damage can occur on one strand, leading to Single Strand Breaks (SSBs) or the formation of bulky adducts (Figure 7A). Alternatively damage can occur simultaneously on both strands, resulting in Double Strand Breaks (DSBs) or interstrand cross-links (ICLs). Further, endogenous errors can occur during DNA replication such as insertion, deletion, or mismatching of base pairs. This variety of damage requires several repair pathways in order to retain genomic integrity. The role of Ub in these pathways can be regulatory through the UPS, which is observed during SSB repair. Alternatively Ub can play a signalling role, assisting the recruitment of DNA repair proteins to sites of damage. Some roles of Ub modifications during DDR are briefly introduced below.

Nucleotide excision repair (NER) is the pathway utilised to repair bulky adducts. NER is achieved *via* two pathways: Transcription Coupled Repair (TCR) and Global Genome Repair (GGR). TCR acts as a rapid but error prone repair mechanism, whilst GGR provides an error free alternative. It is now known that Ub modifications provide the link between these two pathways and trigger GGR in the event of failed repair by TCR.<sup>86</sup> On sensing damage during transcription, TCR is first initiated. Once the repair machinery is assembled, Rpb1 (DNA-directed RNA polymerase II subunit RPB1), a subunit of this repair machinery, is modified with polyUb chains by Rsp5 (E3 ubiquitin-protein ligase RSP5) and UBC5 (Ubiquitin-conjugating enzyme E2-16 kDa). The progression of this polyUb modification

acts as an Ub 'clock'. If the DNA is repaired in time, transcription resumes. However if the polyUb chains grows to four Ub in length, TCR is abandoned, the damage site cleared of TCR machinery, and GGR commences.<sup>87, 88</sup> Other Ub clocks are known to control other repair pathways, such as the repair of damage encountered during replication.<sup>3</sup>

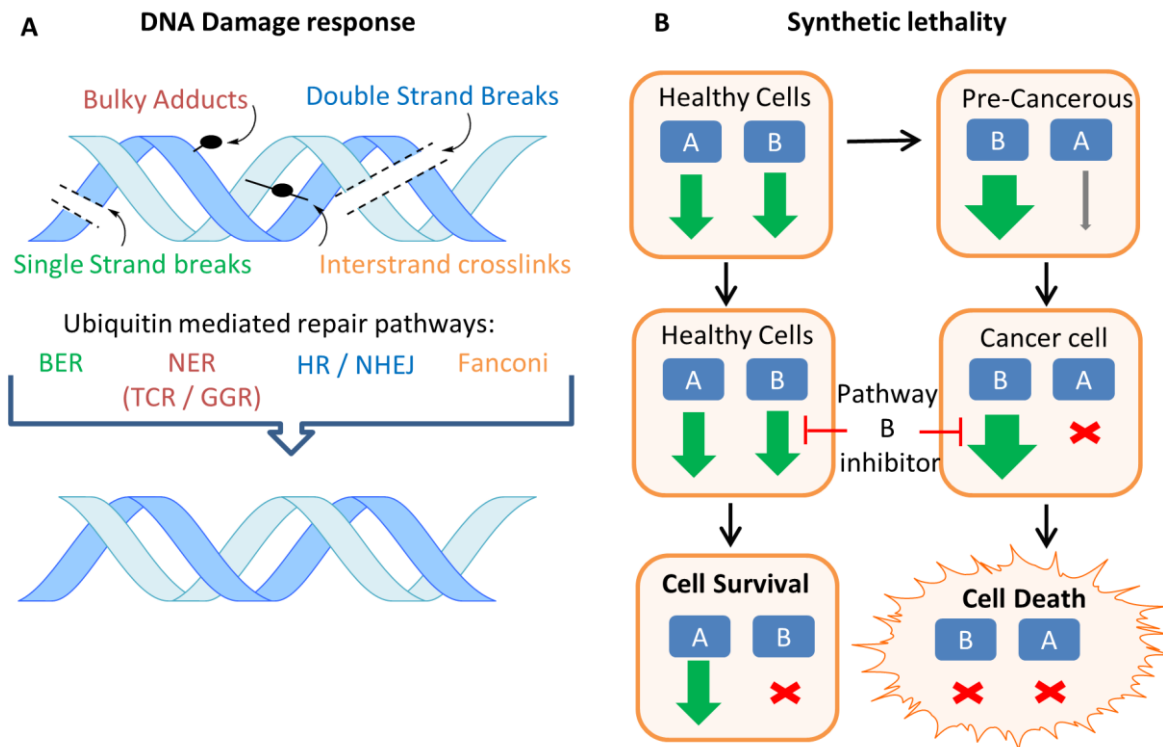
Two cullin-E3 ligases: Cullin-4A (CUL4A) and Cullin-4B (CUL4B) are also associated with NER.<sup>89</sup> As discussed in Section 1.1.5.2, cullin-E3 ligase activity is regulated by Nedd8 modifications. Loss of neddylation leads to an accumulation of cullin-E3 ligase substrates, and ultimately S-phase cell cycle arrest.<sup>90</sup> This results in DNA re-replication which constitutes as DNA damage and triggers DDR.<sup>91, 92</sup> Nedd8 inhibition has been further shown to sensitise cells to DNA damage induced by cisplatin and IR, and recently localisation of Nedd8 at damage sites has been observed, making Nedd8 machinery a novel anti-cancer target.<sup>93-97</sup>

DSBs are the most cytotoxic form of DNA damage. K63 polyUb chains are known to assemble at DSBs following phosphorylation of Histone H2AX.<sup>98</sup> Three independent studies in 2007 discovered that such chains are assembled by E3 ubiquitin-protein ligase RNF8 (RNF8) and the E2 UBE2N.<sup>99-101</sup> In human cells RNF8 was later shown to further interact with phosphorylated E3 ubiquitin-protein ligase HERC2 (HERC2), leading to a stabilised RNF8/UBE2N interaction, and promoting RNF8 dependent ubiquitination events in downstream DSB repair.<sup>102</sup> E3 ubiquitin-protein ligase RNF168 (RNF168), is further required for the recruitment of the repair proteins BRCA1 (Breast cancer type 1 susceptibility protein) and 53BP1 (Tumor suppressor p53-binding protein 1) to sites of damage.<sup>103</sup> Depletion of RNF8 or RNF168 sensitises cells to DSB-induced damage.<sup>98</sup> It is now understood that RNF168, which is negatively regulated by the DUB OTUB1, binds to polyUb motifs created by RNF8 and promotes polyUb chain formation on histone H2A, which in turn leads to the recruitment of repair proteins.<sup>98, 104</sup> In this way, RNF8/RNF168 seemingly controls recruitment of both BRCA1 and 53bp1, even though these are associated with separate DSB pathways: Homologous recombination (HR) and Non Homologous End Joining (NHEJ) respectively.<sup>98</sup> As with NER, these two pathways provide both error free and error prone repair mechanisms. The mechanistic relationship between 53bp1 and BRCA1 and subsequent repair pathway selection remains a complicated puzzle to crack.<sup>98</sup> Recent studies show that E3 ubiquitin-protein ligase RNF169 (RNF169), a close relative of RNF168, negatively regulates DSB repair pathways by competing with 53BP1 and BRCA1 binding to RNF8 assembled polyUb chains. This suggests that RNF169 contributes to the DSB pathway balancing act. BRCA1 itself contains a RING domain, and complexes with BARD1 (BRCA1-associated RING domain protein 1) to function as a heterodimeric RING E3 ligase. Further E3 ligases have also been implicated in DDR and have been extensively reviewed by Brown *et al.*<sup>98</sup>

Finally, DNA damage induced by ICLs is sensed and repaired through the Fanconi pathway in a Ub mediated manner.<sup>105</sup> FANCD2 ( Fanconi anemia group D2 protein) and FANCI (Fanconi anemia group I protein) together form the heterodimeric 'ID complex'. On sensing ICLs, mono-ubiquitination of both components of this complex is achieved by E3 ubiquitin-protein ligase FANCL (FANCL) in combination with UBE2T (Ubiquitin-conjugating enzyme E2 T), leading to the recruitment of the repair complexes to damaged chromatin, with enzymatic turnover regulated by the DUB USP1.<sup>98</sup> This brief introduction offers just a taste of the physiological roles Ub/Ubls play during DDR. For a comprehensive review refer to one of the suggested references.<sup>49, 98</sup>

Targeting DDR pathways is an emerging strategy to improve the efficacy of existing cancer therapeutics. As radio- and chemo-therapies induce DNA damage, DDR pathways offer a mechanism of resistance for cancer cells.<sup>49</sup> As described above, damage can often be repaired by overlapping, error prone and error free, pathways. In cancer, these DDR pathways can become over-activated or defective. This imbalance alters the cell's DDR repertoire leading to a strong reliance on particular pathways, which can be exploited therapeutically; inhibition of critical pathways can prove lethal for malignant cells, whilst healthy cells survive (Figure 7B).<sup>49</sup> This concept is referred to as synthetic lethality, and is also used in the context of therapeutic strategies combining chemotherapeutics with additional genotoxic agents.

PARP (poly ADP ribose polymerase) enzyme inhibitors are a recent addition to the chemotherapeutic arsenal, which nicely illustrate synthetic lethality in action. Current PARP inhibitors inhibit PARP1 and PARP2, essential enzymes for SSB sensing and repair.<sup>106</sup> The PARP inhibitor Olaparib (Lynparza, AstraZeneca), recently approved by the FDA, has proven to be successful in treating hereditary prostate malignancies which display defective BRCA1 or BRCA2 (Breast cancer type 2 susceptibility protein) genes.<sup>107</sup> BRCA2, like BRCA1 is critical for HR DSB repair. PARP inhibition prevents SSB repair, leading to DSB formation during replication, which cannot be efficiently repaired by the defective HR.<sup>98</sup> Following the demonstrated success of PARP inhibitors, multiple companies and research institutions are developing drugs which employ a synthetically lethal strategy, with several ongoing studies targeting Ub machinery associated with DDR. Although no drugs targeting DDR ubiquitination machinery are currently approved, several DDR relevant DUB and E3 ligase inhibitors are undergoing preclinical trials, and a Nedd8 E1 enzyme inhibitor is in phase I clinical trials.<sup>108</sup>



**Figure 7: Cartoon representations of DDR, and synthetic lethality. (A) The major types of DNA damage and the Ub mediated repair pathways associated with them. (B) Certain cancer cells become reliant on particular repair pathways, inhibition of which leads to cancer cell death but healthy cell survival.**

### 1.1.6.3 *NF- $\kappa$ B pathway*

NF- $\kappa$ B signalling refers to a small family of transcription factor proteins that control the expression of proteins critical to cell survival during inflammatory and immune responses.<sup>109</sup> NF- $\kappa$ B proteins (NF- $\kappa$ Bs) are inactively present in cells, enabling a fast response to stimuli without requiring additional protein synthesis. This is achieved through their complexation to inhibitors of NF- $\kappa$ B (I $\kappa$ Bs) that sequester NF- $\kappa$ Bs in the cytoplasm. In response to stimuli, these I $\kappa$ Bs are degraded and the free NF- $\kappa$ Bs released, which subsequently enter the nucleus. Two types of NF- $\kappa$ B signalling are known: the canonical and non-canonical pathway. For the canonical pathway, Met1 polyUb chains have been identified as a novel pathway activator; modification of NEMO (NF- $\kappa$ B essential modulator) by the RBR E3 ligase LUBAC leads to its recruitment, which in turn activates the NF- $\kappa$ B signalling pathway.<sup>25</sup> Consequently I $\kappa$ Bs are targeted for degradation by K11 polyUb modifications, enabling NF- $\kappa$ Bs to translocate to the nucleus. Several DUBs are reported to negatively regulate NF- $\kappa$ B signalling, including CYLD (Ubiquitin carboxyl-terminal hydrolase CYLD), A20, and OTULIN.<sup>109</sup> Conversely, non-canonical NF- $\kappa$ B signalling is not regulated by linear ubiquitination. Seemingly distinct, studies by Basak *et al.* revealed that in fact these two pathways are mechanistically linked; suggesting that loss of canonical function will lead to atypical non-canonical function.<sup>110</sup>

Constitutively active NF- $\kappa$ B signalling is a common signature in a range of human cancer, with unchecked proliferation enabling continued tumorigenesis and metastasis of malignant cells whilst apoptosis is suppressed.<sup>111</sup> NF- $\kappa$ B signalling is also activated by genotoxic agents.<sup>112</sup> Consequently, components of the NF- $\kappa$ B signalling pathway are an interesting anti-cancer target, as their inhibition could inhibit tumour proliferation, whilst also sensitising cells to chemotherapeutics. A recent genetic screen revealed that a component of the E3 complex LUBAC attenuates cisplatin induced apoptotic cell death, implying that NF- $\kappa$ B activation by LUBAC is a contributing factor in cisplatin resistance.<sup>113</sup>

It has been suggested that NF- $\kappa$ B pathway inhibition in combination with UPS inhibition could offer a chemotherapeutic strategy. It was initially believed that UPS inhibition would directly inhibit NF- $\kappa$ B signalling, due to hindered I $\kappa$ B degradation.<sup>114</sup> This hypothesis was supported by the observed accumulation of I $\kappa$ B following Bortezomib treatment in clinical trials.<sup>84</sup> However, further studies which observed activation of NF- $\kappa$ B signalling on UPS inhibition contradict this rationale, and suggest that combined inhibition could prove synthetically lethal.<sup>84</sup>

## **1.2 Studying ubiquitination: a proteomics approach**

### **1.2.1 Proteomics: the challenges of studying PTMs**

Proteomics is the study of proteins. Through genetics we have established the 'blue print' of life: the genome, which details all the genetically encoded proteins of the cell, known collectively as the proteome.<sup>115</sup> The completion of the human genome project and the rise of genetic tests have powered the concept of personalised medicine, however, genomics alone does not provide all the information we need to understand biological processes.

The genome is static and therefore doesn't provide any dynamic information for the proteins that it encodes. Yet protein expression levels can vary dramatically in response to a number of biological factors. For instance, the proteome fluctuates throughout the cell cycle and in response to outside stimuli.<sup>116</sup> It also differs between healthy and diseased cellular states. Add to this a resource of PTMs, plus additional DNA variations such as RNA splicing and isomerism, and the reality of studying proteins becomes far more complex. Consequently, sophisticated techniques that enable spatiotemporal control, coupled with high throughput and high sensitivity, are required to study the proteome and its related PTMs.

There are two main aims when studying PTMs: i) to discover substrates of a given modification; ii) to identify the enzymes required to catalyse this modification. Since the discovery of Ub much work has been undertaken to study the modification and its associated machinery. Various methods, particularly mass spectrometry (MS) approaches have been applied, and will be summarised in the following sections.

### **1.2.2 Mass spectrometry (MS) approaches**

Cracking the human genome and that of other organisms has led to the cataloguing of genome sequences for all encoded proteins to create vast protein databases, which in combination with sophisticated mass spectrometry (MS) technologies fuelled the nascent of MS-based proteomics.

MS is an analytical method used to separate, detect, and measure ionised chemical species according to their mass-to-charge-ratio ( $m/z$ ). Species can be ionised in a variety of ways, generally referred to as soft and hard ionisation techniques, which enables a broad range of chemical moieties to be detected and various experimental designs to be employed. In the 1980s it was demonstrated that soft ionisation techniques, such as MALDI (matrix assisted laser desorption/ionisation) and ESI



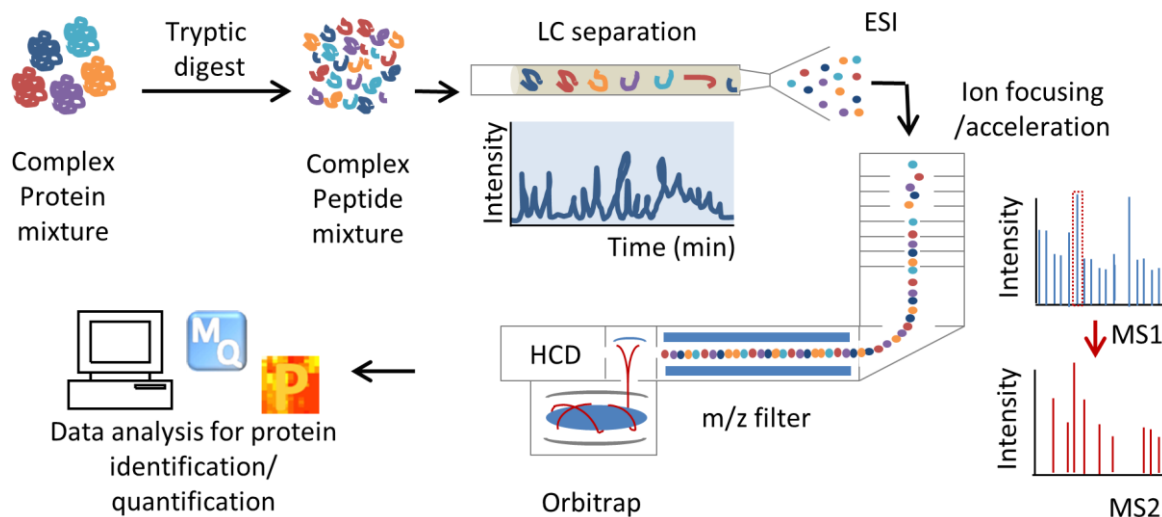
(electrospray ionisation) can be used to ionise proteins and peptides, enabling their identification by MS methods.<sup>117, 118</sup>

Tandem MS (MS/MS) instrumental setups, where both the mass and sequence of multiple peptides can be determined from a single analytical run, has revolutionised our ability to study complex proteomes.<sup>119</sup> In an MS/MS experiment intact ions are detected in the first MS stage (MS1) before some ions are selected for fragmentation and further MS analysis (MS2) to enable sequence determination. Quadrupole Time of flight (Q-TOF) and Orbitrap instruments are commonly used mass detectors for MS/MS proteomic experiments. With Q-TOF, ions are detected separately as they reach the detector at different time points. In the Orbitrap, ions are not separated in time, but can be differentiated by their oscillation frequency, which can be converted to MS spectra by Fourier transform. As with ionisation, ions can be fragmented in a number of ways. Collision Induced Dissociation (CID) in the gas phase is commonly employed for analysis of complex proteomic mixtures.<sup>120</sup> Thus following a database search, proteins and peptides can be identified by their amino acid sequence.

Multiple software packages that facilitate MS spectral processing and protein database searching exist, including the freely available MaxQuant software.<sup>121</sup> Protein assignment is achieved by search algorithms, which fit MS and MS/MS spectral data to sequences recorded in protein databases. This fit is of course a model rather than concrete proof, and as such each 'fit' has an associated error, referred to as a 'false discovery rate' (FDR). Further complex algorithms fit these peptides to proteins, again with an associated FDR. FDR cut off thresholds can be applied to define the confidence of a protein assignment. Additional software, such as Perseus, enables data evaluation and visualisation. Further analytical tools, including the free online tool STRING, provide bioinformatic analysis of MS data, such as the Gene Ontology (GO) annotations associated with identified proteins, the known interacting partners and pathways of a protein, or the known interactions between a subset of proteins.<sup>122</sup>

There are two main methods of protein MS analysis: the top-down and bottom-up approaches.<sup>123</sup> Bottom-up approaches, which identify proteins from their constituent peptides, are well suited for the analysis of complex mixtures, and can be used to identify thousands of proteins and modifications within a few hours. Proteins are typically digested using trypsin, which cleaves proteins at Lys and arginine (Arg, or R) residues, but other enzymes, such as Lys-C which only cleaves at Lys residues, offer alternative peptide generation.

Bottom-up proteomics can be undertaken in a sort-then-break or break-then-sort manner. Sort-then-break requires the separation of proteins prior to their enzymatic digestion and subsequent MS/MS analysis. Sorting is achieved by 1D or 2D gel electrophoresis, with digestion of excised bands providing a peptide ‘fingerprint’ of the protein in question, enabling its identification.<sup>124</sup> This technique is limited by the sensitivity of the protein staining method used and therefore is not ideal for studying low abundant proteins or modifications. Alternatively, break-then-sort involves initial peptide generation with subsequent separation by liquid chromatography (LC).<sup>125</sup> This LC-MS/MS methodology (Figure 8), commonly known as ‘shot-gun’ proteomics provides a much higher throughput, which is capable of detecting low abundance proteins and even membrane proteins. LC-MS/MS is limited by its detection ability, this being subject to the resolution of separation and the abundance of the peptides. Both of these factors are exacerbated as sample complexity increases.



**Figure 8: Cartoon schematic of a bottom-up LC-MS/MS workflow using an orbitrap mass spectrometer. A protein mixture isolated from cells is enzymatically digested, separated by liquid chromatography and subjected to MS/MS analysis. Peptides are ionised by ESI to provide peptide masses in MS1, selected, and fragmented by HCD (Higher-energy collision dissociation, a form of CID commonly used with orbitrap mass spectrometers) before MS2 analysis. Data analysis of MS1 and MS2 spectra, together with peak intensities, enables protein identification through MaxQuant software, and data evaluation using Perseus software.**

In comparison, whole proteins can be separated and analysed by top-down approaches.<sup>126</sup> This enables different biological states of a protein, or ‘proteoforms’, to be detected by their differing accurate masses.<sup>127</sup> In this way a PTM (such as ubiquitination) along with any co-occurring PTMs (such as acetylation, phosphorylation, and further ubiquitination) can be observed simultaneously, in the context of the whole protein. This provides important information beyond the scope of bottom-down approaches; it is almost certain that several PTMs work in unison to control proteins, and this ‘whole picture’ is lost when changes are detected at the peptide level.<sup>128</sup> As MS sensitivity

and separation techniques steadily improve high throughput top-down proteomic approaches become increasingly plausible, especially for lower molecular weight protein mixtures: recently Durbin *et al.* identified 1577 proteomes for cellular extracts of <30 kDa.<sup>129</sup>

### 1.2.3 Quantitative proteomic techniques

Various approaches can be employed to quantify proteins. Quantification can be relative between different samples, or it can provide an absolute measure of protein concentration. Relative quantification, which enables the comparison of proteomes across different biological conditions and disease states, can be achieved through the introduction of stable isotopes into samples. Isotopes exhibit comparable physio-chemical properties whilst differing in mass, thus making them distinguishable by MS techniques.

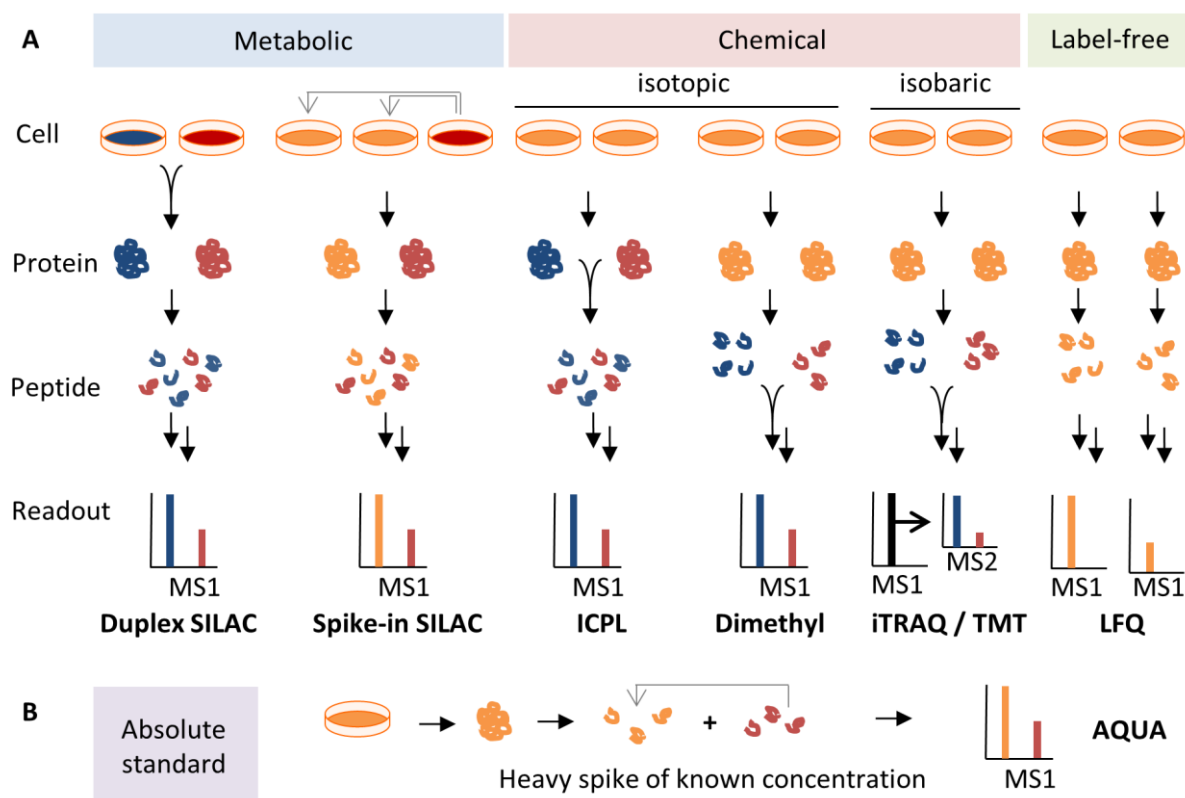
Stable isotopes can be introduced by metabolic or chemical means. SILAC (stable isotope labelling by amino acids in cell culture) is a metabolic approach that exploits the reliance of human cells on two essential amino acids: Arg and Lys.<sup>130</sup> Culturing cells in 'heavy' media, which contain Arg and Lys amino acids labelled with <sup>15</sup>N and <sup>13</sup>C isotopes (<sup>15</sup>N<sub>2</sub><sup>13</sup>C<sub>6</sub>: Lys8, K8 and <sup>15</sup>N<sub>4</sub><sup>13</sup>C<sub>6</sub>: Arg10, R10 ), leads to the stable incorporation of these 'heavy' amino acids into the cellular proteome. Heavy labelled proteomes can be quantified relative to 'light' proteomes incorporating natural amino acids in a classical Duplex SILAC experiment (Figure 9A). Further development of 'medium' isotopically labelled amino acids (<sup>2</sup>H<sub>4</sub>: Lys4, K4 and <sup>13</sup>C<sub>6</sub>: Arg6, R6) has enabled three-way Triplex SILAC comparisons. Alternatively, SILAC methodologies can be used to study *de novo* protein production and protein turnover dynamics in so called 'pulsed-SILAC' experiments.<sup>131</sup> More recently a spike-in SILAC methodology was established (Figure 9A).<sup>132</sup> In spike-in SILAC, conditions to be compared are cultured in the absence of isotopes. Quantification is achieved by introducing a fixed quantity a heavy proteome into each light condition, which acts as an internal standard for relative quantification.

Metabolic labelling introduces isotopic labels early on, thus minimising the impact of any experimental error introduced during sample preparation. However, SILAC reagents are relatively expensive and are not globally applicable; due to the necessity of heavy amino acid incorporation, SILAC is unsuitable for studying human tissue or bodily fluids. Chemical incorporation of isotopic tags offers an alternative approach. Isotopic tags can be incorporated at the protein or peptide level. The isotope-coded protein label (ICPL) method incorporates tags onto all free Lys residues and N-termini, prior to digestion (Figure 9A).<sup>133</sup> Alternatively, tags can be incorporated onto free amines following digestion using a low cost dimethyl labelling strategy (Figure 9A).<sup>134</sup> Both of methods are limited to triplex comparisons. In addition, as isotopic labels are introduced at the end of significant sample handling, error during quantification is a concern.

Both the metabolic and chemical labelling approaches introduced above are quantified during MS1 detection. These labels, therefore, increase sample complexity at the MS1 level, as each isotopically labelled peptide represents an additional analyte. This in turn reduces the likelihood of observing low abundant peptides and PTMs. Methods whereby quantification occurs during MS2, rather than MS1 circumvent these complexity issues. Quantification during MS2 is achieved through the use of isobaric tags. Such tags are of equivalent mass, and co-elute during LC separation. On fragmentation during MS2, however, different isobaric tags reveal characteristic ions of unique mass. This enables the tags to be differentiated and relatively quantified. iTRAQ and TMT are commercially available reagents that allow up to 10-plex sample comparison (Figure 9A).<sup>135, 136</sup> Though offering clear advantages by enabling high multiplexing whilst decreasing MS1 spectral complexity, such reagents still suffer from the errors associated with late stage labelling approaches, and are very expensive. Neutron encoding (NeCode) is a new technique that offers the advantages of both metabolic incorporation and isobaric tagging.<sup>137</sup> NeCode relies on small mass differences (6 mDa) which can be generated between lysine isotopes, enabling up to 18-plex sample comparison. Such small mass defects make these variants nominally isobaric, and resolvable using high resolution MS techniques.

Proteomes can be relatively quantified in the absence of isotopic labels using label-free quantification (LFQ, Figure 9A).<sup>138</sup> LFQ algorithms quantify peptides by integrating their MS spectral signal, and then comparing each peptide signal between different samples.<sup>139</sup> As LFQ requires no labelling it provides a cheap quantification method with a limitless multiplexing range. It also can provide a higher dynamic range of quantification, which is useful for the quantification of proteins which vary dramatically between samples. However, sample variation can result in significant quantification error. Consistent sample handling can be achieved by adopting automated workflows and minimising sample processing.

Absolute quantification of proteins (AQUA) can be achieved by introducing known concentrations of heavy protein or peptide which act as a standard (Figure 9B).<sup>140</sup> AQUA methodologies are applicable in targeted proteomic studies, such as selected reaction monitoring (SRM) MS. In SRM experiments, peptides of interest are preselected. Subsequent LC-MS/MS data acquisition focusses on the detection of these preselected peptides, thus, the addition of isotopically-labelled analogues provides standards for absolute quantification.



**Figure 9: Overview of (A) relative and (B) absolute quantitative proteomic techniques. Isotopic labels can be introduced metabolically or chemically, whilst chemically introduced isobaric tags allow multiplex quantification in MS2. Different methodologies introduce quantification at different workflow stages, and thereby are prone to different extents of sampling error. Label-free quantification compares MS peak intensity in the absence of a heavy standard. Introducing a heavy spike of known concentration for a protein of interest enables absolute quantification of that protein. Red and blue denote the introduction of labels for quantification, at the cell, protein, or peptide stage.**

## 1.2.4 Profiling Ub

### 1.2.4.1 MS analysis of Ub

For successful analysis by MS methods, a PTM needs to be both stable and detectable under MS conditions. At 8 kDa, Ub represents a large modification; however, it meets both these criteria. Furthermore, following tryptic digest Ub modifications leave a characteristic scar at modification sites which can be exploited to identify Ub modification sites. Ub-modified Lys residues cannot be efficiently digested, generating a missed cleavage site on modified peptides. As the C-terminal sequence of Ub is Arg-Gly-Gly (RGG), tryptic digest leads to a Gly-Gly (GG) scar at sites of Lys modification representing a 114.1 Da mass shift. Thus, a secondary database search for this modification on Lys residues enables the identification of both modified proteins and their modification sites.

Furthermore, as each of the seven Ub Lys residues reside on unique, prototypic, peptides different Ub chain isoforms can be distinguished. In 2003 Peng *et al.* demonstrated that all chain isoforms co-

exist in yeast using this technique, and were able to measure their relative abundance.<sup>141</sup> Since this study, the development of AQUA Ub standard peptides has enabled the absolute quantification of Ub chain linkage.<sup>142, 143</sup> Further studies implementing AQUA have compared Ub modifications throughout the cell cycle, and in various disease states. For example K11 chain abundance increases sharply on exiting mitosis, whilst K48 abundance remains constant.<sup>22, 144</sup>

The tryptic signature peptide of SUMO is much longer (19 or 32 amino acids for SUMO1 and SUMO2/3 respectively) making SUMO modifications much harder to detect. The development of SUMO mutants which mimic the RGG C-terminal of Ub, together with the correct controls, have enabled the detection of SUMO sites.<sup>145</sup> An alternative mutant possessing a Lys-Gly-Gly (KGG) C-terminus has also been used in combination with Lys-C digestion to identify SUMO sites.<sup>146</sup>

As these MS-based techniques are limited by their sensitivity and dynamic range, various enrichment strategies are implemented to overcome these issues when studying proteins and PTMs of interest, including Ub. These approaches are discussed below.

#### 1.2.4.2 Affinity enrichment approaches

Genetically incorporated epitope tags provide a useful handle for the affinity enrichment of specific proteins. N-terminal incorporation of His (poly-histidine), HA (Human influenza hemagglutinin), and biotin tags have been used to identify Ub- and Ubl- conjugated substrates. His-tagged Ub can be enriched using a Ni-based resin. His-tags have proved a successful handle for identifying Ub substrates in yeast, and were used in the study by Peng *et al.* mentioned earlier in Section 1.2.4.1.<sup>141</sup> His-tag enrichment suffers from high background in mammalian cells, however, due to the higher numbers of native His rich proteins. HA offers an alternative tag which can be enriched using a HA-specific antibody. Alternatively, biotin forms a strong, non-covalent interaction with avidin, which can be conjugated onto resin and used to enrich biotin-tagged proteins. Biotin can be incorporated chemically or enzymatically by a bacterial ligase: BirA (Bifunctional ligase/repressor BirA). The latter recognises a short N-terminal fused peptide sequence, known as the avi-tag, leading to the covalent attachment of a biotin motif.<sup>147</sup> This approach has enabled the biotinylation of Ub *in vivo* and its subsequent incorporation into native polyUb chains.<sup>148</sup> All these enrichment strategies suffer from background resulting from non-specific resin binding of non-tagged proteins. Therefore, potential substrates require further validation, commonly by immunoprecipitation or western blot with substrate specific antibodies.

UBDs have been used as an enrichment strategy for PolyUb chains. UBD pairs can bind cooperatively to polyUb in series, providing affinity interactions in the region of 1-15  $\mu\text{M}$ .<sup>55</sup> This has led to the development and application of tandem repeat ubiquitin binding entities (TUBEs) by Hjerpe *et al.*<sup>149</sup> As well as enriching polyUb chains, using TUBEs has the added bonus that inherent chain sequestration by TUBEs protects substrates from proteasomal degradation enabling their isolation and further study.<sup>150</sup> Through combining different UBDs, TUBEs can be modulated to enrich for specific chain isoforms of interest.<sup>151</sup> TUBEs have also been used to study the substrates of specific E3 ligases by overexpressing a selected E3 ligase with an epitope-tagged, trypsin-resistant TUBE.<sup>152</sup> In an alternative approach, UBDs have been fused to the E3 ligase of interest to form ligase-traps which have enabled E3 ligase substrate studies.<sup>153</sup> In general, the specificity of TUBEs and UBD enrichment strategies remains a concern, however, due to their nonspecific interaction with Ubl sequence domains in other proteins.<sup>154-156</sup>

#### 1.2.4.3 Ub Antibody enrichment

Ub antibodies have been available since the 1980s, however, although possessing affinity for Ub they suffered from cross reactivity.<sup>157</sup> Since, the development of monoclonal antibodies with specificity for polyub has enabled polyUb sample enrichment.<sup>158</sup> In 2008 specific antibodies for K48 and K63 chains isoforms were reported which have proven useful for western blot applications.<sup>159</sup> However, their application as an enrichment strategy has been unfruitful to date.

Enrichment strategies at the peptide, rather than protein, level have been developed which exploit the K-GG signature peptide formed from the GG scar on Lys residues following tryptic digest of Ub modification sites. In 2010 the first K- $\epsilon$ -GG antibody was reported.<sup>160</sup> This was quickly followed by three landmark papers which demonstrated that enrichment with a K- $\epsilon$ -GG antibody enabled the identification and SILAC relative quantification of >10 000 sites in the largest set of Ub substrates identified to date, all without perturbing the endogenous system.<sup>161-163</sup> The downside of using the GG signature peptide to detect Ub modifications is that this motif is shared by Nedd8 and ISG15 Ubls, making them indistinguishable by this technique alone.<sup>128</sup> In their study, Kim *et al.* used positive controls to estimate the relative contribution of each modification. ISG15 contribution was estimated by  $\beta$ -interferon stimulation, whilst Nedd8 contribution was estimated by cleaving all Ub modification with generic DUBs, so that only Nedd8 modifications remained. In both cases the Ubls were estimated to make a minor impact on the overall K- $\epsilon$ -GG enrichment landscape.<sup>161</sup>



#### *1.2.4.4 Localisation studies*

Genetic incorporation of fused reporter proteins, such as GFP (green fluorescent protein) is common within the proteomic field. Being fluorescent, GFP fused proteins can be applied in both direct imaging to determine cellular localisation, and in co-localisation studies. Due to their size, however, fused proteins present a large perturbation from the endogenous system, which may alter both protein function and localisation.

As PTMs are not genetically encoded, fusion proteins are often not applicable; however, as Ub is a protein, is it possible to design Ub-GFP substrates. These have been used to study proteasomal degradation in live human cells and further incorporated into transgenic mouse models.<sup>164, 165</sup> In addition, GFP-fused UBDs and multi-UBDs have been reported as an alternative approach to 'sense' polyUb chains in live and fixed mammalian cells.<sup>166, 167</sup> More generally, specific antibodies can be fluorescently labelled for localisation and co-localisation studies for proteins of interest in fixed cells.

### **1.2.5 Profiling Ub machinery**

#### *1.2.5.1 Genetic methods*

Early investigation of Ub and its associated machinery was undertaken in both mammalian and yeast models. Approximately 50 % homology with human genes associated with disease is observed, which combined with its ease of manipulation makes yeast a good biological model.<sup>168</sup> Gene deletions, or 'knockouts', can be used to infer the biological function of the encoded enzymes. A recent study used yeast deletion mutants to quantify the effect of nine specific DUBs on the yeast proteome, using a multiplex TMT labelling strategy.<sup>169</sup>

In mammalian models, gene knockouts and transient RNA interference (RNAi) techniques have been used extensively in enzymes studies, including that of Ub machinery. The CRISPR/Cas-9 (Clustered regularly-interspaced short palindromic repeats / Caspase-9) system is increasingly regarded as a superior method of gene knockout, providing a quicker and more robust approach.<sup>170</sup> However, all these genetic approaches suffer from redundancy in pathways, whereby if several genes perform the same function, disruption of one may result in minimal phenotypic effect. This gene redundancy can make the biological roles of certain genes and their encoded proteins difficult to pinpoint.

Genetic mutations, both synthetic and computational, are another important tool for studying enzyme activity. Alanine scanning, whereby amino acids are symmetrically replaced with alanine, has been used to identify key interacting residues in both Ub and its machinery.<sup>171, 172</sup> Furthermore,

mutations to form more stable analogues can aid structural studies of transient enzyme complexes by crystallography methods, as mentioned in Section 1.1.3.2

#### 1.2.5.2 Activity-Based Probes (ABPs)

Activity-Based Probes (ABPs) use chemical entities to capture and investigate a subset of enzymes in an activity-dependent manner.<sup>173</sup> ABPs bind covalently to the active site of enzymes forming probe-enzyme complexes, and can range from small inhibitor-like molecules to protein derivatives. ABPs offer a range of applications, including quantitative proteomic analysis and localisation studies in the context of PTMs. Classical ABPs possess an electrophilic trap to enable the covalent trapping to nucleophilic residues and a tag, such as a biotin moiety, which enables affinity enrichment of probe-enzyme complexes (Figure 10A). Enriched proteins can then be separated and analysed by MS/MS or western blot. The early development of such ABPs was pioneered by the Cravatt and Bogoy lab, and has since been used to profile a variety of enzyme classes.<sup>174, 175</sup>

ABPs derived from Ub have proven applicable for the activity-based protein profiling (ABPP) of DUBs. As Ub is transferred between Ub machinery *via* catalytic cysteine residues, incorporation of a cysteine reactive electrophilic trap at the C-terminal of Ub enables covalent trapping of these enzymes. Various electrophiles show reactivity towards cysteine, with varying degrees of selectivity over other amino acid residues observed.<sup>176</sup> In 2001 Borodovsky *et al.* reported that radiolabelled Ub incorporating a C-terminal vinyl sulphone (<sup>125</sup>I-Ub-VS) could be used as a DUB probe in mammalian cell lysate.<sup>177</sup> The following year the same authors reported the development of a series of HA-tagged Ub probes (Figure 10B), which were synthesised using an intein-based chemical ligation method.<sup>178</sup> Various electrophiles were incorporated and the resulting probes profiled. Subsequent bottom-up mass spectrometry analysis led to the identification of several DUBs by these probes. Since, Ub ABPs have enjoyed widespread use in DUB studies, including quantitative SILAC comparison of various disease states and selectivity profiling of DUB inhibitors.<sup>179, 180</sup>

Further development of DUB reactive Ub ABPs has followed, including the synthesis of di-Ub probes which mimic various Ub chain isoform linkages, as well as the addition of short peptide sequences to form Ub-substrate mimics.<sup>181, 182</sup> Identification of other ubiquitination machinery has proven more difficult by this approach. The incorporation of highly electrophilic traps into Ub ABPs has been reported as a method of identifying E3 ligases.<sup>183</sup> However, attempts to synthesise such ABPs has proven unsuccessful in our hands.<sup>2</sup> Recently, a Ub-E2 conjugate probe incorporating an electrophilic trap was used to study the RBR E3 ubiquitin-protein ligase parkin (Parkin).<sup>184</sup>

The synthesis of Ub ABPs is hindered by the low yields afforded by intein-based methods.<sup>178</sup> The optimisation of solid phase peptide synthesis (SPPS) procedures has enabled total synthesis of Ub and Ub ABPs in substantially improved yield.<sup>185</sup> This synthetic approach also allows increased flexibility in the choice of N-terminal modifications. Ub ABPs synthesised in this manner are now commercially available.

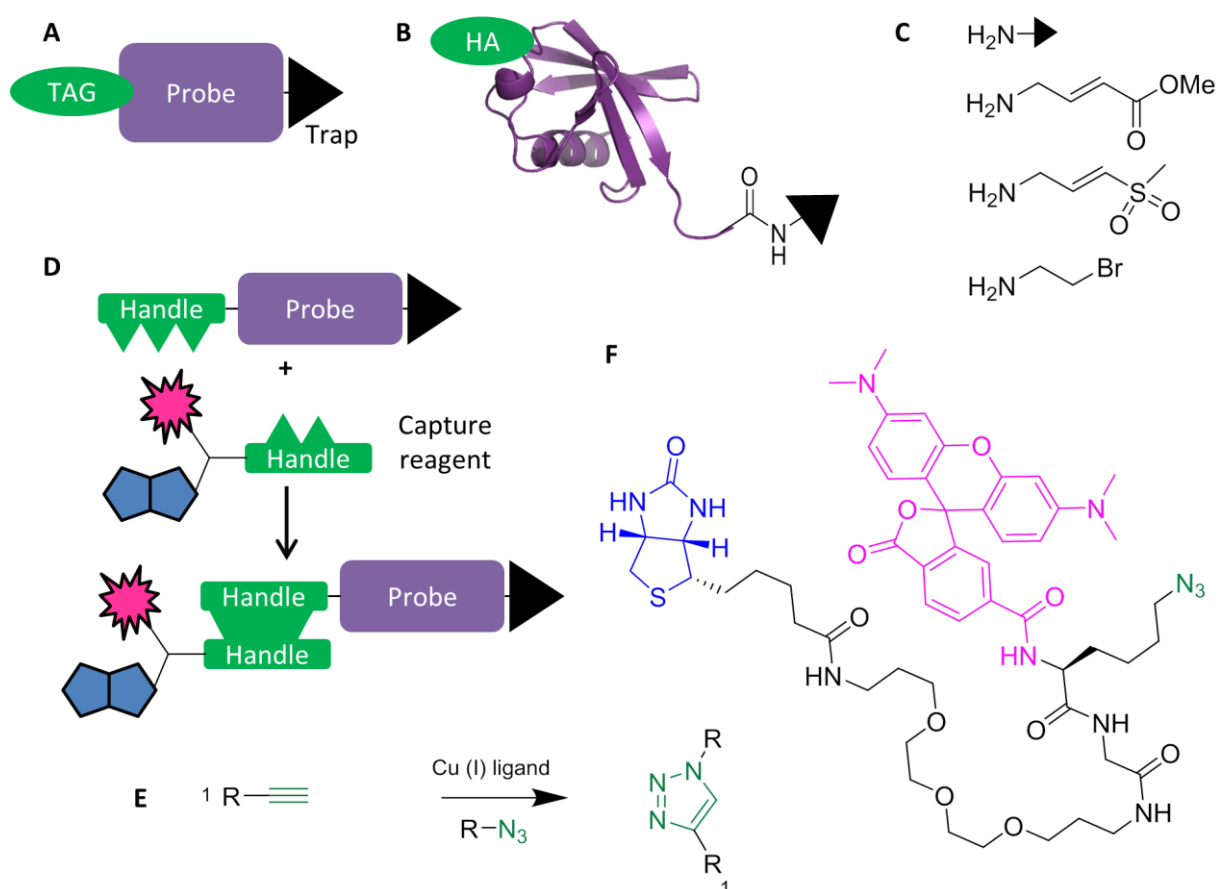
In addition to one-step enzyme capture, the development of bioorthogonal reactions has permitted the use of two-step ABPP strategies (Figure 10D).<sup>186</sup> Such reactions enable the addition of various bulky reporters, such as fluorophores and affinity tags, after the formation of probe-enzyme complexes. Bioorthogonal reactions require incorporation of a small chemical handle onto the ABP that can react with a complementary handle on reporter groups. These handles present a much smaller perturbation than an epitope tag or fusion protein, thereby preventing disruption of native biological processes under investigation.

Bioorthogonal reactions by definition are chemical reactions that can occur without interfering with native biological processes. A bioorthogonal reaction must, therefore, be both selective and biocompatible. Accordingly, reaction components must be stable under physiological conditions, and the reaction performable at physiological temperature and pH. If the reaction is performed in live cells, toxicity from reagents and by-products is a consideration and needs to be avoided. Copper catalysed azide-alkyne cycloaddition (CuAAC) commonly referred to as 'click chemistry' has been used with great success as a bioorthogonal reaction in proteomic applications.<sup>187, 188</sup> In CuAAC, alkyne and azide moieties provide complementary handles, which covalently react through a Cu (I) catalysed cycloaddition reaction (Figure 10E). CuAAC is a fast and highly sensitive reaction, though the use of Cu (I) does result in undesired toxicity. However, due to cell permeability bioorthogonal ligation of reporter groups is often undertaken after cell lysis, reducing toxicity concerns.

Multiple reporters can be ligated onto a probe complex simultaneously using designed 'capture reagents'. These reagents can possess affinity handles, fluorophores, and in some cases isotopic labels for quantification purposes. AzTb is a capture reagent developed by the Tate group combining TAMRA, biotin and azide moieties, to enable labelling of alkyne tagged probes (Figure 10F).<sup>189</sup> Capture reagents that include cleavable linkers have also been designed.<sup>190</sup> Such linkers can be used to improve purification strategies, as well as to validate the site of ABP binding.

Ub ABPs incorporating an alkyne handle have been reported.<sup>191</sup> Such probes have been successfully synthesised in house and further demonstrated to label a number of proteins.<sup>2</sup> A small molecule

inhibitor derived ABP, possessing an alkyne tag, has also been reported that labels Ub and Ubl E1 enzymes through a two-step labelling approach.<sup>192</sup> In a more global approach, ABPP of reactive cysteines has been performed by the Cravatt lab using an iodoacetamide (IA) derived alkyne tagged ABP, IA-Alkyne.<sup>193, 194</sup> IA is a pan cysteine alkylating reagent and IA-Alkyne does identify some ubiquitination enzymes, however, both are toxic making the cellular application of IA-Alkyne problematic. A 'caged' bromoacetamide probe analogue has recently been reported that overcomes this toxicity, enabling the identification of >300 reactive cysteines in cells.<sup>195</sup> Coverage of Ub machinery however still remains low.



**Figure 10: ABPP approaches. (A) Classical ABPs possess an affinity tag for enrichment, and an electrophilic trap for covalent binding. (B) HA-tagged Ub ABPs have been used extensively to study DUBs. (C) Various traps can be incorporated into Ub ABPs, which demonstrate varying selectivity. (D) Two step ABPs possess a bioorthogonal handle to enable the addition of bulky capture reagents after probe-enzyme complex formation. (E) CuAAC is a bioorthogonal ligation reaction between an alkyne and an azide bond, which is suitable for in lysate applications. (F) Structure of the capture reagent AzTb.**

Affinity-based probes (AfBPs), bind in an affinity, rather than an activity, based manner and can also be employed in an analogous manner.<sup>196</sup> Such probes possess a photo-crosslinking moiety, to enable covalent trapping of interacting enzymes and further enrichment. To date, AfBPs have not been used in the proteomic profiling Ub machinery.

### 1.2.5.3 Emerging methods

Methods have been developed which enable the detection of enzyme-probe interactions without requiring any affinity handles or bioorthogonal tags. DARTS (drug affinity responsive target stability) and CETSA (cellular thermal shift assay) are two 'label-free' approaches that exploit the altered stability of proteins on binding to small molecules. DARTS technology distinguishes bound and unbound proteins by their altered proteolytic stability.<sup>197</sup> Alternatively, CETSA technology exploits their altered thermal stability properties.<sup>198</sup>

Fusion of *BirA* ligase onto proteins of interest is a method of identifying an enzyme's substrates *in vivo*. The technique, known as BioID, relies on substrate proteins being biotinylated when they are in close proximity to the *BirA* ligase.<sup>199</sup> This enables subsequent purification and identification of protein substrates. Recently, BioID has been applied to identify the substrates of a cullin-E3 ligase complex.<sup>200</sup>

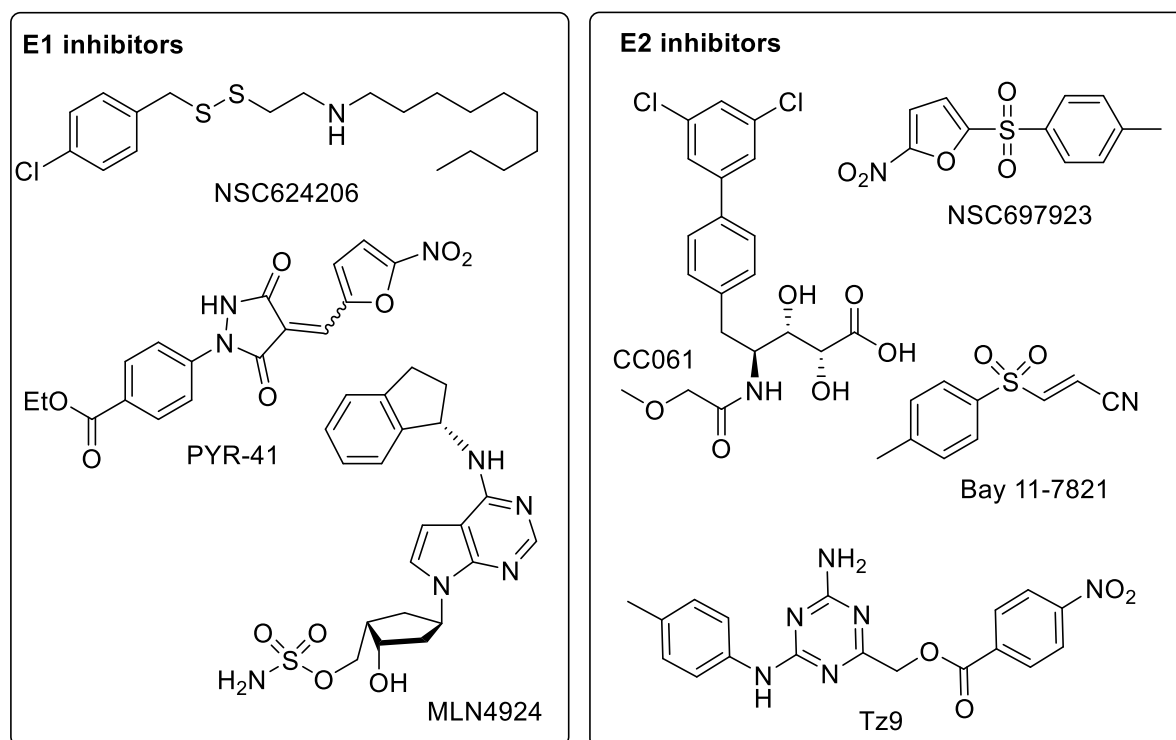
Proteolysis-targeting chimera (PROTAC) technology provides a chemical means of protein knockdown.<sup>201</sup> In 2010 an E3 ligase, Cereblon (CRBN), was identified as the primary target of thalidomide (Figure 12).<sup>202</sup> The ligation of thalidomide or its derivatives to specific inhibitors enables the targeted degradation of proteins of interest. Such technology has been reported as a means of degrading bromodomain-containing protein 4 (BRD4), leading to impeded proliferation and induced apoptosis in a cancer cell model.<sup>203, 204</sup> This approach has been applied to other proteins, through no Ub machinery to date.<sup>205</sup>

## 1.2.6 Inhibitors of Ub machinery

As well as their therapeutic potential, high quality inhibitors can assist our understanding of biological processes, and provide starting points for the design of tool molecules, such as ABPs. Various inhibitors of Ub machinery have been identified through small molecule screens both *in silico* and *in vitro*, as well as through rational structural guided design. A selection are included in Figure 11-13 and further discussed below.

PYR-41 and NSC624206 are two reported inhibitors of the Ub E1 enzyme, UBA1 (Figure 11).<sup>206, 207</sup> Both prevent E1-Ub thioester formation through suggested covalent mechanisms. MLN4924, a selective inhibitor of UBA3 (ubiquitin like modifier activating enzyme 3), the Nedd8 E1 enzyme, has shown promise as an anti-cancer therapeutic and is currently in phase I clinical trials.<sup>208</sup> CC0651 is an inhibitor of UB2R1 (Ubiquitin-conjugating enzyme E2 R1) that binds to an allosteric site, stabilising the E2-Ub conjugate.<sup>209, 210</sup> Active site binding E2 inhibitors have since been reported, including

NSC697923 and Tz9.<sup>211-213</sup> Bay11-7821 is an anti-inflammatory drug, which is now known to form a covalent adduct with NF- $\kappa$ B pathway associated E2 enzymes.<sup>214</sup>



**Figure 11: Structures of reported E1 and E2 inhibitors.**

Due to their large number and imparted substrate specificity, much effort has gone into the development of E3 ligase inhibitors. Several E3 ligase inhibitors target PPIs between E3s and their substrates or other enzyme partners (Figure 12). Particular attention has been paid to the interaction between the E3 ligase MDM2, and its substrate p53. The Nutlin derivative Nutlin3, and MI-219 are two MDM2-p53 PPI inhibitors in phase I clinical trials.<sup>215, 216</sup> Serdemetan (also in phase I clinical trials), SP141 and NSC66811 are known MDM2 binders that inhibit the MDM2-p53 interaction, whilst RITA inhibits the interaction by binding to p53.<sup>217-220</sup>

Other inhibitors have been developed that target various cullin E3-ligases, such as GS143 and SZL P1-41, and as mentioned in Section 1.2.5.3 thalidomide is a CRBN inhibitor.<sup>202, 221-223</sup> SM-406 and GDC-0152 are both inhibitors of cullin E3-ligases that act as IAPs (Inhibitor of Apoptosis Proteins) and are currently in phase I clinical trials due to their potential as selective cancer therapeutics.<sup>224, 225</sup> An inhibitor of the APC/C complex, TAME, has also been reported.<sup>226</sup>

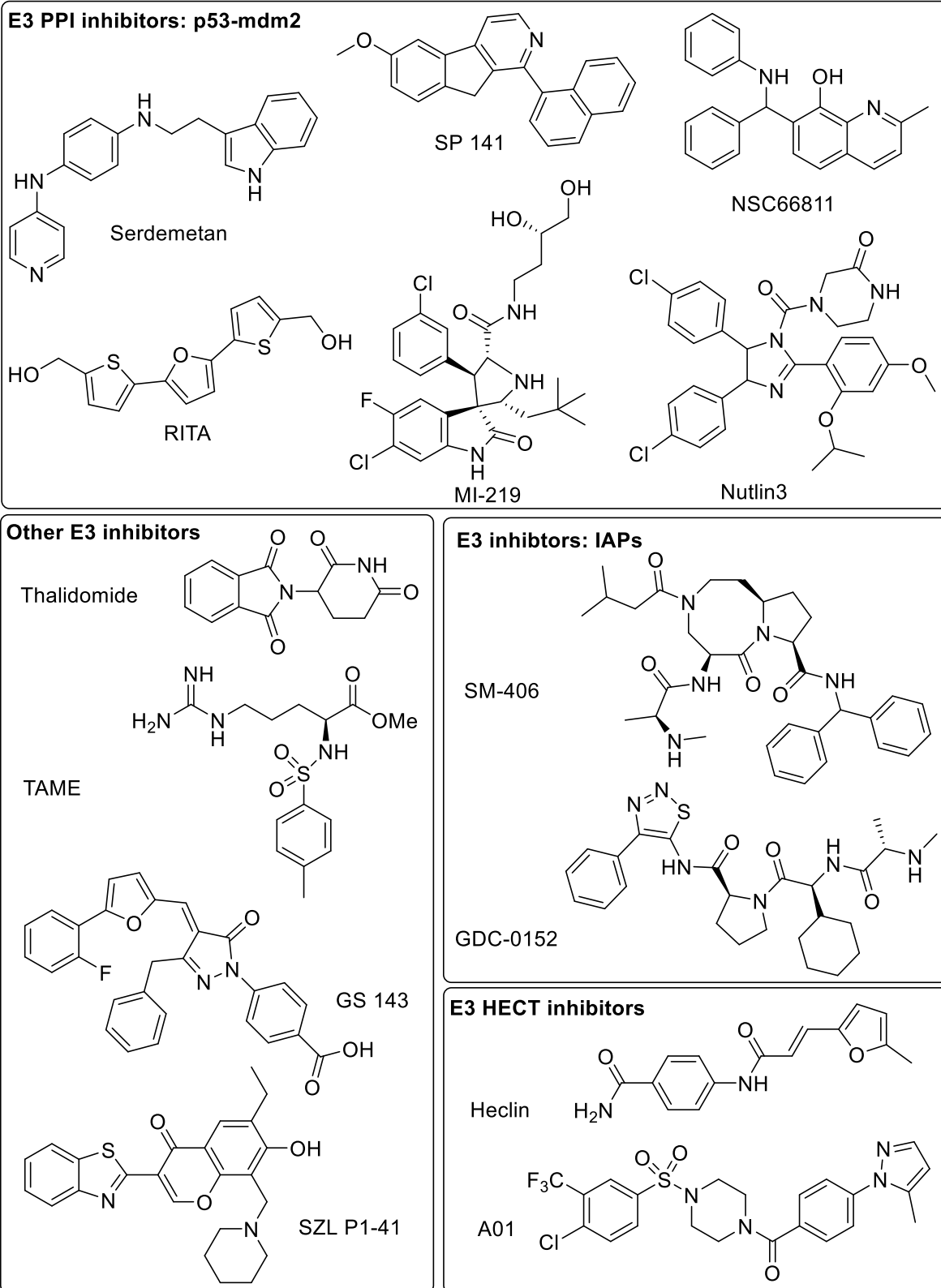
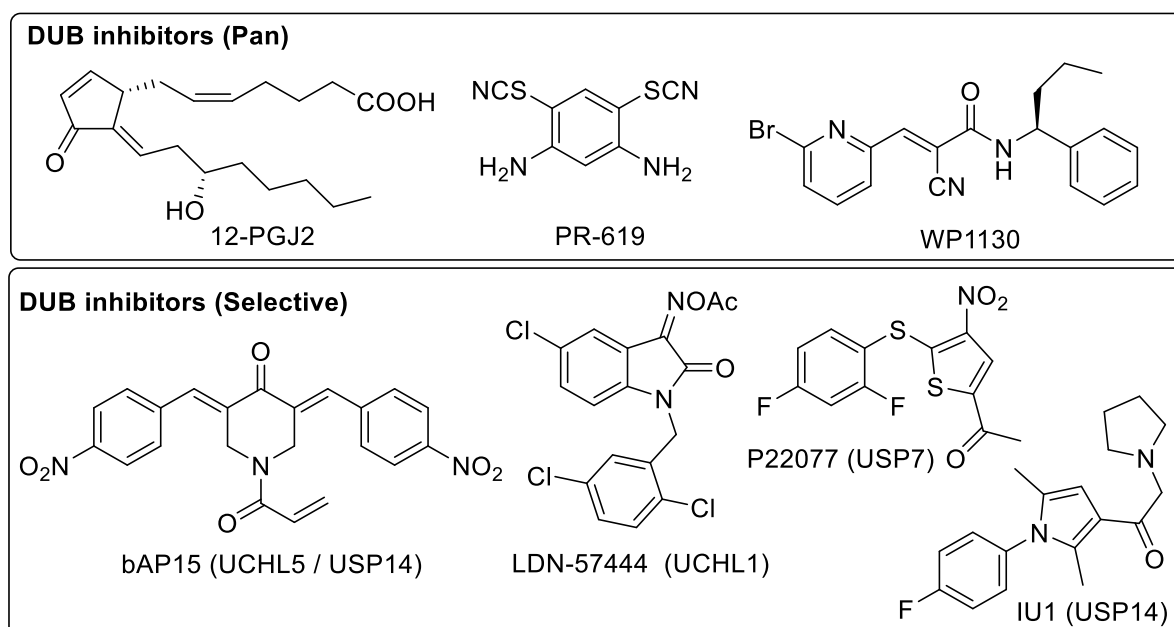


Figure 12: Structures of selected E3 ligase inhibitors.

In 2014 Mund *et al.* reported Heclin, a small inhibitor which shows broad inhibition of HECT E3 ligases in cells, demonstrating their druggability.<sup>227</sup> A01 has also been further reported as a selective E3 ubiquitin-protein ligase SMURF1 (SMURF1) inhibitor.<sup>228</sup>

DUBs, which like E3s impart substrate specificity, have also enjoyed significant attention as drug targets. Several literature DUB inhibitors are known, varying from pan to specific inhibitors (Figure 13). For example, PR-619 and WP1130 are broad inhibitors of USP and UCH family DUBs.<sup>229, 230</sup> P22077 and IU1 are reported as specific inhibitors of USP7 and USP14 respectively.<sup>179, 231</sup> Some inhibitors of cysteine protease DUBs are covalent binders, utilising a Michael acceptor to trap catalytic cysteines. 12PgJ2, WP1130 and bAP-15 fall into this category.<sup>232, 233</sup>



**Figure 13: Structures of selected DUB inhibitors. Enzyme selectivity is indicated in brackets.**

For further details of ubiquitination enzyme inhibitors, please refer to the following comprehensive reviews.<sup>108, 234, 235</sup>



### 1.3 Project objectives

The aim of this PhD project is to develop chemical tools to assist the study of ubiquitination machinery. As mentioned in Section 1.2.5.2, Ub-based ABPs have been used extensively in the literature to study the activity of DUBs. However, the study of other ubiquitination machinery has proven non-facile using these tools. Furthermore, such tools are restricted to in lysate applications.

Several Ub enzyme inhibitors, introduced in Section 1.2.6, have been reported. A number of these compounds provide starting points for novel, cell permeable, ABP design. Once characterised, these ubiquitination probes could be applied to biological models of interest, such as DDR. Ultimately, DDR specific ubiquitination activity could be identified, the inhibition of which could induce synthetic lethality, or in combination could improve the efficacy of current chemotherapeutic strategies (Section 1.1.6.2).

Alternatively, novel starting points for tool design could be identified using high throughput screening methods against specific families of ubiquitination machinery such as RBR E3 ligases. These enzymes form transient complexes with their partner E2 enzymes (Section 1.1.3.1) and their substrates. The development of structural models to study these complexes would further assist their study and evaluation for further drug discovery.

The main objectives of this PhD thesis were to:

- Design and synthesise novel ubiquitination ABPs, derived from peptide and small molecule inhibitors (Chapter 2 and Chapter 3).
- Identify the targets of these ABPs using a chemical proteomics approach, and to evaluate their suitability as ubiquitination probes (Chapter 2 and Chapter 3).
- Apply ubiquitination ABPs in a DDR model (Chapter 2).
- Explore novel starting points for the development of tools to study RBR E3 ligases, using a fragment based drug discovery approach (Chapter 4).
- Develop a model system to study transient E3 ligase complexes (Chapter 5).

## **Chapter 2 Peptide activity-based probes for studying ubiquitination machinery**

---

Chemical proteomics is a powerful tool for elucidating complex biological interactions. This Chapter reports the design, synthesis, and development of 'clickable' peptide derived activity-based probes and their subsequent application in quantitative SILAC experiments in lysate. Their suitability as cell-based probes is further explored by utilising fluorescently labelled analogues and through the incorporation of cell penetrating peptide sequences.

## 2.1 Introduction

As previously mentioned in Chapter 1, Ub-derived ABPs are well established tools for profiling DUB activity in cell lysate. However, such ABPs are not suitable for all types of ubiquitination machinery. Whilst Ub-derived ABPs label more than 60 % of DUBs<sup>†</sup>, they have shown limited capacity to label other ubiquitination machinery. Furthermore, probe application is restricted by ineffective cell permeability. Though application in semi-intact cells has been reported, in the majority of cases the ABP is applied post cell lysis.<sup>236</sup> This presents difficulties if targets of interest are transient or their activity is disrupted during lysis. Therefore, the development of cell permeable probes capable of effective labelling of Ub machinery would significantly enhance the study of these enzymes.

In 2005, Borodovsky *et al.* reported that C-terminal Ub-derived peptide inhibitors possessing an electrophilic vinyl sulphone (VS) trap can inhibit several DUBs *in vitro*.<sup>237</sup> These peptides were evaluated by autoradiography through competition against [<sup>125</sup>I]-Ub-VS (introduced in Section 1.2.5.2), with DUB targets subsequently identified by LC-MS/MS analysis of excised bands using a Ha-Ub-VS ABP (Chapter 1, Figure 10). This target analysis is not comprehensive, however, as it only identifies targets also labelled by Ha-Ub-VS. We hypothesise that such peptide-derived ABPs, far removed from the tertiary structure of whole Ub, will have the ability to label beyond the DUBs reported in this study, and potentially offer a probe for Ub machinery detection. Consequently, ABP analogues of these peptide inhibitors were designed and studied to evaluate their ability to label Ub machinery of interest.

## 2.2 Design and synthesis

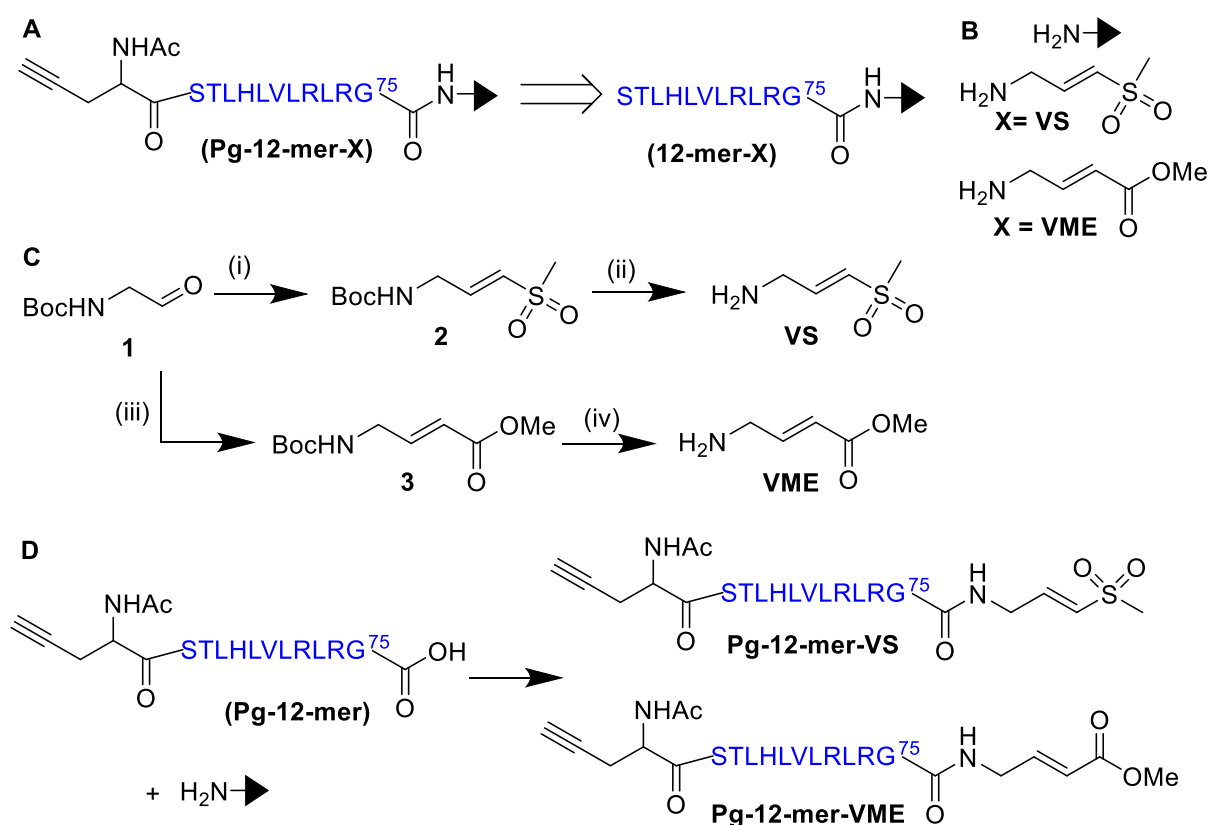
For our preliminary peptide-based probe studies, a peptide consisting of the last 12 C-terminal residues of Ub (**12-mer**) was chosen, as this was the minimum peptide length reported to achieve effective competition.<sup>237</sup> An ABP probe analogue (**Pg-12-mer**) was designed by incorporation of the unnatural amino acid propargyl glycine (Pg) at the N terminus (Figure 14A). Addition of a C-terminal electrophilic trap completed both inhibitor and probe design.

Two electrophilic traps; **VS** and a vinyl methyl ester (**VME**) were tested to scope the effect of electrophile type on target identification (Figure 14B). Boc-protected **VS** (**2**) was successfully synthesised from *N*-Boc-2-aminoacetaldehyde (**1**) and diethyl [(methylsulfonyl)methyl] phosphonate

---

<sup>†</sup> Calculated from published data.<sup>179</sup>

using NaH in THF *via* a Horner-Wadsworth-Emmons type reaction, in 48 % yield (Figure 14C).<sup>238</sup> Protected **VME** (**3**) was synthesised in an analogous fashion from **1** and methyl diethylphosphonoacetate, in 42 % yield. Although initially employed with success, the use of NaH proved unreproducible on repeat synthesis of these compounds. Consequently, K<sub>2</sub>CO<sub>3</sub> was tested as an alternative base, providing **2** and **3** in an improved yield of 66 % and 53 % respectively.<sup>182</sup> Subsequent Boc deprotection with 33 % TFA/DCM afforded **VS** as a TFA salt, whilst overnight treatment with anhydrous *p*TsOH afforded **VME** as a tosylate salt. Both deprotection methods afforded quantitative yields.



**Figure 14:** (A) Retrosynthetic scheme of Pg-12-mer-X from 12-mer-X. (B) Structures of electrophilic traps employed. (C) VS and VME synthetic development: i) diethyl [(methylsulfonyl)methyl]phosphonate, K<sub>2</sub>CO<sub>3</sub>, THF, RT, 16 h, 66 %; ii) 33 % TFA / DCM, RT, 1 h, quant; iii) methyl diethyl phosphonoacetate, K<sub>2</sub>CO<sub>3</sub>, THF, RT, 16 h, 53 %; iv) *p*TsOH, Et<sub>2</sub>O, RT, 16 h, quant. (D) Amide coupling of protected Pg-12-mer to an electrophilic trap in the presence of PyBop and DIPEA, followed by TFA deprotection affords the complete probe structure.

Both **12-mer** and **Pg-12-mer** were successfully synthesized by automated SPPS, using acid cleavable 4-sulfamylbutyryl rink amide AM resin.<sup>237, 239</sup> All N-terminals were acetylated, to mimic an adjacent peptide bond. On first attempt, the yield was significantly reduced due to poor coupling of the Threonine (Thr<sup>66</sup>) residue leaving the peptide capped after Lys<sup>67</sup> as the major product. Nevertheless, C-terminal trap incorporation could be tested using this peptide mixture; the resin was activated by

overnight treatment with TMS-CHN<sub>2</sub><sup>239</sup> and the peptides displaced with concomitant installation of **VS** or **VME**. Subsequent TFA deprotection of the crude mixture, peptide precipitation from *t*BME, and LC-MS analysis confirmed successful formation of **Pg-12-mer-VS** and **Pg-12-mer-VME** probes. However, the crude mass return was extremely poor (~5 %). It was thought that this low mass yield was due to inefficient peptide cleavage from the resin; analysis of the resin post displacement confirmed that peptide still remained attached to the resin.

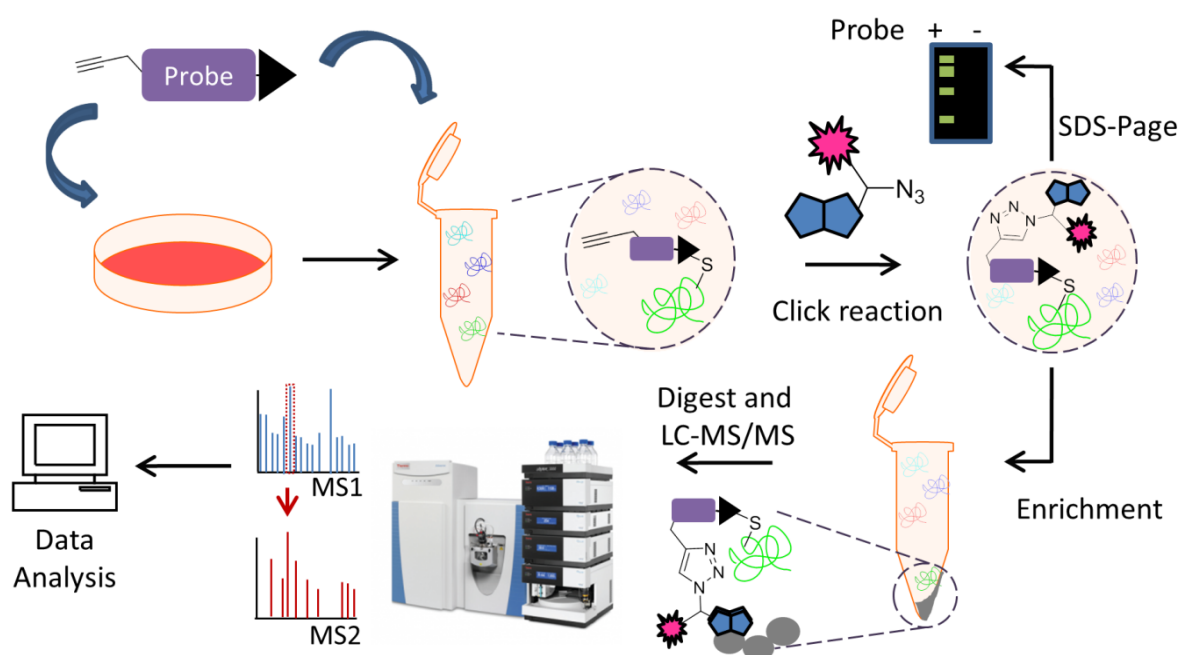
It was postulated that the low mass return was due to poor accessibility of the trap amine to the resin cleavage site. To improve the yield both trap incorporation and peptide coupling strategies were addressed; Thr<sup>66</sup> coupling was improved through the use of a Fmoc-Ser(*t*Bu)-Thr( $\Psi$ ME,Mepro)-OH pseudoproline building block, whilst trap coupling was improved by using a 2-chlorotrityl chloride resin. This resin allows mild peptide cleavage using HFIP whilst maintaining side-chain protecting groups, thus enabling in-solution coupling of the trap in the presence of PyBop and DIPEA, with subsequent TFA deprotection, *t*BME precipitation and LCMS purification, affording the desired product (Figure 14D).<sup>185</sup> This approach proved significantly more efficient, with minimal truncated by-products observed and all peptide cleaved from the resin. In this way **Pg-12-mer-VS** and **Pg-12-mer-VME** probes were successfully isolated in mg quantities representing 8 % and 16 % yield respectively. **12-mer-VS** and **12-mer-VME** were synthesised in an analogous manner in 2.8 % and 13 % yield.

### 2.3 In lysate applications and initial proteomic evaluation

With **Pg-12-mer-VS** and **Pg-12-mer-VME** in hand, work progressed to test their labelling ability in cell extract. HeLa cells were chosen for preliminary tests, as they had been previously used successfully in our hands to evaluate labelling with Ub-based probes.<sup>2</sup> The general proteomic workflow employed is summarised in Figure 15.

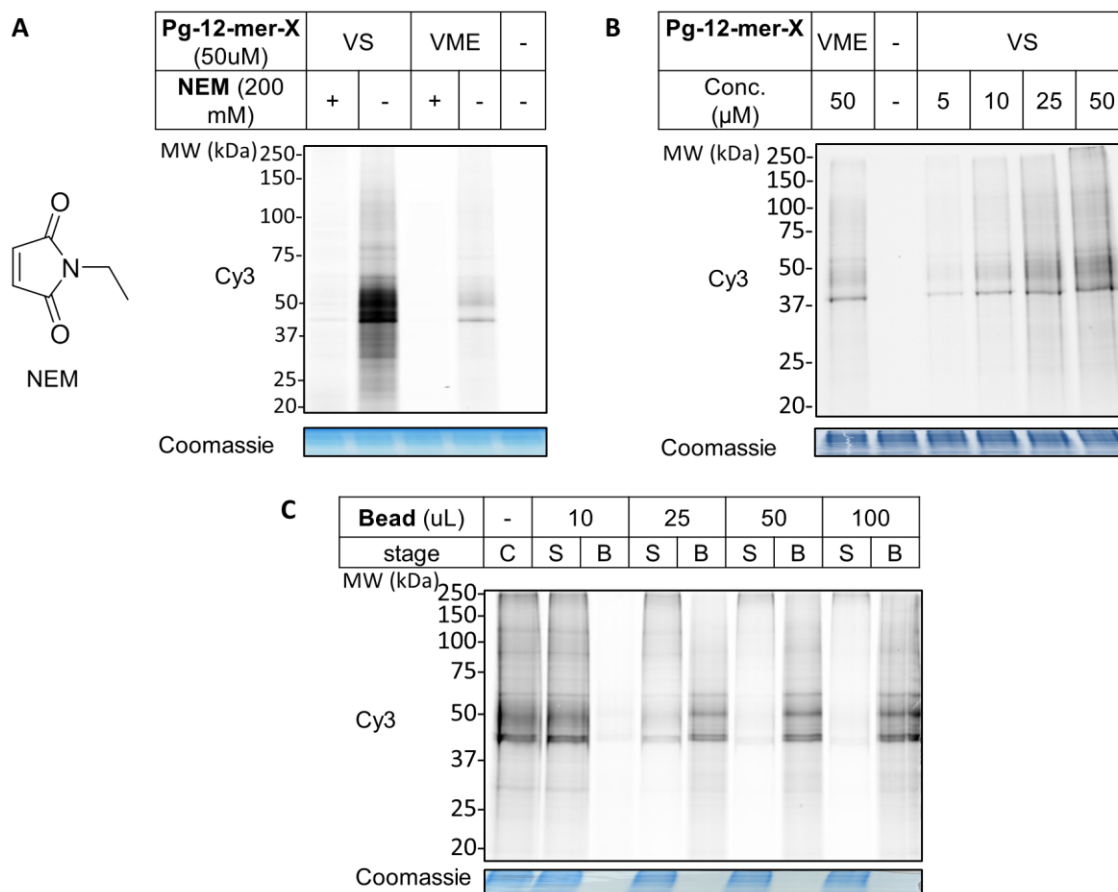
Native HeLa cell lysate, formed by shear lysis, was incubated with 50  $\mu$ M of probe for 1 hour at 37 °C. Pre-incubation with 10 mM *N*-ethyl maleimide (NEM), a cysteine alkylation reagent, for 20 minutes was used as a control for competitive labelling. The probe-enzyme complexes formed were subsequently ligated to AzTB, the capture reagent introduced in Chapter one, by CuAAC ligation chemistry. For primary analysis of labelling, samples were separated by SDS-PAGE and visualised by in-gel fluorescence (Figure 16A). Fluorescent labelling by both probes was observed above background, with labelling significantly stronger for **Pg-12-mer-VS** at this concentration. A depletion

of signal was observed in the NEM control lanes in both cases, suggesting that labelling is dependent on the availability of catalytic cysteines, and thus is activity dependent.



**Figure 15: Cartoon representation of the general proteomic workflow. Following probe incubation with cells or cell lysate, probe-enzyme complexes are ligated to AzTb. Captured complexes are then separated by SDS-PAGE and visualised by in-gel fluorescence, or affinity enriched using NeutrAvidin™ beads (Thermo Fisher), digested with trypsin, and characterised by LC-MS/MS.**

In order to compare the labelling attributes of each electrophilic trap, probe concentrations were adjusted so that the labelling intensity was normalised (Figure 16B). A 10  $\mu\text{M}$  treatment with **Pg-12-mer-VS** was chosen, as this was observed to provide a similar labelling intensity to that of **Pg-12-mer-VME** at 50  $\mu\text{M}$ . Prior to proteomic analysis, optimisation tests were conducted to ensure that a sufficient amount of NeutrAvidin beads was used to enrich the samples. Following ligation to AzTB, 100  $\mu\text{g}$  of probe-treated lysates were incubated with varying amounts of beads for 2 hours. The supernatant was then removed and the beads washed, before boiling the beads to release enriched peptides. Aliquots of the clicked peptide, supernatant and bead enriched lysate were separated by SDS-PAGE and visualised by in-gel fluorescence (Figure 16C). On analysis, 10  $\mu\text{L}$  of beads was found to insufficiently enrich the sample, whilst 25  $\mu\text{L}$  and greater afforded a noticeable depletion in supernatant fluorescence signal. 25  $\mu\text{L}$  was therefore chosen for proteomic applications, as this enabled sample enrichment, whilst minimising background signal arising from non-specific binding to the beads.



**Figure 16: Preliminary investigation of peptide probe labelling on HeLa cell lysate by in-gel fluorescence. (A)** Cell lysate were pre-treated with NEM or DMSO for 20 min before incubation with probe for 1 hr, followed by CuAAC ligation to AzTB, separation by SDS-PAGE, and visualisation by in-gel fluorescence in the Cy3 channel (excitation wavelength 552 nm, emission wavelength 570 nm). Molecular weight (MW) markers are indicated, and Coomassie staining provides protein loading controls. **(B)** Concentration series of Pg-12-mer-VS labelling. **(C)** Following AzTB ligation, Pg-12-mer-VS (10 μM) treated lysates (100 μg) were enriched on NeutrAvidin beads. Stage: C = following CuAAC, S = supernatant following enrichment, B = following release from bead.

With probe concentration and sample enrichment conditions optimised, preliminary proteomic analysis was undertaken for **Pg-12-mer-VS** and **Pg-12-mer-VME**. Briefly, native HeLa lysate was incubated with the aforementioned concentration of **Pg-12-mer-VS** or **Pg-12-mer-VME** for 1 hour at 37 °C. In the case of **Pg-12-mer-VS**, lysate formed by two preparative methods were prepared: shear lysis and chemical lysis, to compare the effect of lysis conditions on probe labelling. The probe-enzyme complexes formed were subsequently ligated to AzTB and enriched as previously described. Once on bead, samples were reduced and alkylated by dithiothreitol (DTT) and IA respectively; the removal of disulphide bridges from peptide mixtures simplifies subsequent peptide identification. Overnight trypsinisation afforded an enriched peptide mixture which, once desalted, was separated by LC-MS/MS using a Q-Exactive Orbitrap mass spectrometer. Control samples lacking probe, were prepared in parallel in order to identify any background signal due to unspecific interactions with the

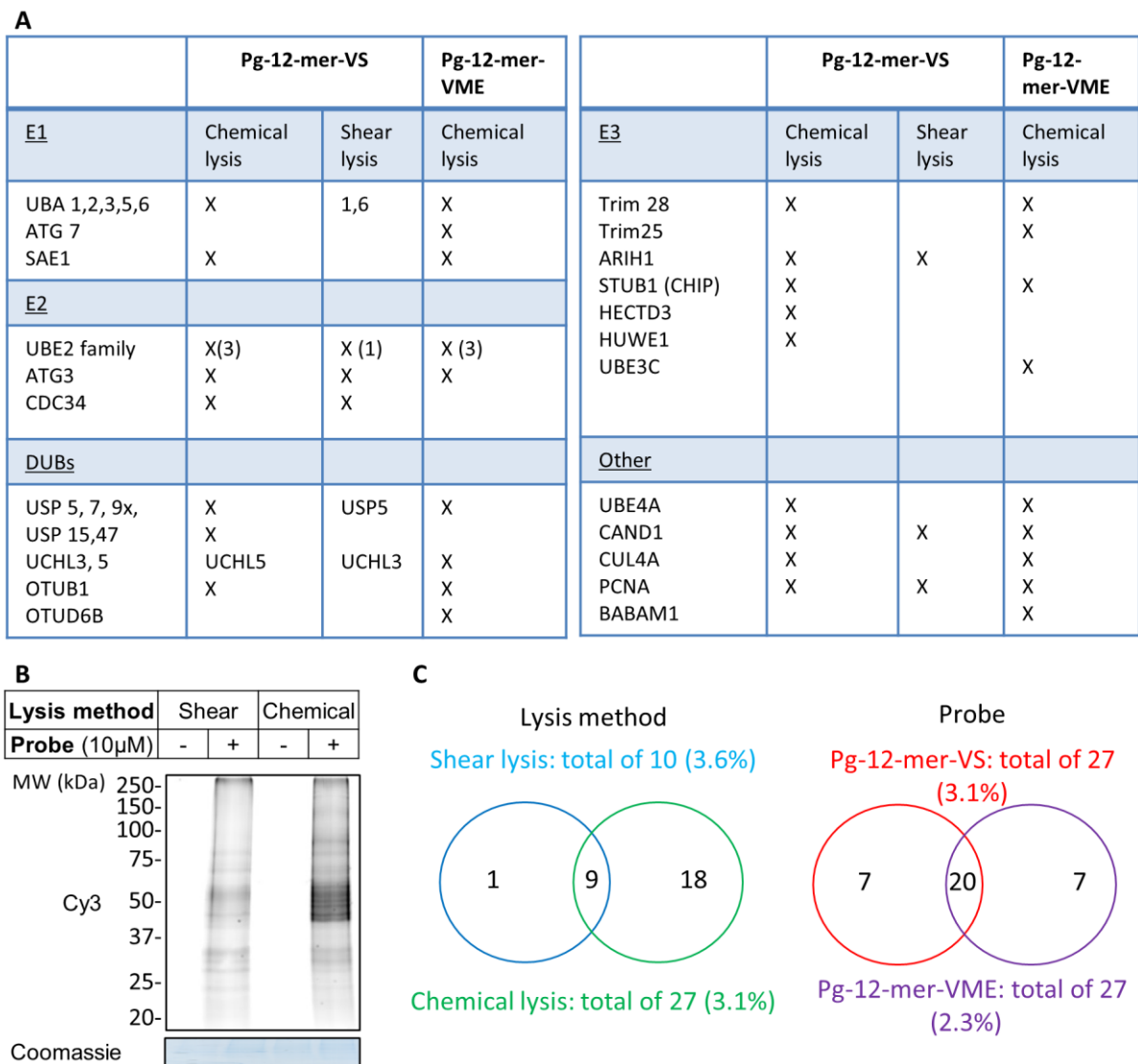
beads. Proteins were identified from their peptide MS data, by processing the raw data using MaxQuant and analysis using Perseus software (both introduced in Section 1.2.2). Identification by at least 2 'unique+razor' peptides with a false discovery rate (FDR) of 0.01 was set as a minimum to define a probe target.

On analysis of the chemically lysed lysate, 57 proteins were identified from the control as background 'non-specific binders', which were subtracted from the **Pg-12-mer-VS** and **Pg-12-mer-VME** probe intensities. 1172 protein targets were identified for **Pg-12-mer-VME** and 882 for **Pg-12-mer-VS**. Of these hits, 27 (2.3 % of proteins identified) and 27 (3.1 %) ubiquitination enzymes were detected respectively with E1, E2, E3 and DUBs all represented (Figure 17A). As Borodovsky *et al.* conducted competition studies in EL4 lysate, a direct comparison cannot be made, however, it is encouraging that four DUB targets; USP5, USP7, USP9x, and UCHL3 are identified in both studies.

**Pg-12-mer-VS** labelling in shear forced lysate was substantially lower; with only 16 background proteins observed, and 279 protein targets identified. Of these hits 10 (3.6 %) ubiquitination enzymes were detected respectively with E1, E2, E3 and DUBs all represented (Figure 17A). There are several possible explanations for the observed variation in protein targets with lysis condition: i) incomplete release of proteins by shear lysis in comparison to chemical lysis techniques; ii) constituent detergents in chemical lysis buffer lead to additional protein-probe interactions; iii) human error during sample preparation. Although ubiquitination machinery represents a higher percentage of observed labelling in shear forced lysate the loss of coverage of E1 and E3 enzymes is somewhat disappointing, therefore, chemical lysis conditions were adopted for future investigations.

In general, labelling associated with ubiquitination machinery is low across all samples, highlighting the promiscuous nature of these probes. As the difference in labelling by **VS** and **VME** electrophiles probes was small, it was decided that only one probe would be taken forward at this stage. **Pg-12-mer-VME** was selected for continued study owing to the higher yield of probe obtained. If quantitative analysis proved interesting for **Pg-12-mer-VME**, **Pg-12-mer-VS** would be revisited.





**Figure 17: Proteomic identification of Ub machinery labelled by Pg-12-mer-VS (10 µM) and Pg-12-mer-VME (50 µM). (A) Table summarising the Ub machinery identified under different lysis conditions by LC-MS/MS analysis. (B) in-gel fluorescence of Pg-12-mer-VS labelling in shear force and chemically lysed lysate. (C) Venn diagrams showing overlap of Ub machinery identified under different lysis conditions and with different probes.**

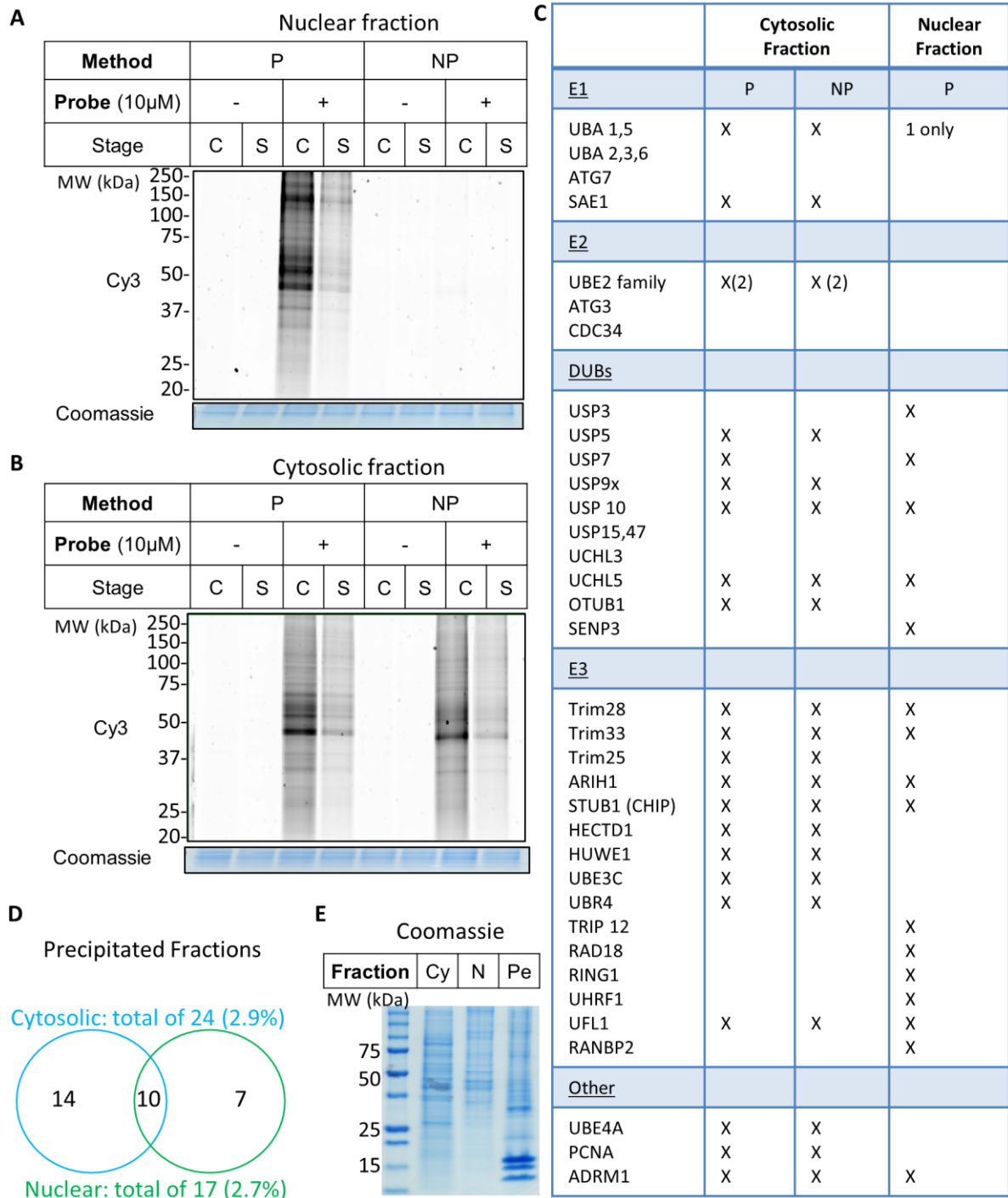
### 2.3.1 Maximising coverage: cellular fractionation

Given our interest in applying **Pg-12-mer-VME** to study ubiquitination during DDR, it was important to maximise the ability of the probe to label relevant proteins. Consequently, cellular fractionation was attempted to determine whether this would increase coverage, particularly of nuclear proteins. Breast cancer cell line MCF7 was chosen for carrying out further optimisation, as these cells were a relevant model system for subsequent DDR studies. A chemical method which was deemed mild enough to maintain native conditions was used. Briefly, the cytosolic fraction was prepared by lysis of MCF7 cells with 0.5 % NP-40 at 4 °C. On centrifugation and removal of this fraction, the remaining pellet was resuspended in a sucrose cushion and the nuclear extract isolated. Labelling was initially

evaluated by in-gel fluorescence: fractionated lysate was treated with 10  $\mu$ M **Pg-12-mer-VME** for 1 hour before being ligated to AzTB, separated and visualised as described in Section 2.3. At first labelling efficiency was observed to be poor in nuclear extract (Figure 18A). Removal of sucrose by protein precipitation, methanol washing, and suspension in 0.2 % SDS/PBS proved to rectify the labelling efficiency in nuclear extracts, whilst analogous precipitation of the cytosolic fraction did not substantially reduced labelling (Figure 18B).

For proteomic analysis, each biological condition was prepared in duplicate and processed and evaluated as described in Section 2.3. On analysis of the cytosolic fraction control samples, 278 proteins were identified as background 'non-specific binders'. UBA1 and USP5 were present in both control and probe samples, but at a greater intensity in the probe samples, and therefore were considered valid targets. 827 protein targets were identified above background. Of these hits 24 (2.9 %) ubiquitination enzymes were detected respectively with E1, E2, E3 and DUBs all represented (Figure 18C). By comparison, in the non-precipitated samples 706 proteins were identified, though only one Ub relevant hit was lost in comparison to precipitated sample. Upon analysis of the nuclear fraction control samples, 271 proteins were identified as background 'non-specific binders' and 637 protein targets were identified above background. Of these hits 17 (2.7 %) ubiquitination enzymes were detected respectively. Interestingly, no E2 enzymes were identified in the extract. Known nuclear localised Ub enzymes, USP3, RAD18, RANBP2 (E3 Sumo-protein ligase RANBP2), RING1 (E3 ubiquitin-protein ligase RING1) and UHRF1 (E3 ubiquitin-protein ligase UHRF1) were identified, which was not possible prior to fractionation. In comparison only 60 proteins were identified in the non-precipitated samples.

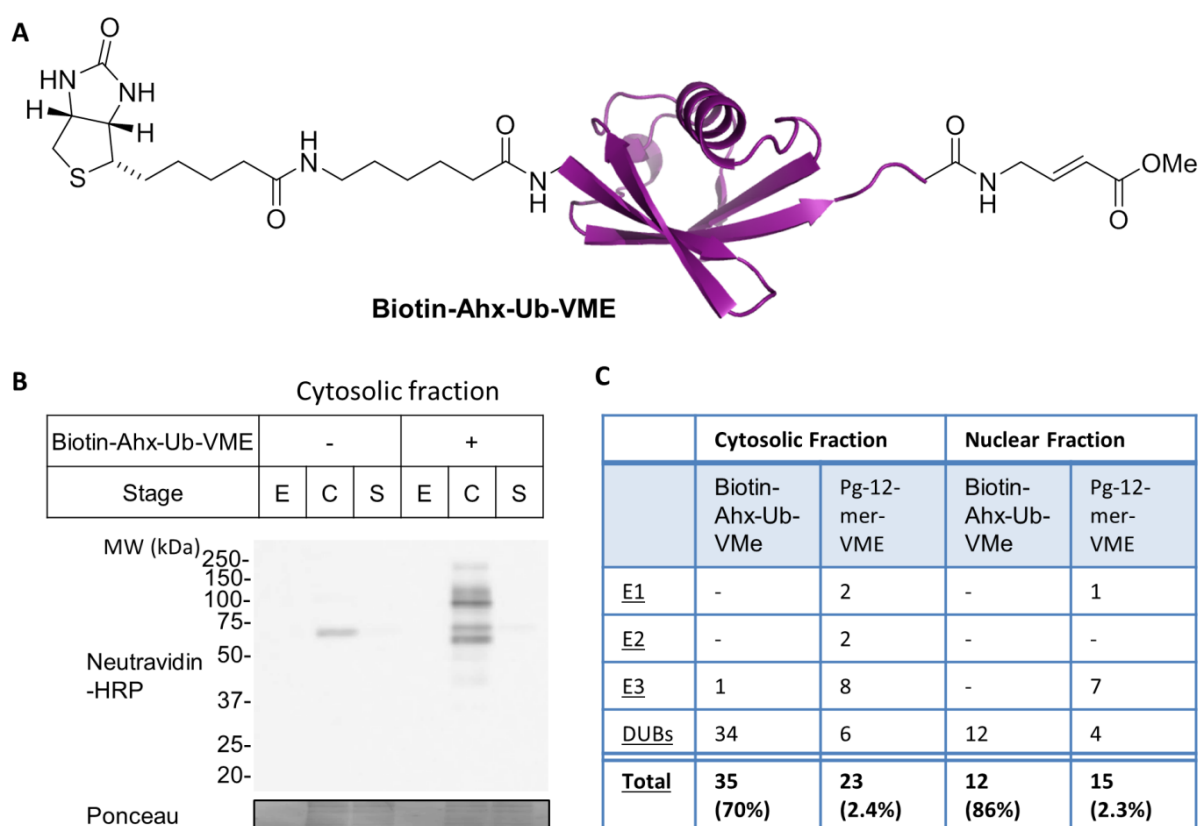
Together this data illustrates two main points: i) fractionation enables the detection of additional proteins of interest, notably USP3 and RAD18. ii) Buffer compatibility is highly important for click efficiency, but precipitation itself does not improve coverage. It should be noted that, due to mild lysis conditions employed to retain native extracts, nuclear extraction is not complete which may limit the identification of probe targets. This can be clearly observed by Coomassie staining as a lack of histones in the nuclear extract (Figure 18E).



**Figure 18: Pg-12-mer-VME labelling (10 µM) in fractionated MCF7 lysate. In-gel fluorescence of Pg-12-mer-VME labelling in (A) nuclear and (B) cytosolic fractions. Method: P = precipitated, NP = not precipitated after CuAAC ligation to AzTB. Stage: C = following CuAAC (and P or NP), S = supernatant following enrichment. (C) Table summarising the Ub machinery identified by LC-MS/MS analysis, in each cell fraction using P and NP methods. (D) Venn diagram showing overlap of Ub machinery identified in precipitated cytosolic and nuclear fractions. (E) Coomassie staining of Cytosolic (Cy) and Nuclear fractions (N), and cell pellet (Pe), all 10 µg.**

### 2.3.2 Evaluation of literature Ub- VME probe

To assess the value of **Pg-12-mer-VME** as an alternative to currently available probes, it was of interest to compare its capability to that of a commercially available Ub-VME probe, **Biotin-Ahx-Ub-VME** (Figure 19A). Fractionated lysate was treated with 200 nM **Biotin-Ahx-Ub-VME** for 1 hour at 37 °C before being spin filtered (10 kDa cut-off) multiple times and enriched on NeutrAvidin beads. Probe labelling and pull down efficiency was initially evaluated for the cytosolic fraction by biotin western blot, blotting against NeutrAvidin-HRP (Figure 19B). **Biotin-Ahx-Ub-VME** labels mostly in the higher MW region corresponding to several USP DUBs. Some background labelling is also observed in the absence of probe, due to naturally biotinylated proteins.



**Figure 19:** (A) Structure of commercially available Biotin-Ahx-Ub-VME. (B) Fractionated MCF7 cell lysate was treated with Biotin-Ahx-Ub-VME (200 nM) for 1 hr, filtered (10 kDa MWCO), enriched on NeutrAvidin beads and analysed by biotin western blot. Ponceau staining was used as a loading control. Stage: E = eluant, C= concentrate following filtration, S = supernatant following enrichment (C) Table comparing targets identified by LC-MS/MS with Biotin-Ahx-Ub-VME and Pg-12-mer-VME.

For proteomic analysis, samples were processed and analysed as described in Section 2.3. After subtraction of background proteins, 50 cytosolic fraction targets were identified for **Biotin-Ahx-Ub-VME**, 35 of which (70 %) were ubiquitination enzymes (Figure 19C). The majority of these proteins were DUB enzymes, with only HUWE1 (E3 ubiquitin-protein ligase HUWE1), not belonging to this

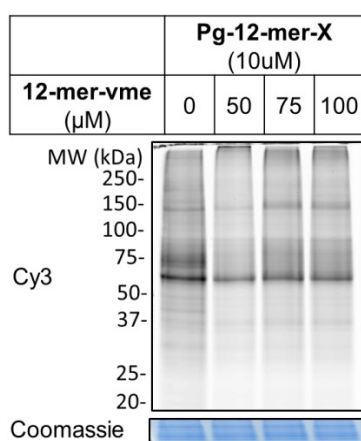
family. 14 protein targets were identified in the nuclear extract by **Biotin-Ahx-Ub-VME**, 12 of which (86 %) were DUBs; no other Ub related enzyme classes were identified in this fraction. In comparison, **Pg-12-mer-VME** identifies Ub machinery from all enzyme classes, but this represents a much lower percentage of overall probe labelling.

Overall, this reflects the literature view that the Ub-derived probes are highly specific to DUBs, and are superior for the study of this enzyme family in lysate models. Clearly the more promiscuous, **12-mer-VME** derived probes interact with several enzyme classes, including E3 ligase families. These findings indicate that such probes offer some potential as a broader ubiquitination machinery probe, though it remains to be seen whether quantitative changes for enzyme of interest can be observed within the context of promiscuous labelling.

## 2.4 Quantitative proteomics

With promising results by non-quantitative analysis, work commenced to determine the competitive targets of **Pg-12-mer-VME** using quantitative proteomics. To reduce instrument time, competition was first analysed using non-fractionated native whole cell lysate, with the rationale that if this provided encouraging results, fractionation could be attempted at a later stage.

First, competition was assessed by in-gel fluorescence: cells were incubated with an excess of **12-mer-VME** for 30 minutes, followed by **Pg-12-mer-VME** incubation at 10  $\mu\text{M}$  for 1 hour before being ligated to AzTB, separated and visualised as described in Section 2.3 Interestingly, a decreased labelling intensity was observed with a 5-fold (50  $\mu\text{M}$ ) excess of inhibitor, but this observed competitive effect was reduced with higher excess of inhibitor. It was postulated that this is due to limited solubility of the inhibitor, which was observed to crash out of the reaction solution when in high excess. Therefore, 5-fold (50  $\mu\text{M}$ ) and 2-fold (20  $\mu\text{M}$ ) competition conditions were taken forward for proteomic evaluation by Spike-in SILAC.



**Figure 20: In-gel fluorescence of Pg-12-mer-VME labelling in competition with parent inhibitor 12-mer-VME. Whole cell MCF7 lysate was pre-treated with inhibitor or DMSO for 30 min and then incubated with probe for 1 hr.**

### 2.4.1 SILAC analysis of competition

MCF7 whole cell lysate was incubated in triplicate with varying concentrations of **12-mer-VME** (0, 20 and 50  $\mu\text{M}$ ) for 30 minutes at 37  $^{\circ}\text{C}$  before incubation with **Pg-12-mer-VME** (10  $\mu\text{M}$ ) for 1 hour. Separately, 'heavy' R10K8-labelled MCF7 whole cell lysate was treated with **Pg-12-mer-VME** (40  $\mu\text{M}$ ) for 1 hour at 37  $^{\circ}\text{C}$  before cell lysis to form the 'spike'. Following protein concentration determination, a fixed amount of spiked lysate was added to the normal lysate in a 1:4 ratio before being ligated to AzTB, enriched, reduced, alkylated, trypsin-digested, desalted, and separated by LC-

MS/MS as previously described in Section 2.3. The raw data was processed using MaxQuant (introduced in Section 1.2.2). As for non-quantitative analysis, peptide identification required at least 2 'unique+razor' peptides with an FDR of 0.01. In addition, at least 2 'unique+razor' peptides were required for a valid quantification. The calculated H/L SILAC ratios were further analysed in Perseus (introduced in Section 1.2.2). The SILAC ratios were inverted, to give L/H SILAC ratios, and transformed with  $\text{Log}_2$  to aid subsequent graphical representation of the data. The triplicates were categorically grouped, and proteins filtered such that at least 2 valid ratio values are required in at least one group across the experiment. Average L/H ratios were taken across each group, and these averages normalised against the group without any competition (**12-mer-VME** = 0  $\mu\text{M}$ ) to generate a ratio of ratios. The lower this ratio of ratio value, the more effect the competition observed. The results are summarised in Figure 21.

On analysis, 811 proteins were identified with valid ratio of ratios. 21 ubiquitination enzymes were identified, the majority of which had previously been identified by non-quantitative analysis. USP14, and E3 ubiquitin-protein ligase RNF14 (RNF14) were identified for the first time, possibly owing to the alternative lysis conditions. Applying a threshold ratio change of greater than 1.5 fold (equivalent to a  $\text{log}_2$  fold change of 0.58), 221 of these proteins were categorised as 'biologically significant' shifters. On further analysis by the student T-test ( $P\text{-value} < 0.05$ ) 67 proteins demonstrated statistically significant shifts. Ubiquitin-like-conjugating enzyme ATG3 (ATG3), USP14, and E3 ubiquitin-protein ligase TRIM33 (TRIM33) all demonstrated biological and statistical competitive labelling by **Pg-12-mer-VME**, with CHIP lying just outside the set threshold (Figure 21A). It should be noted that this is a low statistical threshold, however, and that the quantitative changes observed were not deemed significant when subjected to a more stringent perturbation-based FDR T-test.

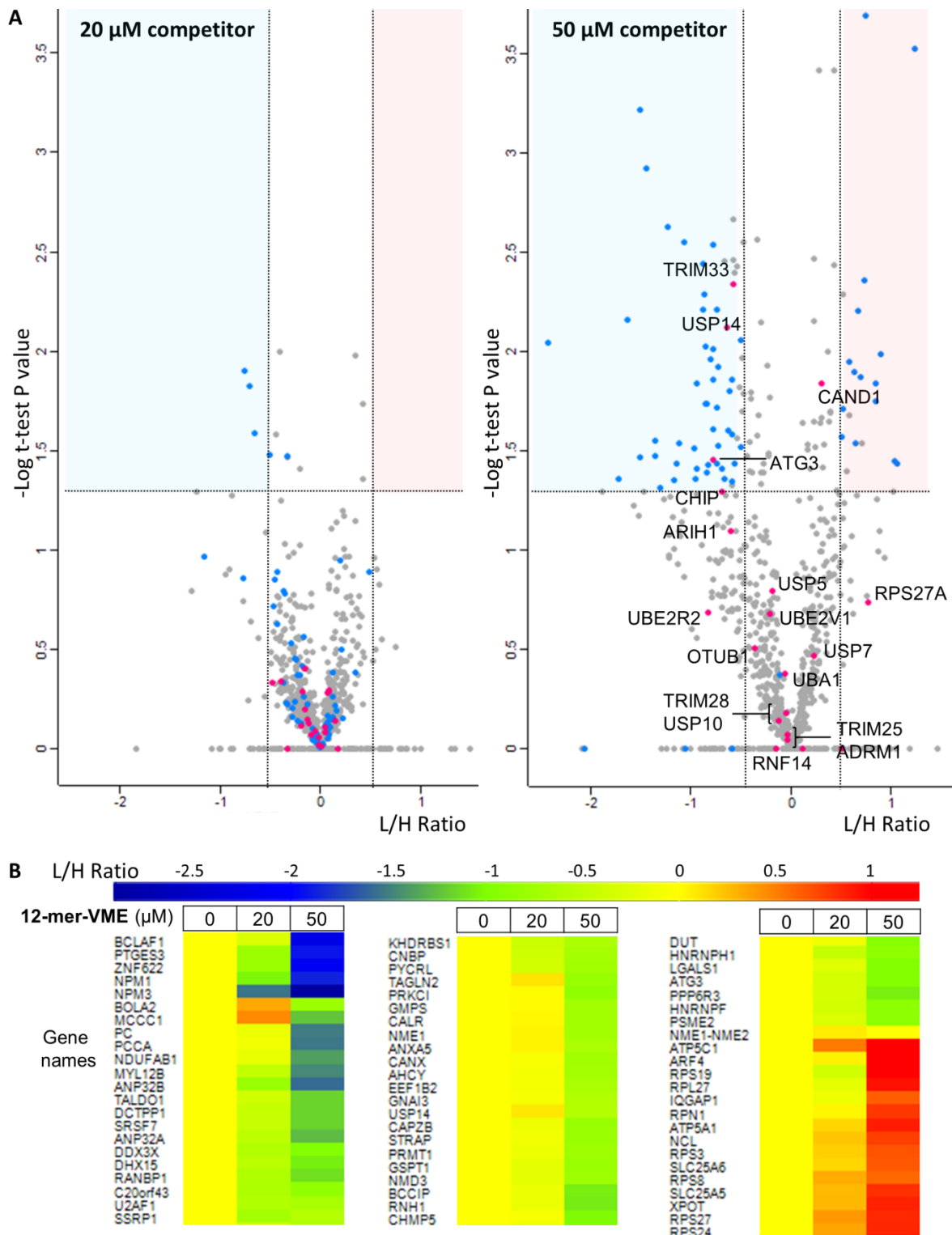
ATG3 is an E2 enzyme, necessary for cytoplasm-to-vacuole transport, autophagy and mitochondrial homeostasis.<sup>240, 241</sup> USP14 is a proteasome associated DUB, which facilitates Ub 'recycling' after proteins have been targeted for degradation.<sup>242</sup> Both of these enzymes utilise a catalytic cysteine. TRIM33 belongs to the RING E3 family and does not possess a catalytic cysteine. However, mutagenesis studies have shown that replacement of either Cys<sup>125</sup> or Cys<sup>128</sup> with alanine abolishes E3 activity, without affecting its interaction with its substrate, SMAD4 (Mothers against decapentaplegic homolog 4).<sup>243</sup> CHIP is an E3 ligase that targets misfolded chaperone proteins, such as heat shock proteins, for degradation.<sup>244</sup> As CHIP is an Ubox protein it also does not utilise a catalytic cysteine, however, mutagenesis studies do indicate that other nucleophilic residues, Lys<sup>30</sup> and His<sup>260</sup>, are necessary to ubiquitinate Forkhead box protein P3 (FOXP3), a target of CHIP.<sup>245</sup> Some significantly induced, rather than competed, labelling of ubiquitination machinery was also

observed. Labelling of E3-cullin ligase activity regulator CAND1 was induced but did not lie in the set range of biological significance.

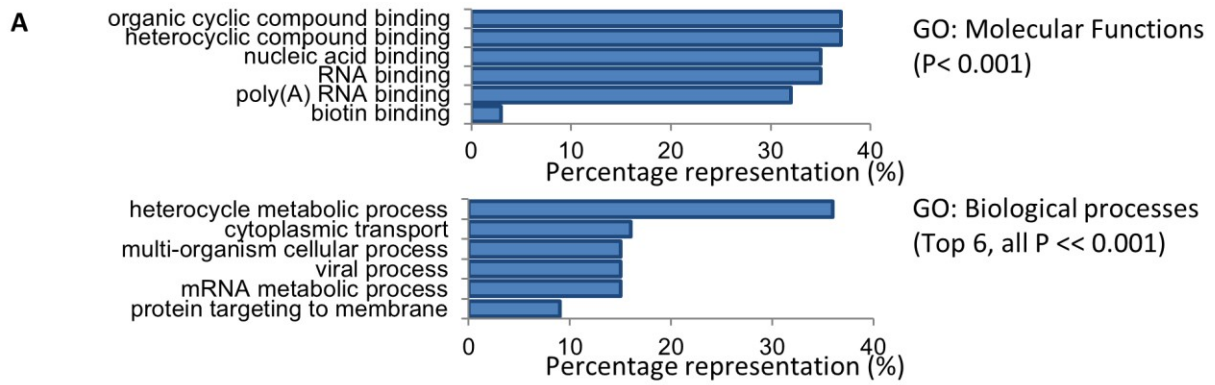
Considering all significant hits, several histone chaperones were identified. Two members of the Nucleoplasmin family: Nucleoplasmin-1 (NPM1) and Nucleoplasmin-3 (NPM3) demonstrated a large ratio change with competition which can be seen most clearly with the heat map representation (Figure 21B). NPM1 is involved in multiple biological processes, including histone assembly and centrosome duplication.<sup>246</sup> Although NPM3 is less well characterised, the homology between NPM1 and NPM3 suggests that they have similar functions. Only three proteins demonstrated significant competition with the lowest concentration of inhibitor: Zinc finger protein 622 (ZNF622), Acidic leucine-rich nuclear phosphoprotein 32 family member B (ANP32B), and FACT complex subunit SSRP1 (SSRP1). ZNF622 has been linked to apoptosis in response to oxidative stress.<sup>247</sup> ANP32B is another histone chaperone, which has been described as an anti-apoptotic protein as well as a cell cycle progression factor.<sup>248, 249</sup> SSRP1 is a component of the FACT complex: a general chromatin factor involved in multiple DNA related processes, including transcription elongation during which FACT acts as a histone chaperone.<sup>250</sup> The FACT complex also plays a role in DSB DNA repair.<sup>251</sup> 15 proteins showed considerable induction of labelling on pre-incubation with inhibitor. Though these are not competitive targets of **Pg-12-mer-VME**, they do suggest that **Pg-12-mer-VME** can label proteins that are activated by parent inhibitor **12-mer-VME**, suggesting that their common scaffold is capable of both inducing and interacting with certain proteins.

Evaluation of GO annotations (introduced in Section 1.2.2) for significant hits revealed a strong association of molecular function with RNA, nucleic acid, and organic cyclic compound binding (Figure 22A). Biological functions were diverse, with 22 strongly enriched ( $P < 0.001$ ) GO terms generally associated with metabolism, protein transport, and protein targeting to membranes. The top 6 of these GO terms are plotted in Figure 22A. Network analysis using STRING (<http://string-db.org/>) also identified multiple connections between the competed targets, with substantial interactions within smaller protein subsets (Figure 22B).





**Figure 21: Spike-in SILAC competition analysis.** Samples were prepared in triplicate and spike added in a 1:4 ratio. (A) Volcano plots of Pg-12-mer-VME (10  $\mu$ M) labelling in competition with indicated concentrations 12-mer-VME. Dashed lines represent statistical (vertical) and biological (horizontal) significance thresholds. Proteins in the shaded regions meet both criteria. Proteins which show significant competition under either condition (blue) and all Ub machinery (Pink, labelled) are highlighted. (B) Heat map of the 67 significantly competed targets, with L/H ratio at each inhibitor concentration indicated.



**B** STRING Network (Interaction / source)

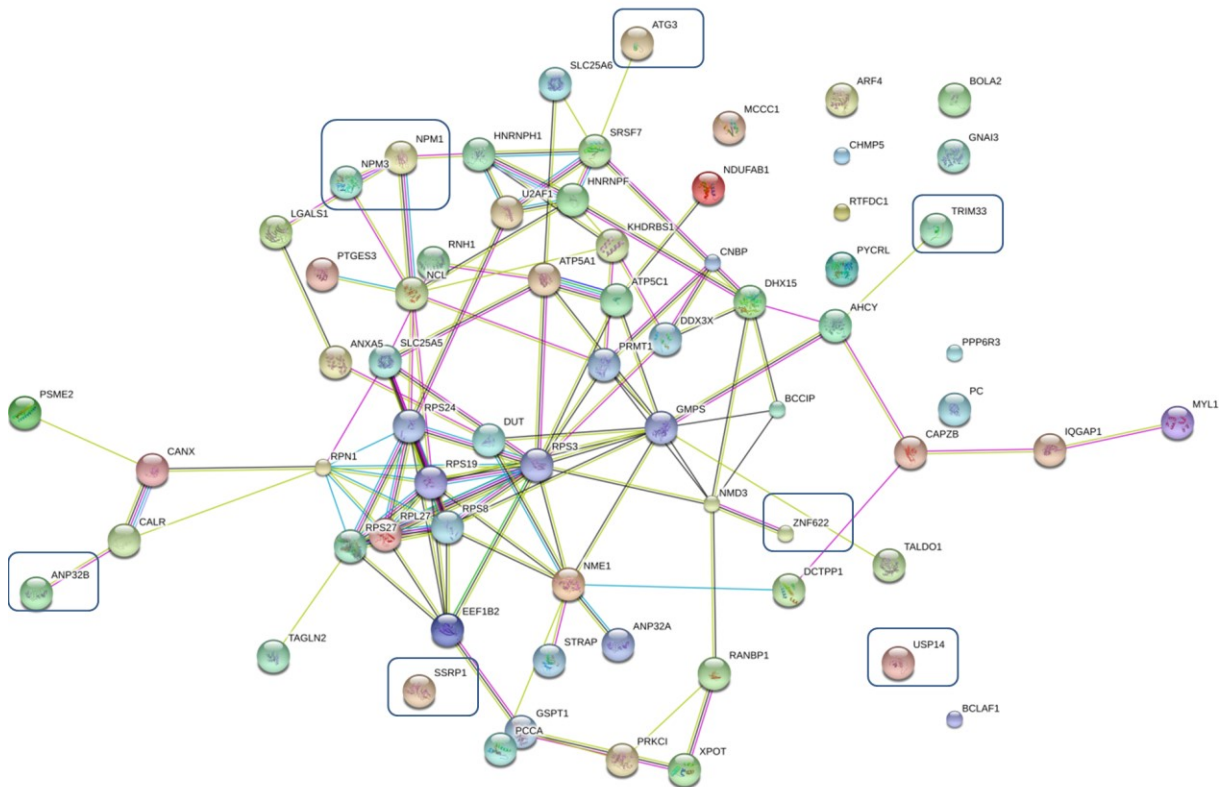
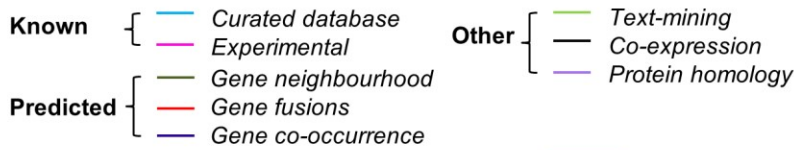
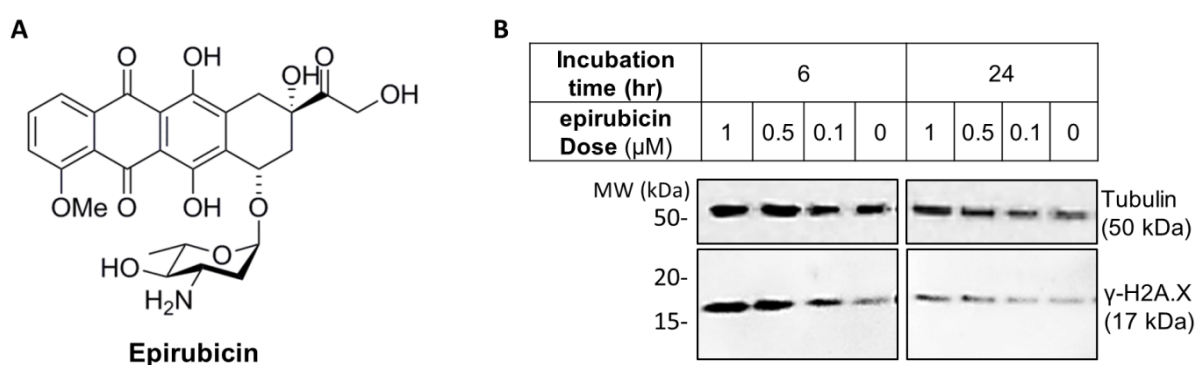


Figure 22: (A) GO terms and (B) STRING network evaluation of the 67 significant targets. Lines in the STRING evaluation represent evidenced interactions between proteins, with evidence source indicated in the key. Proteins discussed in the text are boxed.

## 2.4.2 Studies of a basic DDR model

Having identified some Ub machinery as competitive targets of **Pg-12-mer-VME**, it was of interest to study them in a DDR model. It was further postulated that quantitative changes may be identified for non-competitive target proteins, due to pathway activation in response to DNA damage.

First, a basic DDR model was generated by treatment with epirubicin. Epirubicin is a chemotherapeutic used in the treatment of several cancers, including breast cancer.<sup>252</sup> It is a DNA damaging agent which, amongst other mechanisms of damage, forms double strand breaks.<sup>253</sup> Histone H2A.X is phosphorylated on residue Ser139 ( $\gamma$ -H2A.X) at the start of double strand break repair and can therefore act as a biomarker for this process.<sup>254</sup> MCF7 cells were treated with varying concentrations of drug and incubated for 6 or 24 hours. Following cell lysis, samples were separated by SDS-PAGE, transferred onto PVDF membrane and analysed by western blot. On blotting for  $\gamma$ -H2A.X, accumulation of  $\gamma$ -H2A.X was observed with higher concentrations of epirubicin at both time points tested, signifying higher levels of induced DNA damage. The difference in  $\gamma$ -H2A.X signal across the selected concentration range was greater at 6 hours, suggesting that this is still within the active DDR timeframe. Therefore, quantitative proteomic analysis was undertaken for the 6 hour time point to compare labelling by **Pg-12-mer-VME** in epirubicin treated (0, 0.1 and 1  $\mu$ M) lysate.



**Figure 23: (A) Structure of epirubicin. (B) MCF7 cells were incubated with varying doses of epirubicin and incubated for 6 or 24 hr. Following cell lysis, samples were analysed by western blot. Tubulin is included as a loading control.**

Following cell treatment and native whole cell lysis, lysates were treated with **Pg-12-mer-VME** for 1 hour. Separately, 'heavy' R10K8-labelled MCF7 whole cell lysate was treated with **Pg-12-mer-VME** (40  $\mu$ M) for 1 hour at 37  $^{\circ}$ C before cell lysis to form the 'spike'. Following protein concentration determination, a fixed amount of spiked lysate was added to the normal lysate in a 1:4 ratio before being ligated to AzTB, enriched, reduced, alkylated, trypsin-digested, desalted, and separated by LC-MS/MS as described in Section 2.3. The raw data was processed using MaxQuant and Perseus as described in Section 2.4.1. The results are summarised in Figure 24.

On analysis, 902 proteins were identified with valid ratio of ratios, including 26 ubiquitination machinery. Applying a threshold ratio change of greater than 1.5 fold, 136 of these proteins were categorised as 'biologically significant' shifters. On further analysis by the student T-test (P-value < 0.05) 30 proteins demonstrated a statistically significant shift across the 3 biological conditions. RNF14, an RBR E3 ligase, fell into this category with another RBR, ARIH1, lying just outside the biological threshold. RNF14 has been reported as a positive regulator of Wnt signalling in colon cancer cells, whilst ARIH1 is reported to protect cells from genotoxic stress by mediating DNA damage-induced translation arrest.<sup>255, 256</sup> To reduce the number of hits, a higher statistical threshold was applied ( $p < 0.01$ ), leaving 10 hits. Although all of these proteins had previously been identified in the competition experiment with **12-mer-VME** none had been previously identified as significant. Furthermore, only one of the 10 targets, SUPT16H (FACT complex subunit SPT16), is associated with DDR. SUPT16H is a component of the FACT complex, which was first mentioned in Section 2.4.1. Oliveira *et al.* report that FACT plays an important role at the start of the HR pathway of DSB repair. The same authors showed that depletion of SUPT16H decreases HR activity.<sup>251</sup> Ribosome biogenesis protein BOP1 (BOP1), which is linked to ribosome assembly and cell cycle progression, showed the greatest ratio increase in the DDR model.<sup>257, 258</sup> ATP-dependent RNA helicase DDX39A (DDX39A) showed the greatest statistically significant change. DDX39A is required during nuclear export of mRNA, but to date it has not been linked to DDR.<sup>259</sup>

In summary, some targets of **Pg-12-mer-VME** were observed to be depleted or induced on introduction of DNA damage; however, none of them were previously identified as competitive probe targets. It should be noted that a low statistical threshold has been applied, however, and that the quantitative changes observed were not deemed significant when subjected to a more stringent perturbation-based FDR T-test.

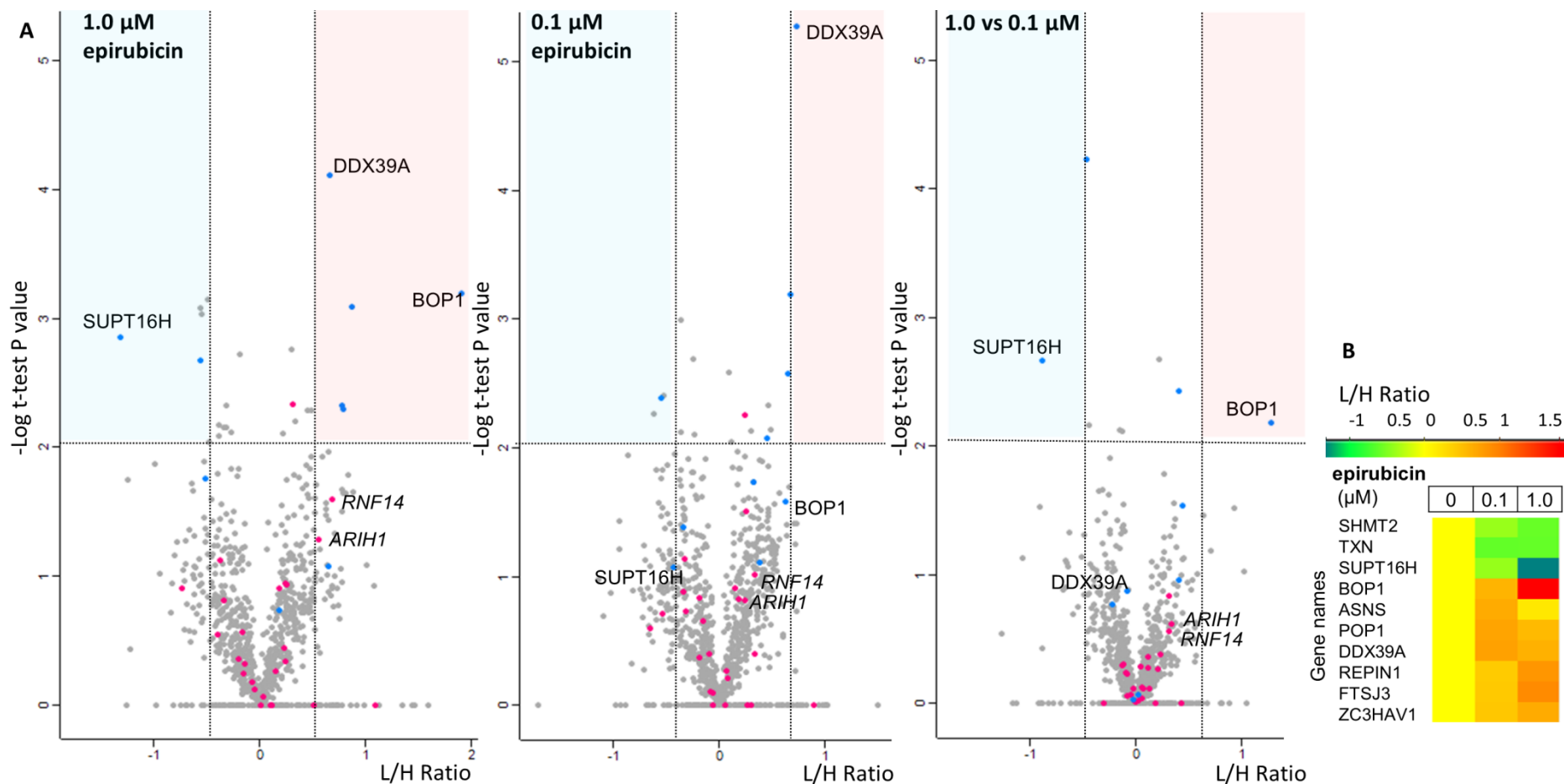
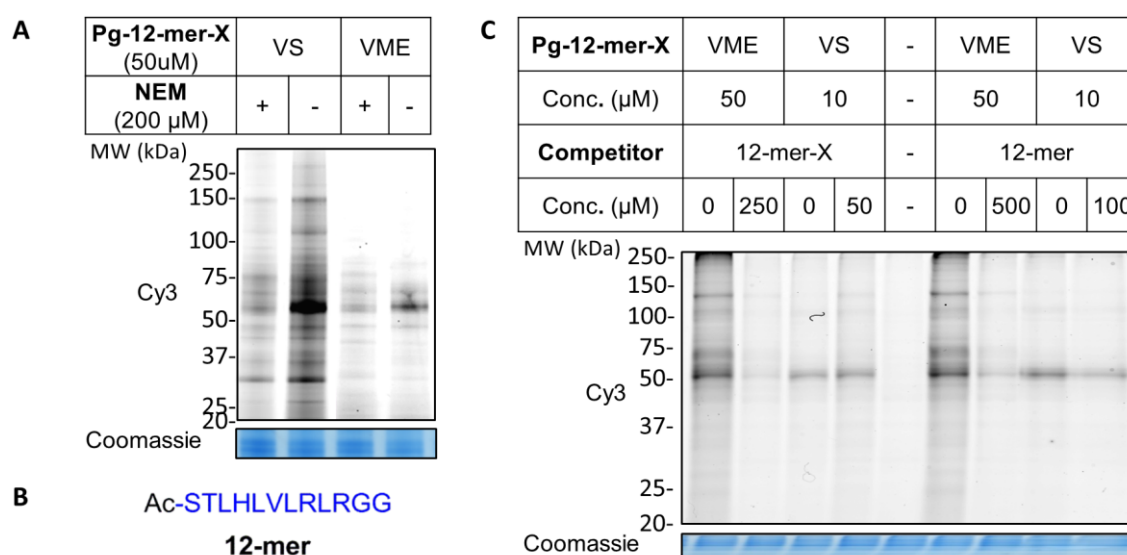


Figure 24: Spike-in SILAC DDR analysis. Samples were prepared in triplicate and spike added in a 1:4 ratio. (A) Volcano plots of Pg-12-mer-VME (10  $\mu$ M) labelling following epirubicin (0 0.1 or 1.0  $\mu$ M) treatment. Dashed lines represent statistical (vertical) and biological (horizontal) significance thresholds. Proteins in the shaded regions meet both criteria. Proteins which show significant competition under either condition (blue) and all Ub machinery (Pink, labelled) are highlighted. (B) Heat map of the 10 significantly competed targets, with L/H ratio at each epirubicin dose indicated.

## 2.5 Preliminary in cell application

In parallel with experiments in lysate, it was of interest to determine whether this activity could be translated into intact cells. Initially, both **Pg-12-mer-VME** and **Pg-12-mer-VS** were tested, to determine the effect, if any, of the electrophilic trap on cell permeability.

First, HeLa cells were incubated with 50  $\mu\text{M}$  of **Pg-12-mer-VME** or **Pg-12-mer-VS** for 1 hour prior to cell lysis. Lysates were then ligated to AzTB, separated and visualised as described in Section 2.3. Pre-incubation with NEM was used as a control for competitive labelling. Although 10 mM NEM was used for analogous in lysate applications, 200  $\mu\text{M}$  was chosen for in cell treatments, as this did not cause cell viability issues. It was observed by in gel fluorescence that both probes are capable of labelling in-cell, in an activity dependent manner (Figure 25A). Further studies were then conducted to ascertain probe competition in cells against the parent inhibitor peptide sequence in cell. In addition to **12-mer-VME** and **12-mer-VS**, the probes were also competed against peptide lacking an electrophilic trap for comparison (**12-mer**, Figure 25B). MCF7 cells were incubated with an excess of inhibitor for 30 mins at 37  $^{\circ}\text{C}$ , before treatment with **Pg-12-mer-VME** or **Pg-12-mer-VS** for 1 hour prior to cell lysis. Lysates were then ligated to AzTB, separated and visualised as described in Section 2.3 (Figure 25C). Labelling by **Pg-12-mer-VME** was clearly out competed by 5-fold (250  $\mu\text{M}$ ) excess of **12-mer-VME**, or 10-fold (500  $\mu\text{M}$ ) excess **12-mer**. Analogous competition for **Pg-12-mer-VS** was less convincing however, with minimal loss of labelling intensity observed with both inhibitors.



**Figure 25: Probe application in intact cells. (A)** HeLa cells were pre-treated with NEM or DMSO for 20 min, incubated with probe for 1 hr, lysed, and analysed by in-gel fluorescence. **(B)** MCF7 cells were pre-treated with inhibitor or DMSO for 30 min and then incubated with probe for 1 hr, lysed, and analysed by in-gel fluorescence.

**Pg-12-mer-VME** was taken forward for further non-quantitative proteomic analysis in intact cells. Briefly, MCF7 cells were treated with **Pg-12-mer-VME** (50  $\mu$ M) for 1 hour at 37 °C. Following cell lysis, the probe-enzyme complexes formed were ligated to AzTB, enriched, and processed for LC-MS/MS analysis as previously described in Section 2.3. Control samples lacking probe, were prepared in parallel in order to identify any background signal due to unspecific interactions with the beads. Proteins were identified from their unique peptide MS data, by evaluating the raw data was using MaxQuant and Perseus software. Disappointingly, after multiple attempts, no ubiquitination machinery was detected by **Pg-12-mer-VME**.

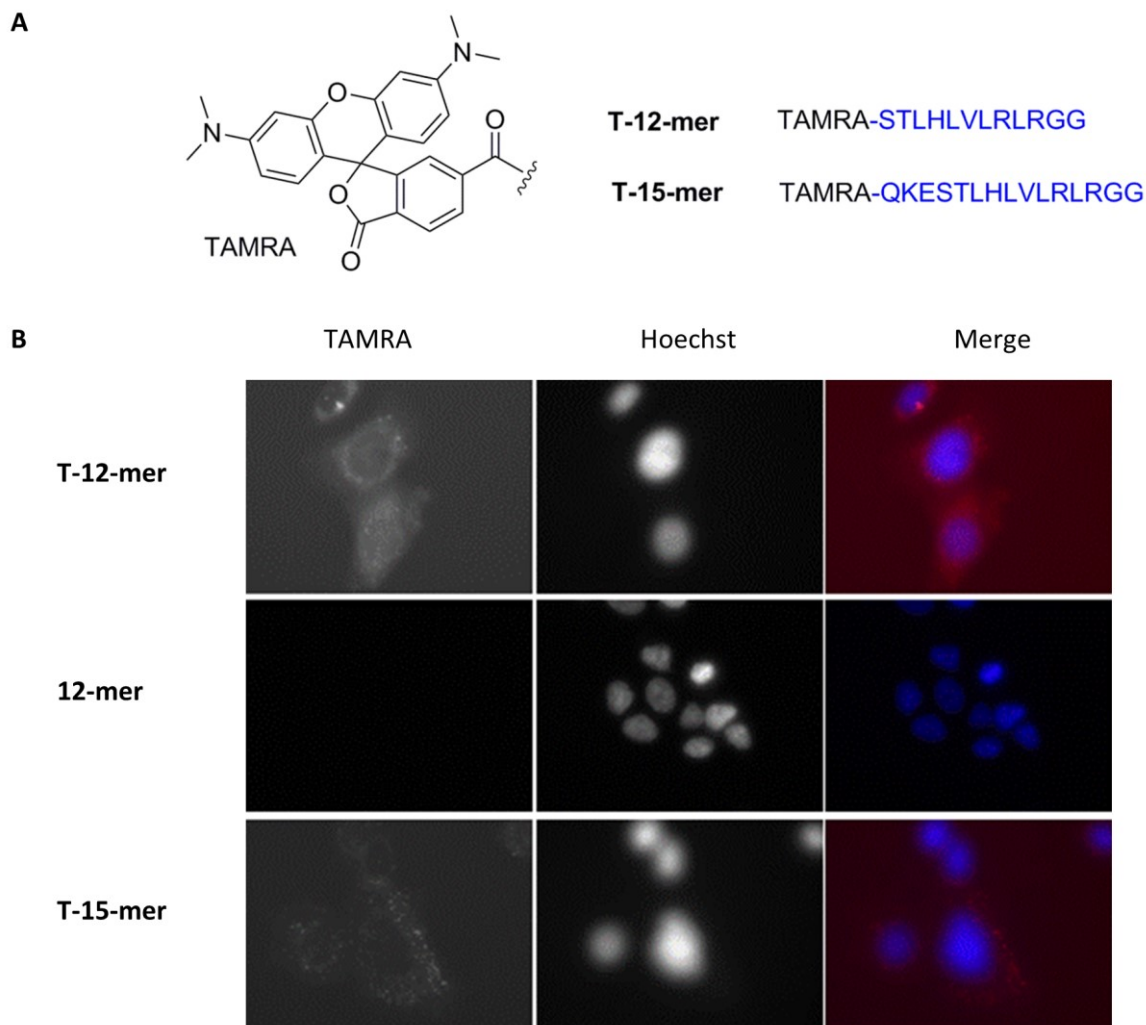
### 2.5.1 Determining peptide cellular localisation

It was hypothesised that probe labelling ability was being hindered by the cellular localisation of the probe peptide sequence. Therefore, further study into the cellular localisation of **Pg-12-mer-VME** was undertaken using live cell fluorescent imaging. Accordingly, fluorescently labelled analogues **T-12-mer-VME** and **T-12-mer** (Figure 26A) were synthesised. Visualisation of such analogues would allow the influence of the electrophilic trap and the peptide sequence on cellular localisation to be decoupled. TAMRA was chosen, due to its relatively low cost in comparison to other fluorophores and its reported cell permeability in isolation.<sup>260</sup> In a cross comparison study Fischer *et al.* further demonstrated that cellular import is peptide, rather than fluorophore, driven.<sup>261</sup> It is possible that a longer peptide sequence would improve probe selectivity towards ubiquitination machinery, therefore, in addition to the **12-mer** sequence **T-15-mer-vme** and **T-15-mer** analogues were also synthesised to ascertain their suitability as cell-based probes.

TAMRA was successfully coupled to the N-terminus of both peptide sequences in the presence of HATU and DIPEA. The reaction proceeded cleanly as confirmed by LCMS analysis, thus, subsequent TFA deprotection, *t*BME precipitation and LCMS purification, afforded analogues **T-12-mer** and **T-15-mer** in high yields of 20 % and 30 % respectively. In order to synthesise **T-12-mer-VME** and **T-15-mer-vme**, mild HFIP resin cleavage was utilised as previously described in Section 2.2, followed by in-solution coupling of the trap in the presence of PyBop and DIPEA. However, although this coupling was attempted several times, it proved unsuccessful. It is unlikely that the N-Terminal TAMRA is interfering with HFIP cleavage ability; HFIP is frequently reported as a solvent in peptide reactions, including TAMRA coupling, and the crude mass obtained was not unusually low.<sup>262</sup> Alternatively, it is possible that the peptide is too hindered for the in-solution coupling reaction.

Although **T-12-mer-VME** and **T-15-mer-vme** synthesis was unsuccessful, **T-12-mer** and **T-15-mer** localisation was still assessed by live cell imaging. **12-mer** was included as a TAMRA-free control.

Briefly, MCF7 cells were incubated at 37 °C with 10 μM of **T-12-mer** or **T-15-mer** or **12-mer** for 1 hour, before addition of Hoechst nuclear stain and further incubation for 30 minutes. The cells were then gently washed with Hank's PBS and imaged on the plate (Figure 26B).



**Figure 26: Localisation studies. (A) Structures of T-12-mer and T-15-mer probes. (B) MCF7 cells were treated with probe or control (10 μM) for 1 hr, before incubating with Hoechst nuclear stain for a further 30 min. cells were washed and TAMRA and Hoechst fluorescence visualised using Cy3 (excitation wavelength 552 nm, emission wavelength 570 nm) and DAPI (excitation wavelength 358 nm, emission wavelength 461 nm) channels respectively.**

It was observed by Cy3 labelling that **T-12-mer** and **T-15-mer** display distinctive localisation patterns. On analysis of the merged images, **T-12-mer** displays broad cellular localisation in the cytoplasm and potentially in the nuclear envelope. This suggests that the observed loss of labelling of ubiquitination machinery in cell is not attributed to cellular localisation. Conversely **T-15-mer** displays weaker and more localised labelling. Neither probe demonstrated nuclear localisation. It should be noted that these results are preliminary and further replicates using glass slides are required in order to



comment significantly about the localisation observed. Due to time restraints, work continued with the 12-mer peptide sequence, though 15-mer does offer an interesting alternative for future study.

## 2.5.2 Peptide development for in cell applications.

In parallel to cell localisation studies, the addition of cell penetrating peptide sequences to **Pg-12-mer-VME** was considered. It was hypothesised that directing the probe would increase labelling of proteins of interest. Addition of a nuclear localisation sequence, for example, may enhance labelling of nuclear proteins. Alternatively, the addition of a cell penetrating sequence could be used to generally improve peptide permeability and therefore improve cellular labelling. To test this hypothesis a small series of alternative probes based on **Pg-12-mer-VME** were designed. N-terminal additions of a nuclear localisation signal (**NLS**) and two patented cell penetrating peptide sequences (**CCP1** and **CCP2**) were synthesised and tested (Figure 27).<sup>263, 264</sup> All sequences were made successfully following standard SPPS procedures and coupled to **VME** as described in Section 2.2. Due to the high prevalence of Arg residues in these sequences, a longer TFA deprotection time of six hours was required to ensure complete side chain deprotection. Following LC-MS purification **NLS**, **CCP1** and **CCP2** were obtained in 1.7 %, 4.2 %, and 2.3 % yields respectively. Though low, this provided sufficient material for preliminary testing.

MCF7 cells were incubated with 10  $\mu$ M of each probe for 1 hour prior to cell lysis. Lysates were then ligated to AzTB, separated and visualised as described in Section 2.3 (Figure 27B). Unfortunately, labelling ability of the new probes was lower than that of **Pg-12-mer-VME**, with the strong  $\sim$ 60 kDa signal not observed. Therefore this avenue was not further studied.

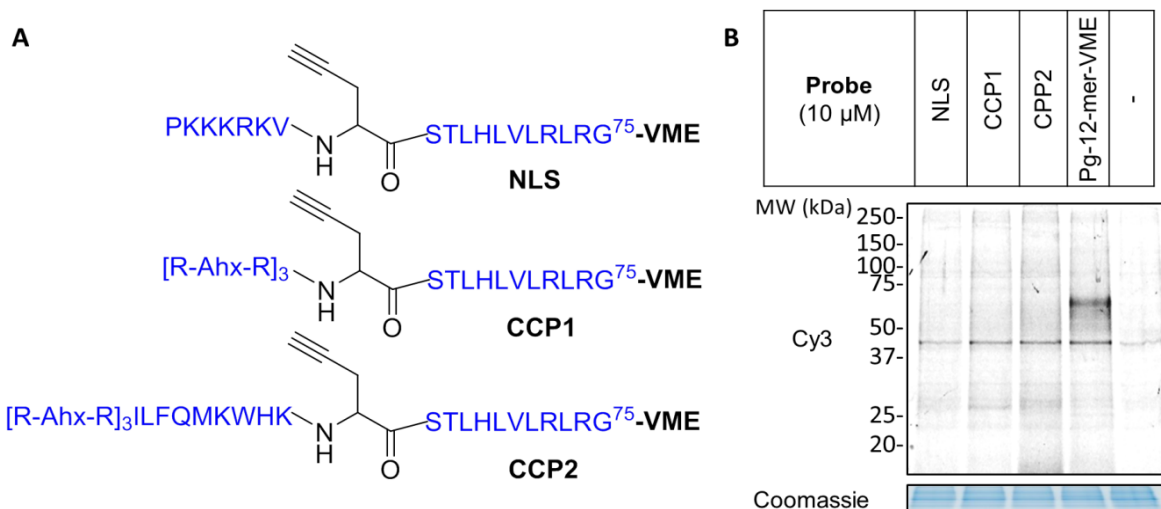


Figure 27: (A) Structures of NLS, CCP1, and CCP2 probes. (B) MCF7 cells were incubated with each probe (10  $\mu$ M) for 1 hr, lysed, and analysed by in-gel fluorescence.

## 2.6 Conclusions

The work in this Chapter focuses on the design, synthesis, and analysis of ABPs derived from the 12 C-terminal residues of Ub. Alkyne tagged probes containing an electrophilic trap, **VS** or **VME** were designed and synthesised by in-solution phase coupling of the trap. Labelling of Ub machinery was observed for both probes in lysate. As lysate proteomic analysis showed a minimal difference in labelling between the two traps, only VME was taken forward for the purposes of further optimisation. The coverage of nuclear targets was somewhat improved by utilising a mild cell fractionation method. On comparison, Ub-derived probes showed high selectivity for DUBs, whilst **Pg-12-mer-VME** was highly promiscuous. Though unselective, **Pg-12-mer-VME** did label a number of E3 ligases, which were not identified by its protein counterpart. Quantitative proteomic analysis on competition and in a DDR model revealed several quantifiably engaged targets with diverse biological roles. Some ubiquitination machinery demonstrated competitive labelling, however, the confidence in these findings is restricted to the non-stringent statistical thresholds applied.

**Pg-12-mer-VME** showed initial promise as a cellular probe by in-gel fluorescence; however, no ubiquitination machinery was labelled in cell. To determine whether this was a result of peptide cellular localisation, TAMRA labelled probes were applied in live cell fluorescent imaging. Preliminary results suggested that the 12-mer peptide widely localised across the cell, whilst the longer 15-mer peptide demonstrated more localised labelling. Probes with an additional cell penetrating peptide or NLS were synthesised and tested, but did not improve probe labelling observed by in-gel fluorescence.

Though this work has successfully shown labelling of ubiquitination machinery in lysate, robust quantification of this labelling has proven non-facile. Furthermore, the labelling observed was not translatable into cells. Chapter 3 discusses small molecule derived probes as an alternative approach for the development of probes suitable for in cell applications.

## **Chapter 3    Small molecule activity-based probes for studying ubiquitination enzymes**

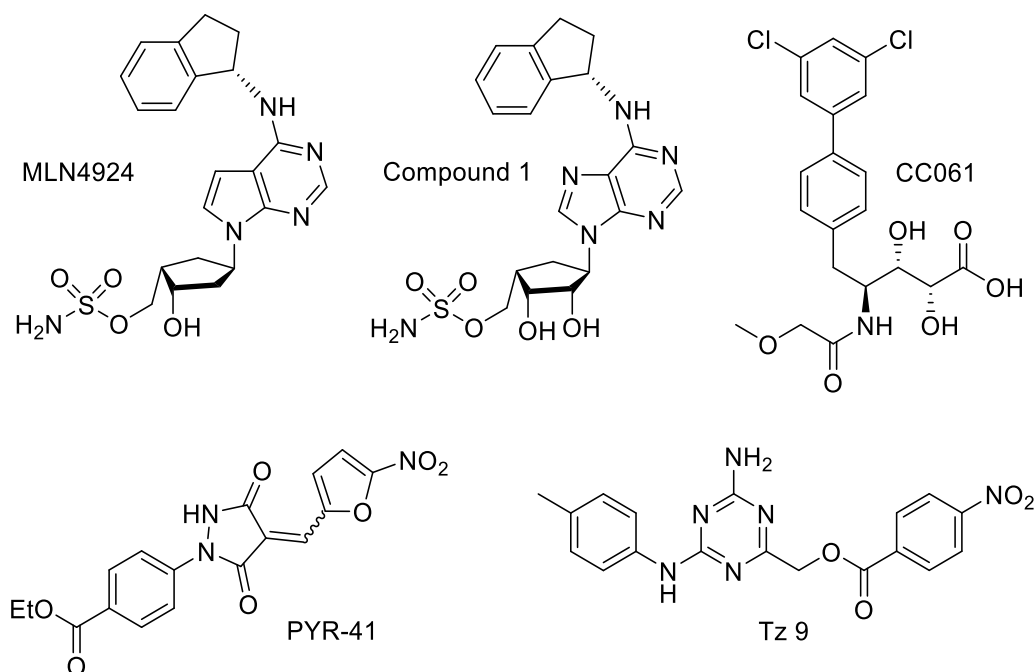
---

In contrast to Chapter 2, this Chapter reports the design, synthesis, and development of ‘clickable’ small molecule derived activity-based probes and their subsequent application in quantitative SILAC proteomic experiments both in cell and in lysate. Two different inhibitor scaffolds are explored, and further compared to alternative electrophilic probes from other projects within the Tate group.

## 3.1 Introduction

### 3.1.1 Small molecule inhibitor leads for probe development

As discussed in Chapter 2, peptide-based ABPs have proven unsuccessful tools to study cellular ubiquitination in our hands. It would be beneficial to develop small molecule ABPs, which are synthetically tractable and translatable into cells. There are numerous small molecule inhibitors reported for E1, E2, E3 and DUB enzyme classes, which provide a starting point for ABP development. Covalent, active site binding inhibitors provide an ideal starting point for probe development as they possess an intrinsic electrophilic trap. Conversion into an ABP, therefore, only requires the addition of a clickable tag. Several DUB and E1 acting covalent inhibitors have been reported in the literature. MLN4924 (Figure 28) is a neddylation selective E1 inhibitor which acts as an AMP-Ub substrate mimetic, resulting in a covalent interaction with the catalytic cysteine.<sup>208, 265</sup> In 2013 An *et al.* successfully converted 'Compound 1', a pan E1 selective analogue of MLN4924, into an alkyne tag possessing activity based probe and demonstrated its ability to label a number of E1 enzymes.<sup>192</sup> This represents the only literature reported cellular ubiquitination ABP to date.



**Figure 28: Overview of the small molecule inhibitors discussed in this Chapter.**

**PYR-41** is a reported cell permeable covalent inhibitor of the human E1 enzyme, UBA1, with a recombinant  $IC_{50}$  of 10  $\mu$ M (Figure 28).<sup>206</sup> The presence of two covalently reactive groups leads to various possible mechanisms of covalent interaction. Indeed, in addition to UBA1 inhibition reduced activity against NEDD4 (E3 ubiquitin-protein ligase NEDD4), a HECT E3 ligase, and significant

inhibition of DUBs USP5, USP9x, USP14, UCHL3 and UCH37 has been reported *in vitro*.<sup>206, 266</sup> In addition, Kapuria *et al.* observed higher MW adduct formation for the two most potent DUB targets, USP5 and USP9x, in cells treated with **PYR-41**. This suggests that **PYR-41** has the capacity to crosslink multiple enzymes at once. This effect further extends to non-Ub pathway related proteins such as Jak2 (Tyrosine-protein kinase JAK2) and RUNX1 (Runt-related transcription factor 1). Irrespective of this published promiscuity, **PYR-41** is sold commercially as an UBA1 selective probe. It was of interest, therefore, to design and synthesise a **PYR-41** derived probe in order to determine its cellular targets, as well as evaluate its value as a pan ubiquitination probe.

Probe formation through the adaption of non-covalent inhibitors is feasible; however the design and synthetic tractability of trap and alkyne tag incorporation needs to be carefully considered. Although several E3 ligase inhibitors are known, many target protein-protein interactions (PPIs) rather than the active site making them unsuitable starting points for ABP development.<sup>108</sup> Similarly allosteric binders, such as the E2 inhibitor CC061 (Figure 28) are unsuitable starting points for ABP design.<sup>209,</sup><sup>210</sup> **Tz9** is a RAD6B (Ubiquitin-conjugating enzyme E2 B) inhibitor, and member of the first known active site binding E2 inhibitor series, reported by Sanders *et al.* in 2013.<sup>211</sup> This provides the first small molecule starting point for E2, and potentially other ubiquitin enzyme, probe design. **Tz9** and its analogues were identified as inhibitor candidates by *in silico* screening against the known X-ray structure of RAD6B before being functionally characterised through their application in both recombinant and cellular assays. Although **Tz9** is not reported as an irreversible probe, inhibition was measured by evaluating covalent Rad6B-Ub thioester formation. Further, the reported *In silico* docking implies that the ester group of the inhibitor lies close to the active site cysteine residue. Analysis of published structure-activity relationship (SAR) data further showed that nitro-containing members of the series, such as **Tz9**, presented improved binding affinities in comparison to those containing electron-donating benzyl ester moieties, an effect which could be due to increased ester reactivity. As there is literature precedent for esters being attacked by ubiquitin ligases<sup>183</sup>, we hypothesised that **Tz9** could be acting in a similar manner. Consequently, Tz9 probes were designed and synthesised to study whether or not this inhibitor offers a novel probe scaffold to identify ubiquitination machinery.

The development and application of probes for both **PYR-41** and **Tz9** are discussed in the ensuing sections.

## 3.2 PYR-41 probe

### 3.2.1 Design and synthesis

Initial work focussed on the design and synthesis of a **PYR-41** derived ABP, **PYR-41-P** (Figure 29). As **PYR-41** intrinsically possesses an electrophilic Michael acceptor, conversion to a probe molecule only required incorporation of an alkyne tag for subsequent bioorthogonal ligation. The ester position was chosen for alkyne tag insertion, as the available structure activity relationship (SAR) data suggested that this was non-essential for inhibition.<sup>206</sup> A retrosynthetic route to **PYR-41-P** is shown in Figure 30. It was envisioned that **PYR-41-P** could be synthesized in an analogous manner to its parent inhibitor by using an alkyne containing analogue (**4**) of intermediate **5**, which is utilised in the reported synthesis of **PYR-41**.

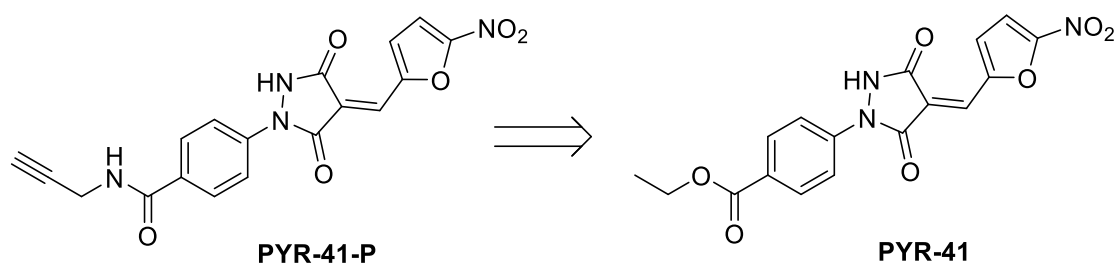


Figure 29: Design of **PYR-41-P** from parent inhibitor **PYR-41**.

The synthetic development of **PYR-41-P** is outlined in Figure 31. At first, a patented synthetic procedure was adopted.<sup>267</sup> Formation of **7** from **6** was achieved in 73 % yield by refluxing in ethanol in the presence of thionyl chloride, followed by salt neutralisation using  $\text{Na}_2\text{CO}_3$ .<sup>268</sup> Further reaction with ethyl malonyl chloride afforded intermediate **8** which, due to instability to silica column and LC-MS conditions, was carried forward without purification. Unfortunately subsequent cyclisation under basic conditions proved unfruitful, therefore, direct formation of the cyclised intermediate **5** from **7** with malonyl chloride in the presence of DMAP was attempted with success, affording **5** in 94 % yield.<sup>269</sup>

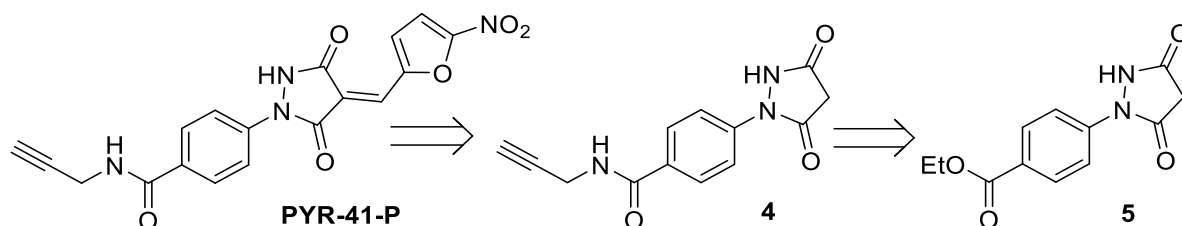


Figure 30: A retrosynthetic approach to **PYR-41-P** synthesis.

Incorporation of the alkyne tag was achieved by ester hydrolysis of **5**, followed by amide bond formation with propargyl amine, EDCI and HOBt to afford intermediate **4**. Due to insufficient solubility of intermediate **9** for reaction in DCM or MeCN, DMF was required for the reaction to proceed. Although formation of **4** was identified by LC-MS analysis, subsequent removal of DMF through aqueous washing was highly inefficient with significant transfer of **4** into the aqueous wash. Consequently, DMF was removed *in vacuo* at 60 °C. Due to poor product migration in organic phase, purification by flash column chromatography proved an unsuitable method of purification, whereas a good trace could be seen by LC-MS analysis. Accordingly, it was decided that **4** was best isolated by preparative LCMS. As there is no risk of racemisation, HOBt is strictly not required for the peptide coupling reaction to occur. Furthermore, HOBt side products are undesirable in this instance, as they may precipitate under the acidic conditions employed during LC-MS purification. Hence, the reaction was repeated in the absence of HOBt and the crude residue after DMF removal successfully purified in 9 % yield. Although this yield is still significantly lower than hoped, it was decided that no further synthetic optimisation would be investigated prior to probe suitability having been determined.

Finally, **PYR-41-P** was formed by Knoevenagel condensation between **4** and 5-nitro-2-furaldehyde. Initially the reaction was achieved in 24 % yield by reflux in ethanol overnight, which was improved to 69 % by refluxing under microwave in MeOH for 20 minutes. **PYR-41** was formed in an analogous manner, by Knoevenagel condensation between **5** and 5-nitro-2-furaldehyde under microwave conditions in MeOH for 20 minutes, in 91 % yield. Both compounds were formed in sufficient yield and purity to undertake biological evaluation.

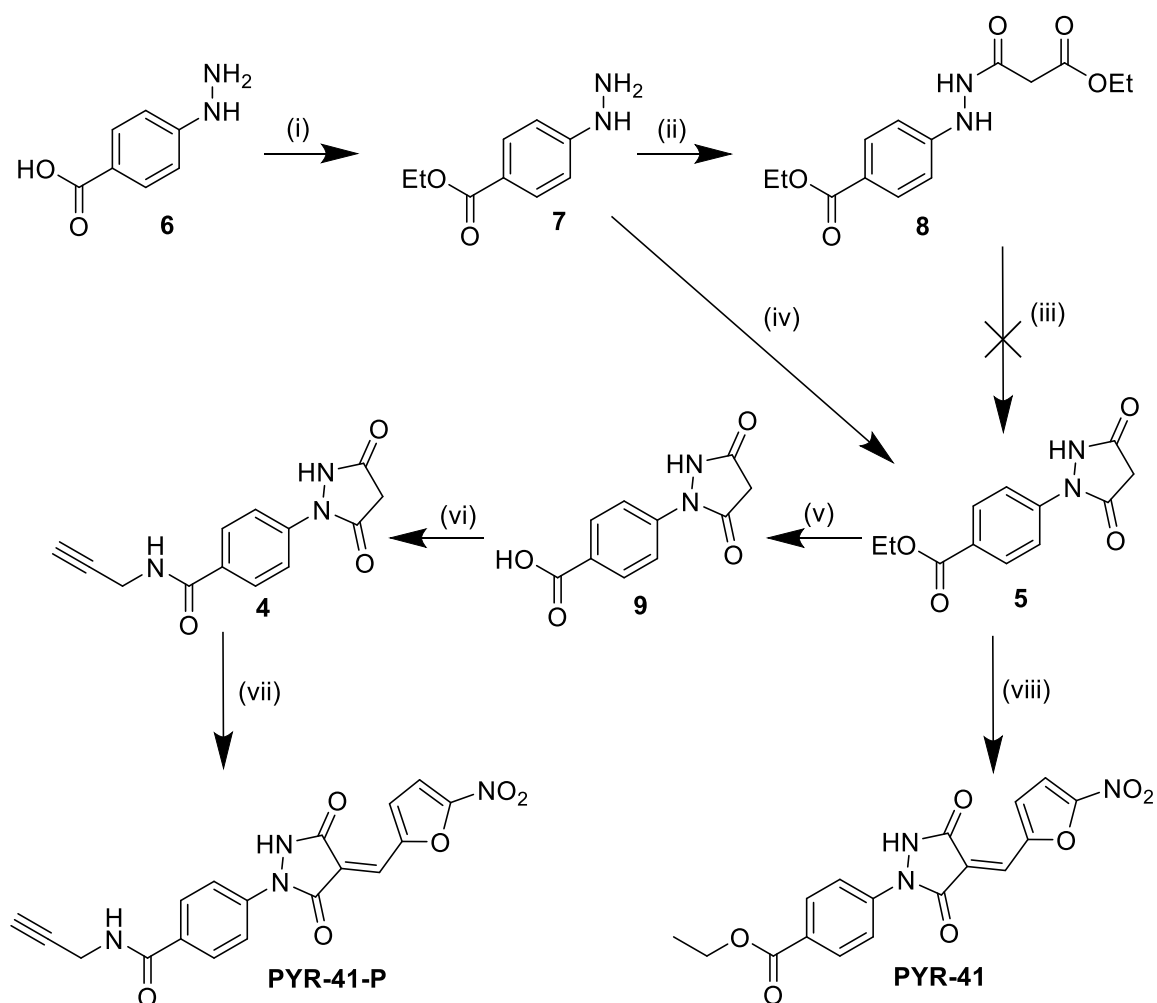


Figure 31: PYR-41 probe synthetic development; i)  $\text{SOCl}_2$ , EtOH, Reflux 16 h, 73 %. Salt neutralisation was achieved using 10 %  $\text{Na}_2\text{CO}_3/\text{DCM}$ , RT, 15 min; ii) ethyl malonyl chloride,  $\text{NEt}_3$ , THF,  $-10^\circ\text{C}$  to RT, 16 h (crude); iii) 1 M NaOH / EtOH RT, 30 min; iv) malonyl chloride, DMAP, DCM, RT, 1 h, 94 %; v) 1 M NaOH / MeOH, 1 h, RT, 86 %; vi) EDCI-HCl, DIPEA, propargyl amine, DMF, RT, 16 h, 9 %; vii) 5-nitro-2-furaldehyde, MeOH, reflux under microwave, 20 min, 69 %; viii) 5-nitro-2-furaldehyde, MeOH, reflux under microwave, 20 min, 91 %.

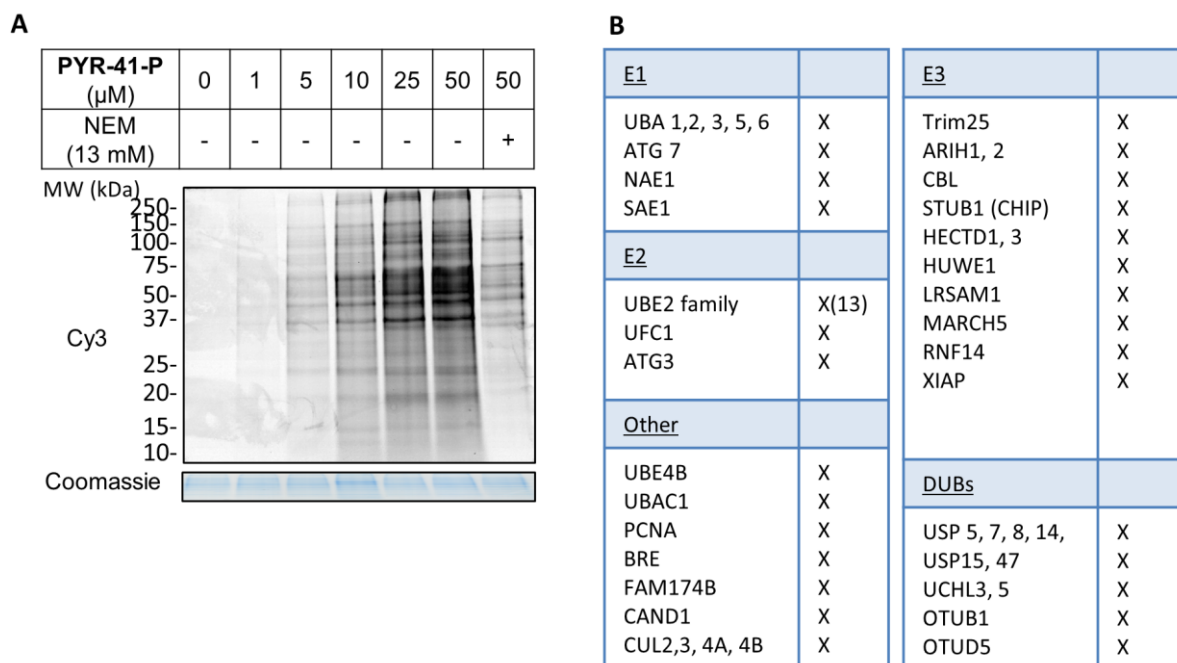
### 3.2.2 In lysate application and initial proteomic evaluation

With **PYR-41-P** in hand, work moved forward to test its ability to label enzymes in cell extract. Akin to Chapter 2, HeLa cells were used for preliminary tests.

Firstly, HeLa cell lysate was incubated with varying concentrations of **PYR-41-P** for 1 hour at  $37^\circ\text{C}$ . Pre-incubation with 13 mM NEM, for 20 minutes was used as a control for competitive labelling. The probe-enzyme complexes formed were subsequently ligated to Az-TB, separated by SDS-PAGE and visualised by in-gel fluorescence as described in Section 2.3 (Figure 32A). Strong, concentration dependent labelling in cell lysate was observed, both at and below  $50\ \mu\text{M}$ : a concentration which is used routinely for studies with **PYR-41** in the literature.<sup>206, 266</sup> A depletion of signal was observed in the NEM control lane, suggesting that labelling is dependent in the availability of reactive cysteines,



and thus can be described as activity dependent. Clearly, **PYR-41-P** is labelling several targets in cell lysate, beyond its expected UBA1 and DUB targets. The presence of high molecular weight labelling (>250 kDa) is also in agreement with the ability of **PYR-41** to cross link proteins.<sup>266</sup>



**Figure 32: Preliminary investigation of PYR-41-P labelling on HeLa cell lysate (A)** Cell lysates were pre-treated with NEM or DMSO for 20 min before incubation with probe for 1 hr, followed by CuAAC ligation to AzTB, separation by SDS-PAGE, and visualisation by in-gel fluorescence (Cy3). Molecular weight (MW) markers are indicated, and Coomassie staining provides protein loading controls. **(B)** Table summarising the Ub machinery identified by LC-MS/MS analysis of PYR-41-P labelling (50  $\mu$ M) .

In order to identify the targets of the **PYR-41-P**, preliminary proteomic analysis was undertaken. Briefly, HeLa lysate was incubated with 50  $\mu$ M of **PYR-41-P** for 1 hour at 37 °C. The probe-enzyme complexes formed were subsequently ligated to Az-TB, enriched, reduced, alkylated, trypsin digested, desalted, separated, and analysed by LC-MS/MS as described in Section 2.3. Identification by at least 2 ‘unique+razor’ peptides with an FDR of 0.01 was set as a minimum to define a probe target. A control sample, lacking **PYR-41-P**, was also prepared in parallel.

On analysis of the control sample, 107 proteins were identified as background non-specific binders, which were subtracted from the **PYR-41-P** intensities. UBA1, the reported target of **PYR-41**, was identified in both samples though at a higher intensity in the probe sample. Treated at its literature concentration of 50  $\mu$ M, 2038 **PYR-41-P** protein targets were identified. Of these hits 54 ubiquitination enzymes were detected, including UBA1, and several E2, E3 and DUBs, which in total represented 2.6 % of labelling (Figure 32B). These initial in-lysate proteomic results were a promising

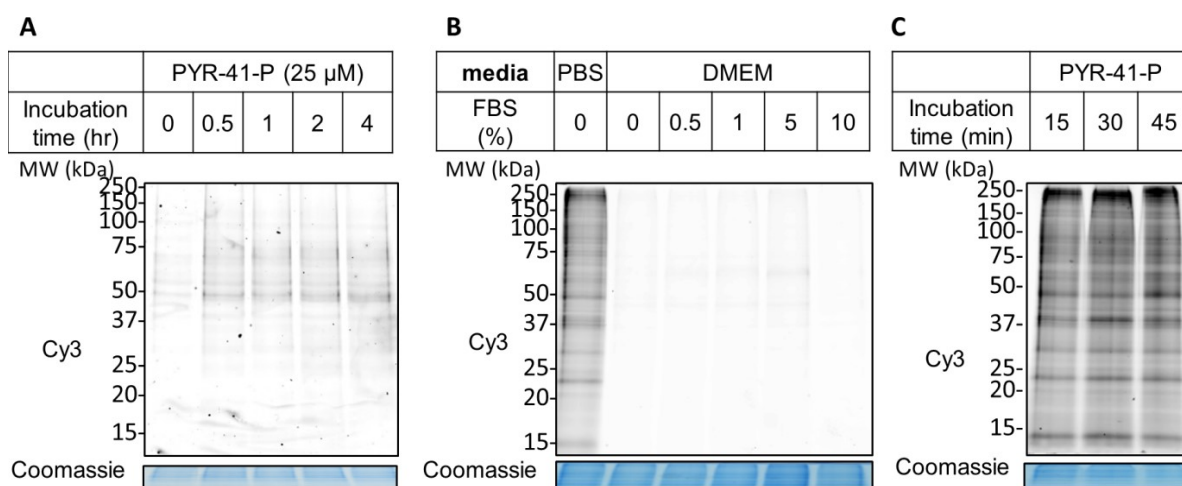
indicator for **PYR-41-P** as a pan Ub probe; although the probe clearly demonstrates promiscuous binding, this may be acceptable if its competitive hits are identifiable.

### 3.2.3 Optimisation of in-cell treatment

With promising results from in-lysate analysis, work commenced to determine the cellular labelling of **PYR-41-P**. Based on the strength of in-lysate labelling, 25  $\mu\text{M}$  was chosen as a suitable concentration to test cellular uptake of the probe. HeLa cells were incubated with 25  $\mu\text{M}$  of **PYR-41-P** for 0 to 4 hours prior to cell lysis. Lysates were then ligated to Az-TB, separated and visualised (Figure 33A). In comparison to in-lysate labelling, in-cell labelling was much weaker, with observed bands poorly defined against the fluorescent background and not significantly affected by incubation time.

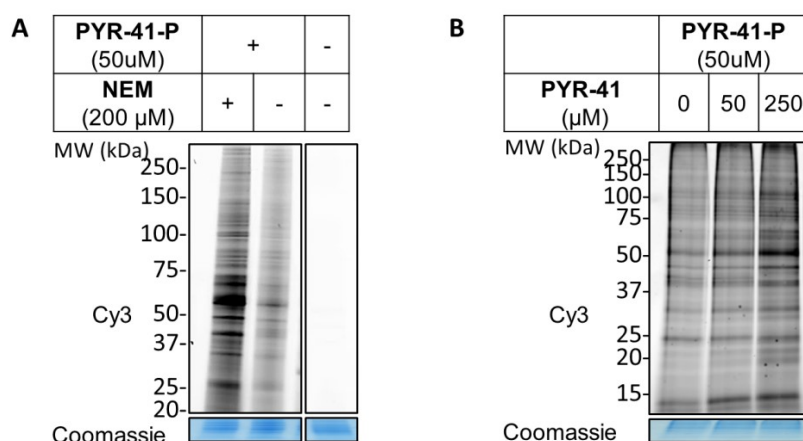
Numerous factors could be contributing to the observed loss of labelling in cells, in particular: i) low cell permeability of **PYR-41-P**; ii) quenched probe reactivity in the growth media; iii) quenched probe reactivity in cell. The calculated clogP values of **PYR-41** and **PYR-41-P** are 1.97 and 0.77 respectively (calculated using imolecular website. <http://www.molinspiration.com/services/logp.html>). Although the clogP of **PYR-41-P** is lower due to its additional H-bond donating amine, its cLogP remains positive, making it unlikely that this is sufficient to significantly hinder cell permeability.

In order to investigate the effect of media on the probe's labelling capability HeLa cells were treated with 50  $\mu\text{M}$  of **PYR-41-P** for 1 hour in PBS and compared to analogous treatment in DMEM (Figure 33B). Strikingly, the labelling intensity was much higher in PBS compared to DMEM. This was regardless of the percentage of FBS supplemented in the DMEM. Though isotonic, over time PBS buffer causes cells to detach from the culture plate, making it unsuitable for extensive application to cells. It is possible that PBS affects cellular membrane properties and contributes to the improved labelling through improved cell permeability. Alternatively, a constituent of DMEM is interfering with the probe. A time course indicated that labelling occurs very quickly, with no significant increase in protein labelling after 15 minutes (Figure 33C), therefore 15 minutes was chosen at a time point to further study **PYR-41-P** cellular labelling when treated in PBS. As there is literature precedent for compound treatment in PBS further tests were conducted to scope these conditions for competition experiments.<sup>270</sup>



**Figure 33: Probe application in intact cells.** HeLa cells were incubated with probe, lysed, and labelling analysed by in-gel fluorescence (A) time course of PYR-41-P labelling treated in DMEM media. (B) Comparative study of feeding media and FBS percentage on probe labelling. (C) Time course of PYR-41-P labelling treated in PBS.

The activity of **PYR-41-P** in cells was tested by pre-incubating cells with NEM in PBS for 20 minutes prior to a 15 minute treatment with **PYR-41-P** at 50  $\mu\text{M}$  (Figure 34A). Intriguingly, as opposed to in lysate studies, labelling increased after pre-incubation with NEM. This could be due to: i) NEM quenching redox active thiols in the cell; ii) NEM activating biological pathways which are subsequently labelled by **PYR-41-P**. Due to the possibility that NEM could be activating pathways in the cell, competition against the parent compound, **PYR-41** was attempted as an alternative indication of cellular activity. Cells were incubated with an excess of **PYR-41** for 15 minutes, followed **PYR-41-P** incubation at 50  $\mu\text{M}$  for 15 minutes (Figure 34B). A similar, though less striking increase in labelling was observed as the excess of **PYR-41** increased. Though unexpected, this phenomenon could be explained if the moiety shared by **PYR-41-P** and **PYR-41** induces and targets the same proteins. If this is the case, pre-incubation of inhibitor would induce targets which are subsequently labelled by the probe species. Alternatively, pre-incubation with either **PYR-41** or NEM quenches free thiols in the media or cells, effectively 'shielding' probe activity.

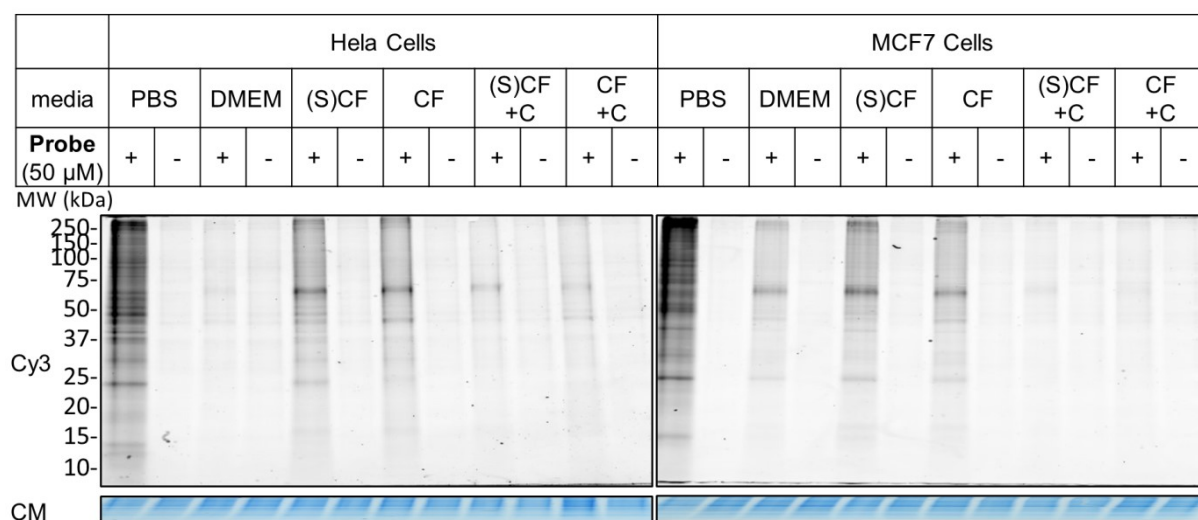


**Figure 34: In cell PYR-41-P competition studies. (A)** HeLa cells were pre-treated with NEM or DMSO for 20 min, incubated with probe for 15 min, lysed, and analysed by in-gel fluorescence. **(B)** HeLa cells were pre-treated with PYR-41 or DMSO for 15 min, incubated with probe for 15 min, lysed, and analysed by in-gel fluorescence **B) Competition series against parent inhibitor, PYR-41.**

When **PYR-41** was first reported by Yang *et al*, quenching of **PYR-41** activity was observed when treated at 50  $\mu\text{M}$  against recombinant UBA1 in the presence of 1 and 10 mM concentrations of glutathione.<sup>206</sup> Although cellular levels of glutathione generally lie in the micromolar range of 0.5-10  $\mu\text{M}$ <sup>271</sup>, it seems reasonable that cellular glutathione is at least in part responsible for the quenching of labelling observed, possibly in combination with other free thiols. With this in mind the constituents of DMEM were considered in detail, on which it was noted that the formulation in use contained cystine. Under reducing conditions, cystine can be reduced to two molecules of cysteine.<sup>272</sup> Amongst its several roles, cystine serves as a specific ligand for the cystine-glutamate antiporter, by which mechanism it is transported into the cell where it is reduced, causing an increase in cellular cysteine levels. It was postulated that using DMEM lacking cystine could improve the cellular activity of **PYR-41-P**, thereby reducing the cellular levels of free thiol.

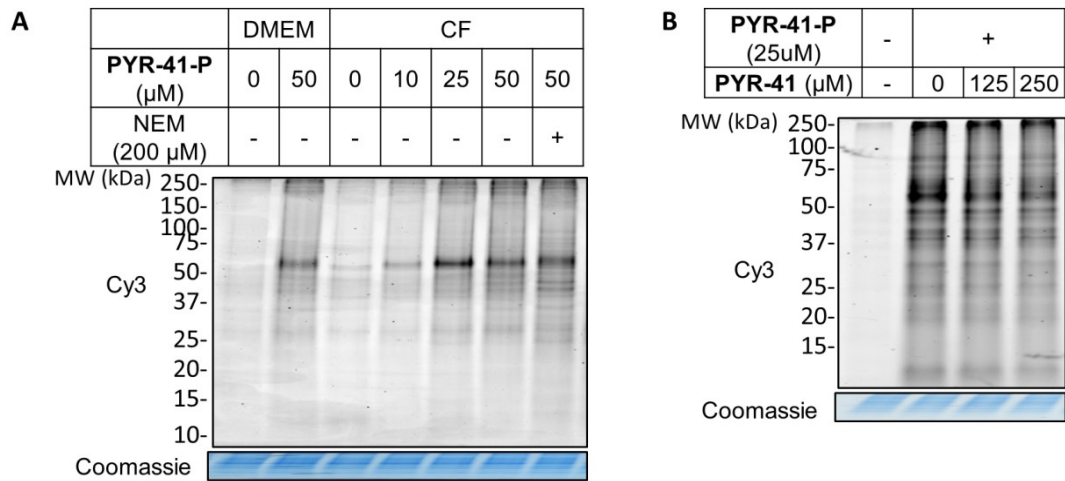
Cells were incubated in either PBS, DMEM or cystine free (CF) DMEM supplemented with 50  $\mu\text{M}$  of **PYR-41-P** or DMSO for 15 minutes. Pre-starving the cells of cystine, by pre-feeding with CF media overnight, was also tested. Control samples, where the HCl salt of cystine was re-added to cystine free media were included, to confirm that any effect observed was due to cystine and not another formulation difference (Figure 35). Although still weaker than that observed with PBS, a marked increase in labelling intensity was observed in HeLa when comparing CF DMEM against DMEM, suggesting that the presence of cystine in the media is adversely affecting probe labelling. The effect of pre-starving the cells on observing this amelioration was minimal, and re-addition of cystine reversed the observation in both cases. Interestingly, expanding the study to include alternative cell lines demonstrated that this effect has some cell line dependence. The labelling pattern observed for HeLa and MCF7 were similar, but interestingly labelling was much more hindered by cystine in the

case of HeLa cells than MCF7 cells. Further testing in HEK293 cells also showed improved labelling on cystine removal (data not shown).



**Figure 35: Comparative study of PBS, DMEM and cystine free (CF) DMEM media on PYR-41-P labelling.** HeLa and MCF7 cells were incubated with probe in different media conditions for 15 min, lysed, and analysed by in-gel fluorescence. Media: PBS = phosphate buffered saline, DMEM = Dulbecco's modified eagle media, CF = cystine free, (S) = cystine starved overnight prior to probe incubation, C = cystine. Addition of cystine to depleted media was used as a negative control. CM = Coomassie.

Further in-cell competition experiments were carried out in MCF7 cells. Labelling of MCF7 cells in CF media by **PYR-41-P** after 30 minutes incubation was shown to be concentration dependent (Figure 36A). Although pre-incubation with NEM did not show loss of labelling, neither did it result in the large increase in labelling observed previously. Competition against **PYR-41** was reassessed: cells were incubated with an excess of **PYR-41** for 30 minutes, followed by **PYR-41-P** incubation at 50  $\mu$ M for 30 minutes (Figure 36B). Although modest, a decreased labelling intensity was observed as **PYR-41** concentration increased; a marked improvement on original attempts to conduct competition analysis in cells. These competition conditions were taken forward for proteomic Spike-in SILAC analysis.

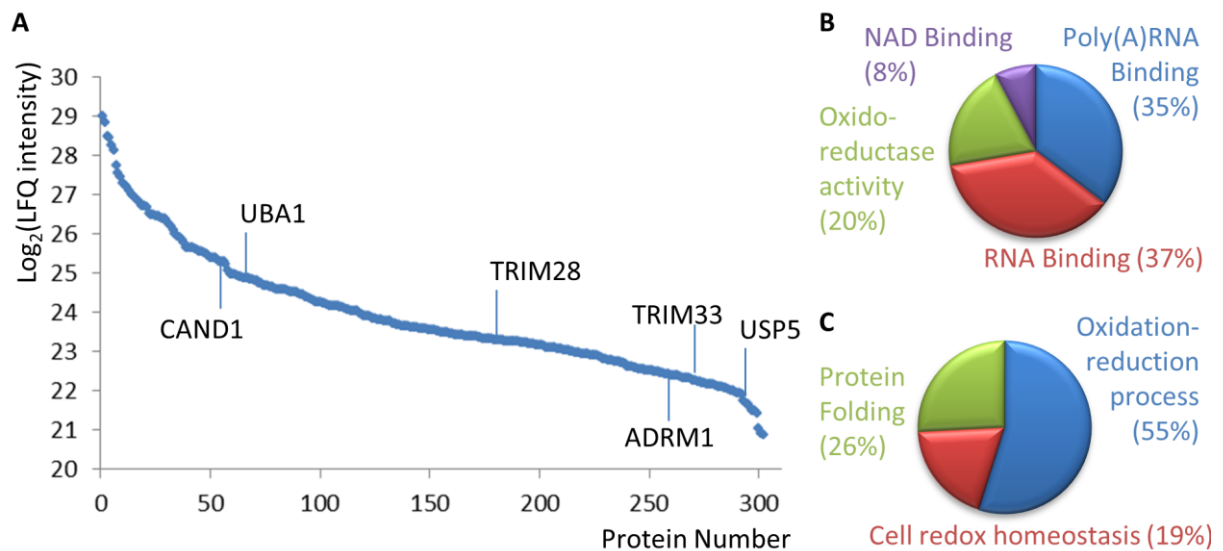


**Figure 36: In cell competition studies in CF media. (A)** MCF7 cells were pre-treated with NEM or DMSO for 30 min, incubated with probe for 30 min, lysed, and analysed by in-gel fluorescence. Probe treatment in DMEM is included for comparison **(B)** MCF7 cells were pre-treated with PYR-41 or DMSO for 30 min, incubated with probe for 30 min, lysed, and analysed by in-gel fluorescence.

### 3.2.3.1 In-cell probe application

Prior to commencing in-cell competition proteomics, the cellular targets of **PYR-41-P** obtained in the absence of any isotopic labelling were analysed. Triplicate plates of MCF7 cells were incubated with 50  $\mu\text{M}$  of **PYR-41-P** or DMSO control for 1 hour at 37 °C prior to cell lysis. The probe-enzyme complexes formed were subsequently ligated to Az-TB, enriched, reduced, alkylated, trypsin-digested, desalted, separated, and analysed by LC-MS/MS. Identification by at least 2 ‘unique+razor’ peptides with an FDR of 0.01 was set as a minimum to define a probe target, and proteins were relatively quantified by LFQ .

On analysis of the DMSO control samples, 101 proteins were identified as background ‘non-specific binders, which were subtracted from the **PYR-41-P** intensities. 302 protein targets of **PYR-41-P** were identified (Figure 37A). Although this included some ubiquitination enzymes, including UBA1, there was a substantial loss of labelling when compared to cell lysate. Ranked by LFQ intensity, GO annotations of the top 50 targets were associated (P-value <0.01) with RNA binding and oxidoreductase activity, and are implicated (P-value < 0.001) in cell redox homeostasis (Figure 37B and Figure 37C). Although the observed decrease in ubiquitination hits was disappointing with respect to the development of a Ub machinery probe for use in intact cells, it was still of interest to determine the observed competition against **PYR-41**.



**Figure 37: Cellular targets of PYR-41-P. (A) Plot of proteins ranked by LFQ intensity, with ubiquitination machinery labelled. GO term (B) Molecular functions (P < 0.01) and (C) biological processes (P < 0.001) associated with the top 50 targets of PYR-m41.**

### 3.2.3.2 In-cell competition against PYR-41

In order to study in-cell competition against **PYR-41**, MCF7 cells were incubated in triplicate with varying concentrations of **PYR-41** (0, 125 and 250  $\mu\text{M}$ ) for 30 minutes at 37 °C before incubation with **PYR-41-P** (25  $\mu\text{M}$ ) for 30 minutes followed by cell lysis. Separately, ‘heavy’ R10K8-labelled MCF7 cells were treated with **PYR-41-P** (50  $\mu\text{M}$ ) for 1 hour at 37 °C before cell lysis to form the ‘spike’. Following protein concentration determination, a fixed amount of spiked lysate was added to the normal lysate in a 1:2 ratio before being ligated to Az-TB, enriched, reduced, alkylated, trypsin-digested, desalted, and separated, and analysed by LC-MS/MS. The raw data was further processed using MaxQuant and Perseus software as described in Section 2.4.1 The results are summarised in Figure 38.

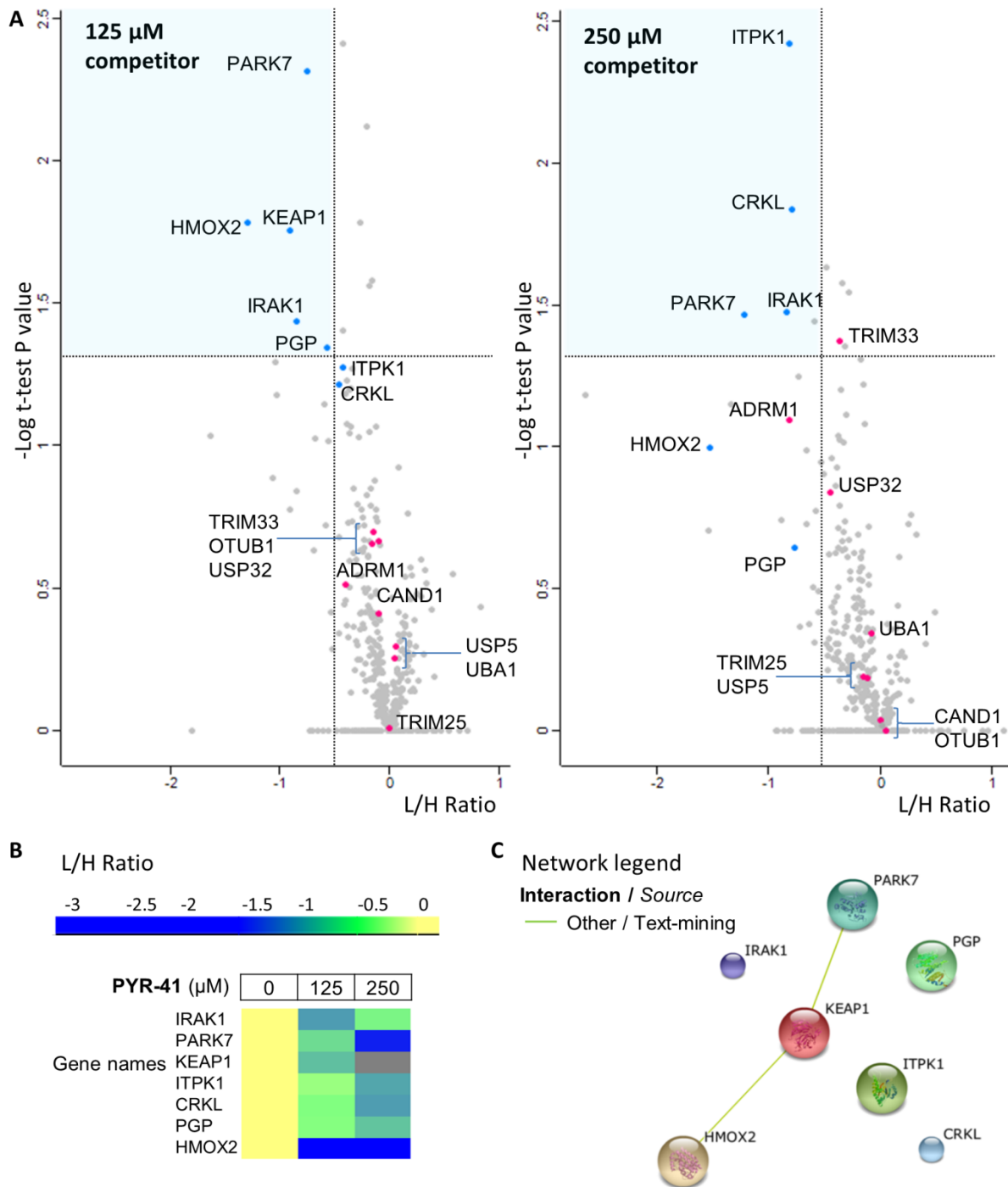
On analysis, 422 proteins were identified with valid ratio of ratios, including all the previously identified Ub relevant enzymes labelled in Figure 37. Applying a threshold ratio change of greater than 1.5 fold (equivalent to a  $\text{log}_2$  fold change of 0.58), 69 proteins were categorised as ‘biologically significant’ competed targets. However, on further analysis by the student T-test (P-value < 0.05) only 7 of these proteins demonstrated statistical significant shifts, none of which represent Ub machinery (Figure 38A).

This result, though disappointing, is not all together unexpected when compared to the in-gel analysis of competition. Although **PYR-41-P** is observed to promiscuously label proteins, it is possible that it is the high promiscuity of **PYR-41** and **PYR-41-P** which makes them difficult to analyse. Of the

7 significantly competed targets, Heme oxygenase 2 (HMOX2) demonstrated the greatest ratio change with competition, which can be seen most clearly with the heat map representation (Figure 38B). HMOX2 is an oxidoreductase that catalyses Heme metabolism through oxidation to form biliverdin, which is subsequently reduced to bilirubin.<sup>273</sup> It has also been linked to cellular response during hypoxia.<sup>274</sup> Protein deglycase DJ-1 (PARK7) showed the next greatest ratio change. PARK7 repairs and releases glycated proteins. Amongst several cellular functions, PARK7 plays an important role as an oxidative stress sensor and redox-sensitive chaperone and protease, protecting the cell from oxidative stress and related pathways of cell death.<sup>275, 276</sup> As the name suggests, PARK7 is associated with Parkinson's disease, with more than 25 PARK7 gene mutations having been linked to the disease.<sup>277</sup> Kelch-like ECH-associated protein 1 (KEAP1) was competed at 5-fold (125  $\mu$ M) inhibitor excess, but was lacking a valid value at 10-fold (250  $\mu$ M) competition. KEAP1 acts as a substrate adaptor protein for the Cullin-3-RBX1 E3 ligase complex.<sup>278</sup> KEAP1 associates with the nuclear factor (erythroid-derived 2)-like 2 (NRF2) transcription factor, sequestering it in the cytoplasm and targeting it for proteosomal degradation. NRF2 regulates expression of several enzymes which maintain cellular homeostasis. As KEAP1 Possesses 27 cysteine residues (9 of which are predicted to be reactive), it can be modified by reaction with numerous electrophiles, leading to dissociation from NRF2, NRF2 nuclear translocation and related gene expression. In this manner, KEAP1 senses the redox status of a cell resulting from electrophilic activity.<sup>279</sup> **PYR-41-P** also competitively labels two kinases: IRAK1 (Interleukin-1 receptor-associated kinase 1) and ITPK1 (Inositol-tetrakisphosphate 1-kinase). As these enzymes do not utilise a catalytic cysteine this further demonstrates the pan reactivity of **PYR-41** and **PYR-41-P**. Crk-like protein (CRKL) and Phosphoglycolate phosphatase (PGP) were also identified as further targets of uncharacterised function. STRING analysis of the 7 hits revealed minimal interaction between the proteins identified (Figure 38C), with only a weak association between HMOX2, PARK7, and KEAP1 identified, which is solely evidenced by text-mining for co-mentioned proteins in PUBmed abstracts.<sup>122</sup>

Although ubiquitination targets were identified in this experiment, none demonstrated both biological *and* statistical significance, though two satisfied the criteria in part. At 10-fold competition proteasomal ubiquitin receptor ADRM1 (ADRM1) did show biologically significant competition. ADRM1 acts an ubiquitin receptor, recruiting the DUB Ubiquitin carboxyl-terminal hydrolase isozyme L5 (UCHL5) to the proteasome and promoting its activity. TRIM33 showed statistically significant competition under the same experimental conditions. Due to a lack of Ub- related competitive probe targets, DDR studies are not reported.





**Figure 38: Spike-in SILAC competition analysis of PYR-41-P.** Samples were prepared in triplicate and spike added in a 1:2 ratio. (A) Volcano plots of PYR-41-P (25  $\mu$ M) labelling in competition with indicated concentrations 12-mer-VME. Dashed lines represent statistical (vertical) and biological (horizontal) significance thresholds. Proteins in the shaded region meet both criteria. Proteins which show significant competition under either condition (blue) and all Ub machinery (Pink, labelled) are highlighted. (B) Heat map of the 7 significantly competed targets, with L/H ratio at each inhibitor concentration indicated. Grey denotes a missing value. (C) STRING network evaluation. Lines in the STRING evaluation represent evidenced interactions between proteins, with evidence source indicated in the key.

### 3.2.4 Summary

**PYR-41-P**, a probe analogue of **PYR-41**, has been designed, synthesised, and shown to label ubiquitination machinery in cell lysate. For in-cell applications, however, low labelling efficiency and potential probe quenching was observed, which was improved through the use of cystine free DMEM media. In-cell competition experiments showed highly promiscuous labelling, which poorly engaged ubiquitination machinery, and showed limited competition against **PYR-41**. In summary, although early promise was seen from in-lysate applications, **PYR-41** has proven unsuitable as a cellular ABP molecule for studying Ub machinery.

### 3.3 Tz9 probe series (T probe series)

#### 3.3.1 Design and synthesis

As the ester group of **Tz9** can be attacked either directly or at the  $\beta$  position, probes with alkyne tags either side of the ester bond were designed (Figure 39). However, with several interactions predicted by *in silico* docking, it was difficult to predict suitable positions for tag incorporation. As the published SARs data implied the methyl group on the benzyl ring minimally effects inhibition compared to proton, this position was chosen as a point of alkyne insertion to probe  $\beta$  attack on the probe (**T1**).<sup>211</sup> On the basis of chemical feasibility and starting material availability the tag was added to the benzyl moiety, *ortho*- to the nitro group, to probe direct interactions with the ester group. Ether (**T2**) and amine (**T3**) chain groups were designed, in order to determine the effect, if any, of hydrogen bonding on probe labelling.

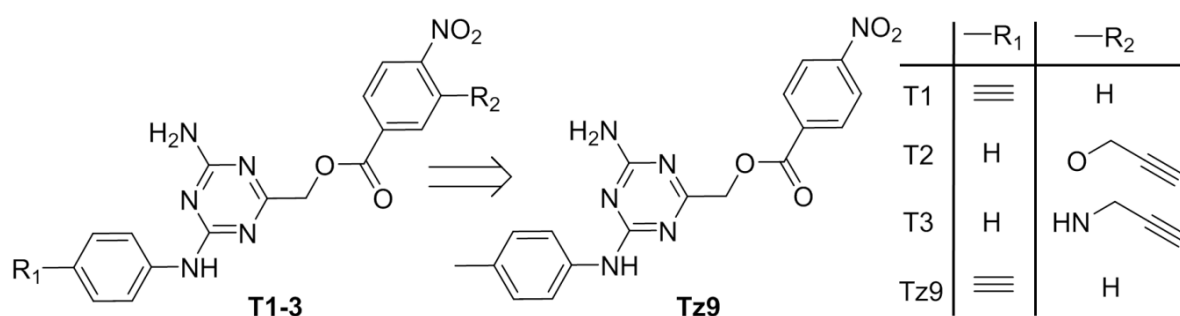
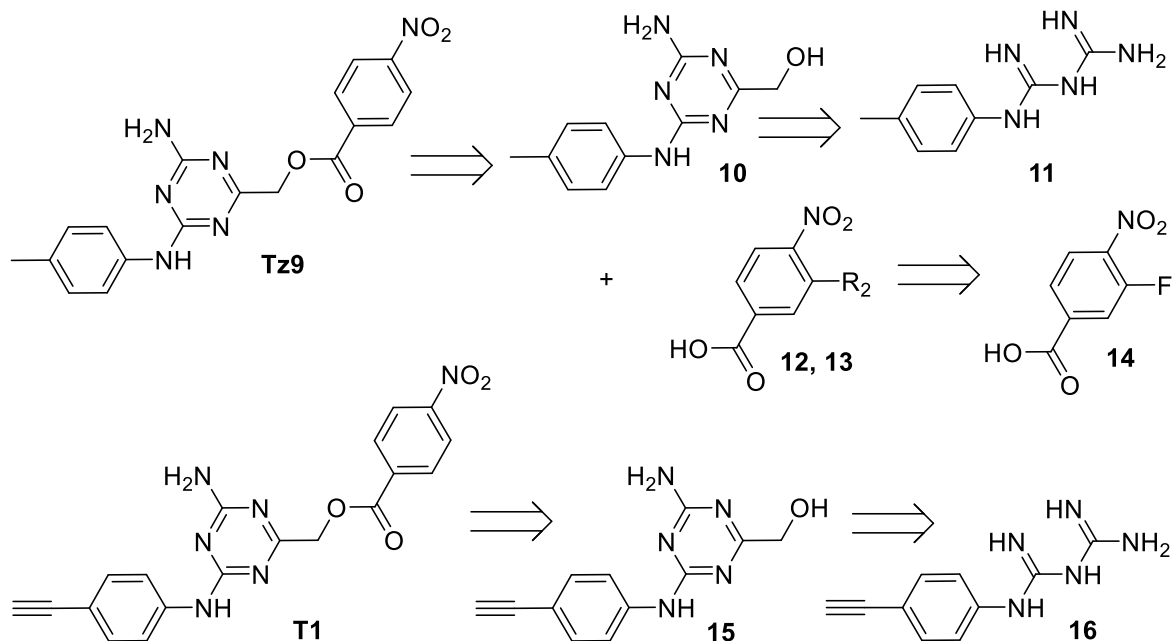


Figure 39: Design of Tz9 derived T series ABPs: **T1**, **T2**, and **T3**.

A retrosynthetic analysis of **Tz9** is shown in Figure 40. It was hypothesised that the probes could be synthesized in an analogous manner. For **T1**, an alkyne containing starting material could be used, whilst for **T2** and **T3** the alkyne moiety could be incorporated onto intermediate **15** using *ipso*-substitution.

The synthetic development of **T1-3** and **Tz9** is outlined in Figure 41. A literature procedure was first adopted to synthesise the parent compound **Tz9**.<sup>211</sup> **11** was formed as an HCl salt from 4-methylaniline (**18**) in 92 % yield by refluxing overnight with dicyandiamide in aqueous HCl. Following salt neutralisation with MeONa in MeOH, refluxing with ethyl glycolate overnight formed **10** in 39 % yield after purification by silica gel chromatography. Subsequent overnight reaction of **10** with 4-nitrobenzoyl chloride in DCM under reflux and purification by silica gel chromatography afforded **Tz9** in 70 % yield.



**Figure 40: Retrosynthetic analysis of Tz9, and postulated incorporation of alkyne tags into probes T1-3.**

**T2** and **T3** were both formed by reacting intermediate **10** with the appropriate nitro-benzoyl chloride derivative: **12** and **13** respectively. Initial attempts to synthesise **13** by reaction of 3-fluoro-4-nitrobenzoic acid (**14**) with propylamine in the presence of triethyl amine in DCM proved unfruitful, most likely due to the poor solubility of **14** in DCM. Repeating the reaction under reflux in H<sub>2</sub>O, however, was successful affording **13** in 72 % yield.<sup>280</sup> **12** was formed using an alternative strategy, by reaction of **14** with propargyl alcohol in the presence of NaHMDS in THF, which proceeded with 44 % yield.<sup>281</sup> The activated acyl chloride derivatives of **12** and **13** were formed by refluxing for 2 hours in toluene in the presence of thionyl chloride.<sup>211</sup> Separate reactions with **10** in the presence of triethyl amine in DCM afforded **T2** and **T3** in 75 % and 64 % yield respectively after silica gel chromatography.

Initial attempts to form **T1** utilised 4-ethynylaniline (**18**) as a starting material, however, refluxing overnight with dicyandiamide in aqueous HCl led to the formation of 4-acteylaniline (**22**, Figure 42) rather than the desired alkyne derivative **16**. On re-assessment of the literature, this outcome is not all that surprising; there are several papers citing such a transformation of the starting material under aqueous acidic conditions.<sup>282</sup> The reaction was reattempted multiple times under microwave conditions in oven dried flasks using dry acetonitrile as an alternative solvent system to minimise the interference of water. Both HCl and glacial acetic acid were trialled as candidate acids, and the resulting precipitates analysed.

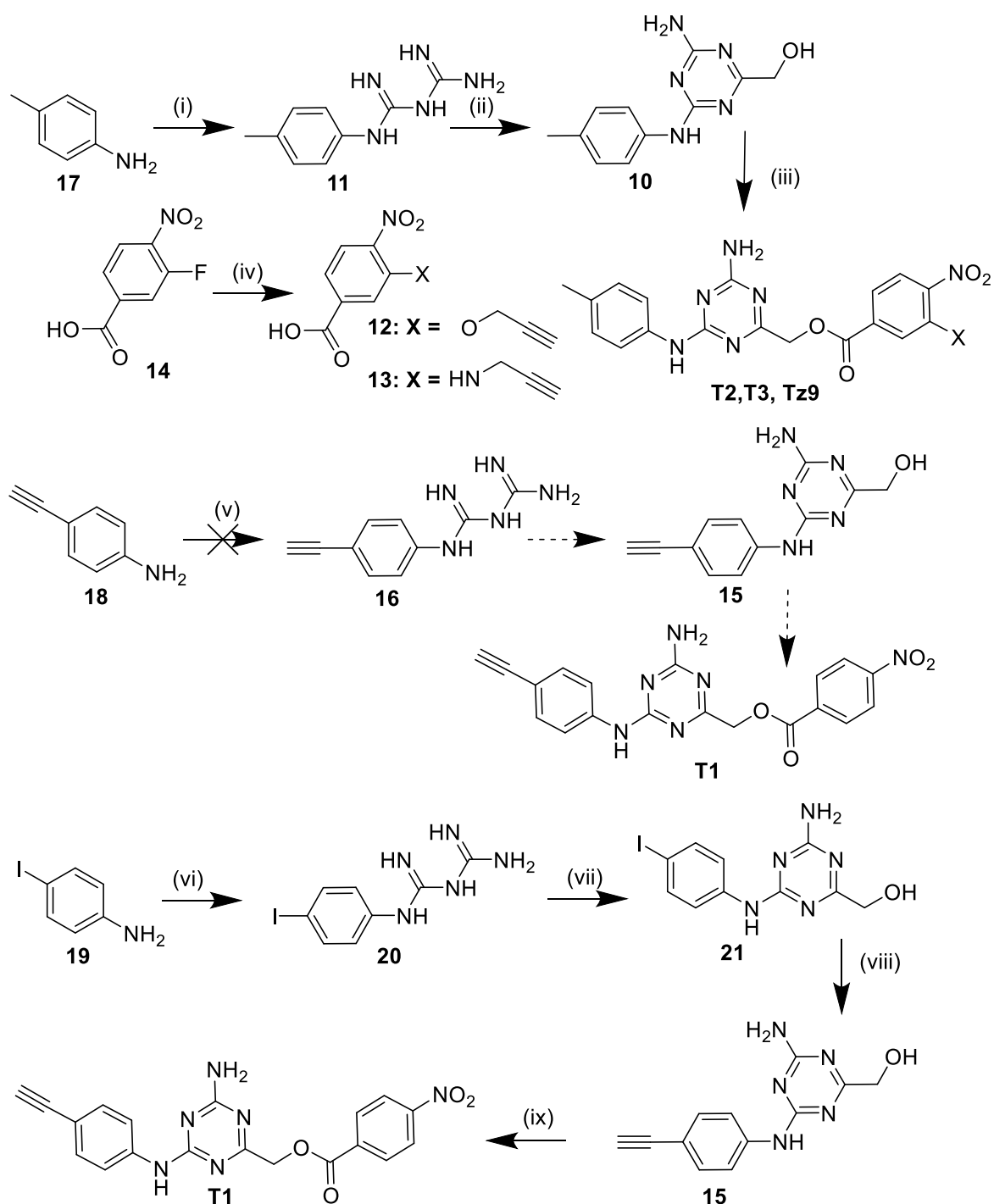
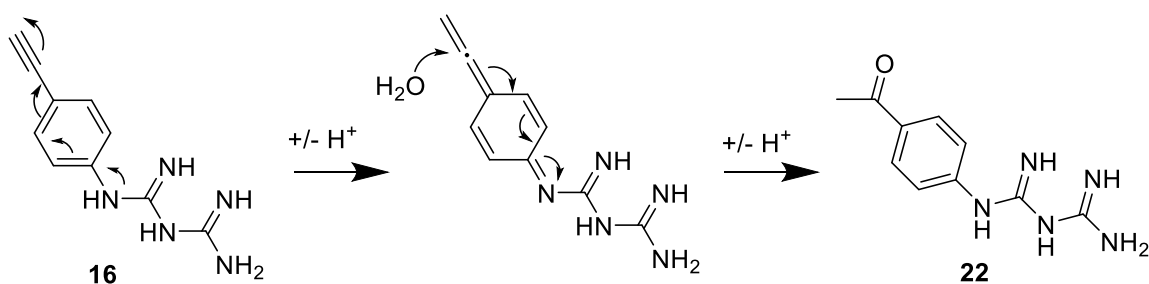


Figure 41: T probe series synthetic development; (i) dicyandiamide, 3 M HCl (aq), 90 °C, 16 h, 92 %. Salt neutralisation was achieved using 37 % MeONa/MeOH, RT, 3 h; (ii) Ethyl glycolate, MeONa/MeOH, reflux, 16 h, 39 %; (iii) **Tz9**: 4-nitro benzoyl chloride, NEt<sub>3</sub>, DCM, reflux, 16 h, 70 %; **T2**: SOCl<sub>2</sub>, THF 75 °C, 2h, then **12**, NEt<sub>3</sub>, DCM reflux, 16 h, 75 %; **T3**: SOCl<sub>2</sub>, THF 75 °C, 2h, then **13**, NEt<sub>3</sub>, DCM reflux, 16 h, 64 %; (iv) **12**: propargyl alcohol, NaHMDS, THF, RT, 16 h, 44 %; **13**: propargyl amine, H<sub>2</sub>O, 85 °C, 16 h, 73 %; (v) dicyandiamide, 3 M HCl (aq), 90 °C, 16 h, no product formation; (vi) dicyandiamide, 3 M HCl (aq), 90 °C, 16 h, 63 %. Salt neutralisation was achieved using 37 % MeONa/MeOH, RT, 3 h; (vii) ethyl glycolate, MeONa /MeOH, reflux, 16 h, 80 %; (viii) bis(triphenylphosphine) palladium (II) dichloride, CuI, NEt<sub>3</sub>, ethynyltrimethylsilane, DMF, 100 °C, 16 h. TMS was subsequently removed using K<sub>2</sub>CO<sub>3</sub> / MeOH, RT, 16 h. overall crude product, 64 %; (ix) 4-nitro benzoyl chloride, NEt<sub>3</sub>, THF, reflux, 16 h, 6.4 %



**Figure 42: Proposed mechanism of ketone formation from compound 16 under acidic conditions.**

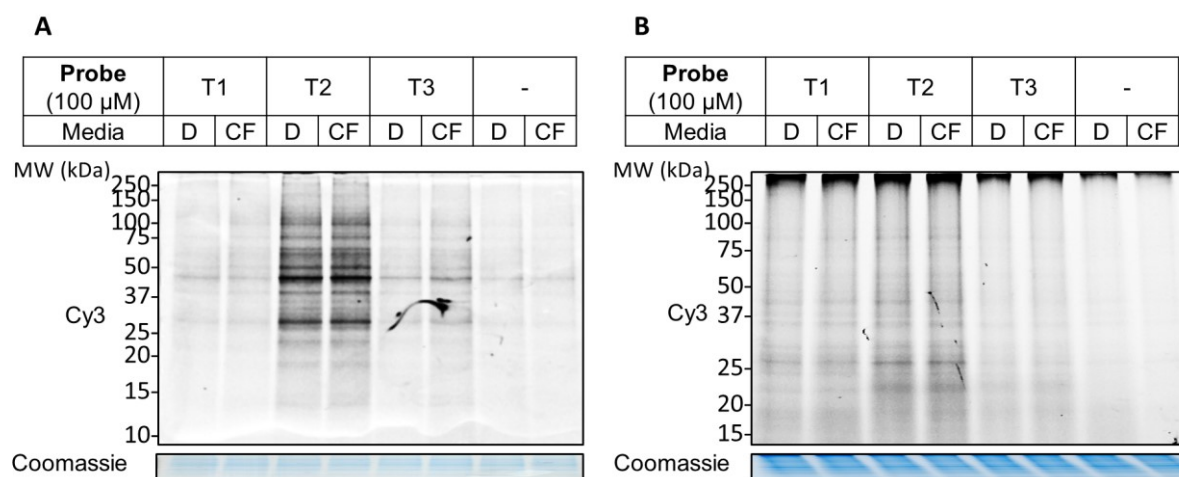
The reaction with HCl was unfruitful, with formation of the acetyl **22** observed as the major product by LCMS analysis. Addition of HCl in dioxane, was also unsuccessful. Glacial acetic acid did lead to some formation of **16**, observable by both LCMS and NMR spectroscopy, however, the product was difficult to solubilise. Milder acidic conditions, using TMS-Cl in acetonitrile in the presence of isopropanol also did not work in our hand.

As there was no literature precedent for the desired reaction preceding under basic conditions this was not attempted. Instead, an alternative synthetic route was devised, utilising 4-iodoaniline (**19**) to form an analogous imidodicarbonimidic diamide intermediate (**20**), thereby allowing the installation of the alkyne moiety at a later stage of the synthesis under non-acidic conditions (Figure 41). **20** was formed as an HCl salt from **19** in 63 % yield by refluxing overnight with dicyandiamide in aqueous HCl. Following salt neutralisation with MeONa in MeOH, refluxing with ethyl glycolate overnight formed **21** as a precipitate in 80 % yield. Due to the poor solubility of **21**, this crude material was carried through. A Sonogashira coupling in hot DMF enabled the installation of TMS alkyne, which on slurrying in  $K_2CO_3$  formed the alkyne derivative (**15**). Subsequent overnight reaction of the crude material with 4-nitrobenzoyl chloride in THF under reflux in the presence of triethylamine and purification by silica gel chromatography afforded **T1** in 6.4 % yield. This low yield is most likely due to the limited solubility of intermediate **21**, leading to the subsequent use of DMF which reduced compound extraction efficiency for small scale reactions. **T1** was also not isolated with optimal purity, with residual 4-nitro benzoic acid observed by NMR. Being non-nucleophilic, it was not expected that the presence of this impurity would impede the ability of the probe to label reactive cysteines. Therefore, as < 2 mg of material had been formed, the ability of the probe to label was first assessed at high concentration, with the intention to optimise **T1** isolation if this probe design demonstrated preferable labelling ability in comparison to **T2** and **T3**.

### 3.3.2 In cell application of T- probe series

#### 3.3.2.1 In-gel fluorescent analysis of T-probe series cellular labelling

Due to the discrepancy observed between cellular and in-lysate labelling by **PYR-41-P**, for analysis of the T probe series it was decided to profile directly in cells. MCF7 cells were incubated with 100  $\mu\text{M}$  of **T1**, **T2**, or **T3** for 1 hour at 37  $^{\circ}\text{C}$  prior to cell lysis and in-gel fluorescence analysis as described in Section 2.3.

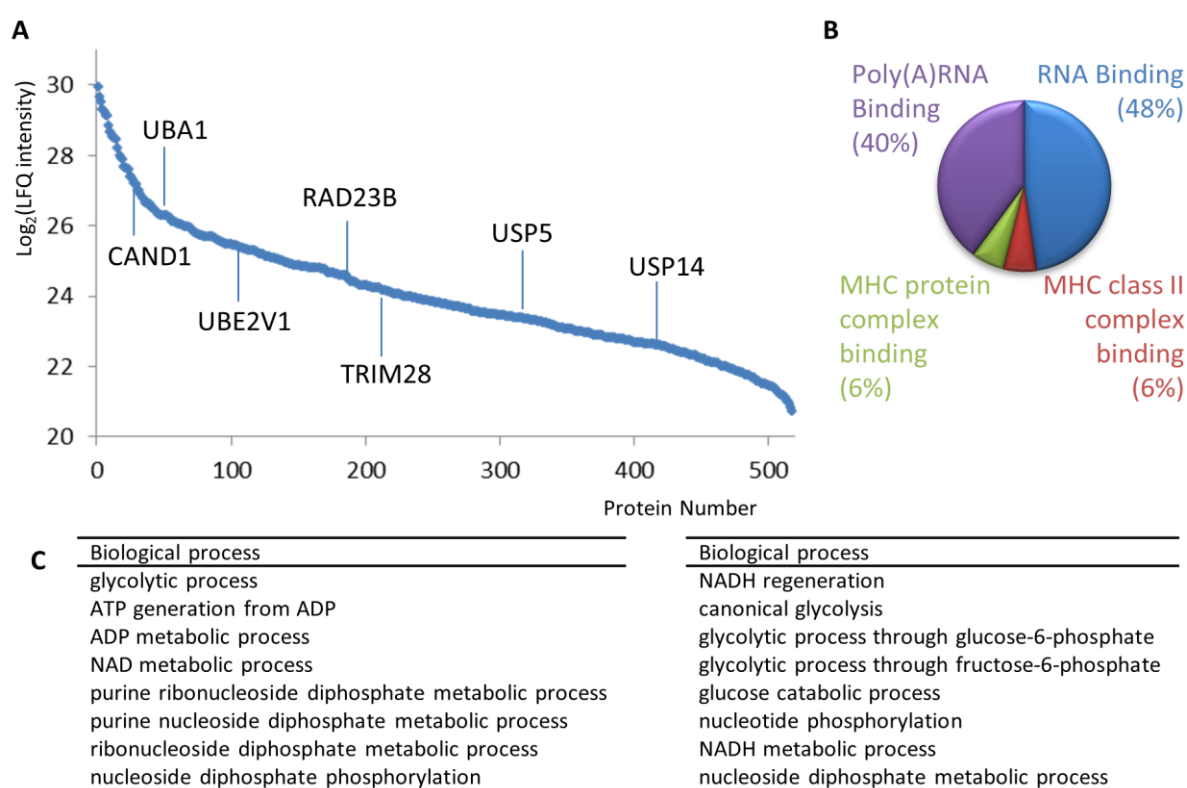


**Figure 43: Preliminary investigation of T probe series. (A)** MCF7 cells incubated with probe (100  $\mu\text{M}$ ) or DMSO in DMEM (D) or cystine free (CF) media for 1 hr, lysed, and analysed by in-gel fluorescence. **(B)** non-reducing gel of the same samples.

It was observed that **T2** labelled strongly at this concentration, whilst **T1** and **T3** showed weak labelling which was poorly defined against the fluorescent background (Figure 43A). The weak labelling observed with **T1** suggests that nucleophilic attack on the probe scaffold is occurring by direct addition to the ester bond. Due to the improved labelling of **PYR-41-P** observed in CF media, this was also tested for the T series, but no significant difference was observed. Whilst **T1** labelling of cysteines would result in the formation of a stable thioether bond, **T2** and **T3** labelling would result in the formation of a labile thioester bond. Reducing and non-reducing gel running conditions were compared, but the relative labelling of **T2** and **T3** was not affected (Figure 43B). This suggests that the lack of labelling observed with **T3** is not due to bond instability under reducing conditions. On the other hand, the lack of effect on **T2** labelling does lead us to question whether **T2** is in fact labelling cysteines. Esters are by no means cysteine selective electrophilic traps; therefore it is not surprising that other nucleophilic acids significantly contribute to the observed labelling. Accordingly, **T2** targets were evaluated by proteomic analysis.

### 3.3.2.2 In-cell probe application

The cellular targets of **T2** (100  $\mu$ M) were identified and quantified by LFQ as previously described in Section 3.2.3.1. On analysis of the DMSO control samples, 126 proteins were identified as background 'non-specific' binders. 24 of these were significantly enriched by **T2** and therefore retained during analysis, leading to 518 identified targets (Figure 44A). Encouragingly, several ubiquitination enzymes were labelled; ranking the protein targets by LFQ intensity, both UBA1 and CAND1 lie within the top 50 targets. On GO term analysis, the top 50 targets were found to be associated ( $P$ -value  $<0.01$ ) with RNA binding, and implicated ( $P < 0.001$ ) in multiple biological pathways that generally relate to glycolysis (Figure 44B and Figure 44C).



**Figure 44: Cellular targets of T2. (A) Plot of proteins ranked by LFQ intensity, with ubiquitination machinery labelled. GO term (B) Molecular functions ( $P < 0.01$ ) and (C) biological processes ( $P < 0.001$ ) associated with the top 50 targets of T2. All listed biological processes are enriched by 6-7 %.**

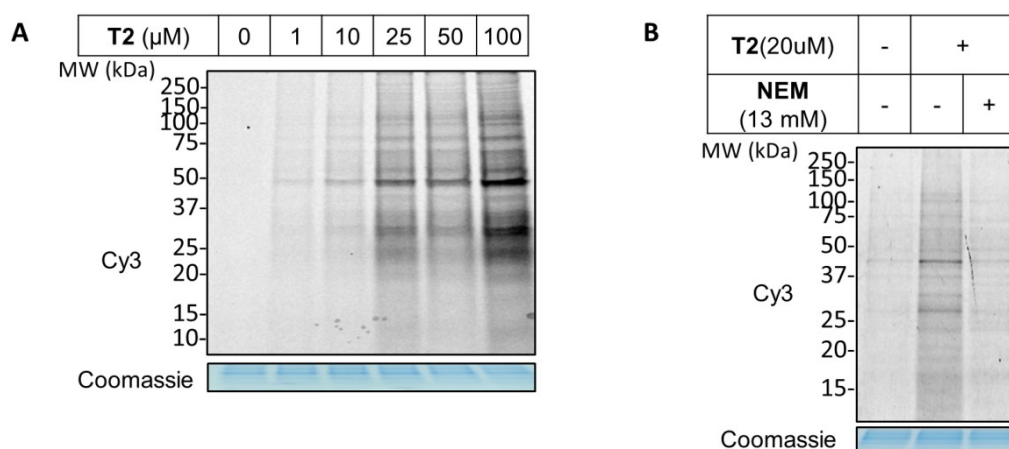
Further ubiquitination machinery identified with **T2** includes two DUBs, an E2, TRIM28, and RAD23B. RAD23B is a multi-ubiquitin chain receptor involved in modulation of proteasome degradation during NER DDR.<sup>283</sup> However, it is not known to possess E2 activity. The expected target of the inhibitor scaffold, RAD6B, was not observed, even though **T2** was applied at 10-fold concentration excess in comparison to literature reported used of **Tz9**. According to the human protein atlas<sup>284</sup> (<http://www.proteinatlas.org/>) RAD6B is expressed at moderate levels in MCF7 cells, therefore, its absence is most likely attributed to i) **T2** is not interacting with RAD6B due to structural change from



**Tz9**; ii) inhibitory action of **Tz9** on RAD6B is not governed by a covalent mechanism. Both hypotheses could be tested through direct incubation of **T2** and **Tz9** with recombinant RAD6B and subsequent ES-MS analysis. However in the interest of time this avenue was not pursued. Though not interacting with RAD6B, **T2** offers a potential probe for several Ub targets and as such warranted further evaluation as a proteomic tool.

### 3.3.2.3 Competition studies with **T2**

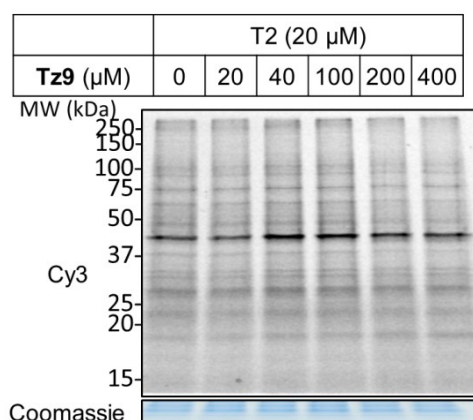
Prior to competition studies, **T2** concentration was optimised to improve the competitive range available for analysis. MCF7 cells were incubated with varying concentrations of probe for 30 minutes at 37 °C prior to cell lysis and in-gel fluorescence analysis. Protein labelling by **T2** was observed to be concentration dependent (Figure 45A). 20 µM was chosen as a sufficient labelling concentration; whilst this concentration ensures widespread protein labelling, it is also low enough to permit competition against high-fold inhibitor excesses. Furthermore, activity-based labelling was confirmed by competing against NEM in cell lysate, using an analogous method to that described in Section 3.2.2. This demonstrates that target labelling is through an electrophilic mechanism that can be inhibited by NEM.



**Figure 45: (A) Concentration series of T2 labelling after 1 hr incubation in MCF7 cells. (B) MCF7 cell lysate was pre-treated with NEM or DMSO for 20 min, incubated with T2 for 1 hr and analysed by in-gel fluorescence.**

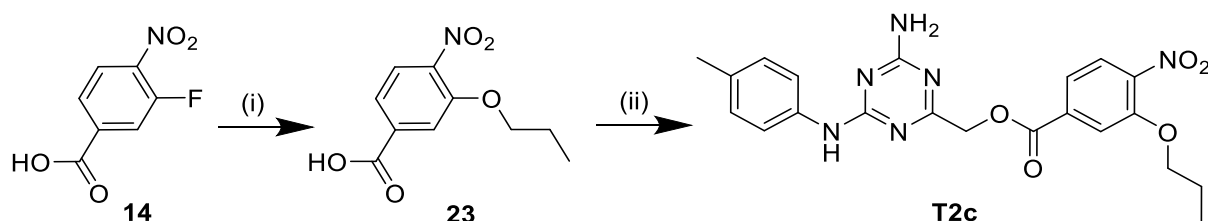
Competition against the parent compound, **Tz9**, was initially assessed by in-gel fluorescence. Cells were incubated with an excess of **Tz9** for 30 minutes, followed by **T2** incubation at 20 µM for 30 minutes and further processing as described in Section 2.3. Disappointingly, minimal competition was observed (Figure 46). It was hypothesised that this lack of competition was a result of the relatively large scaffold modification caused on addition of the alkyne containing group to form **T2**, and that this modification could prevent effective competition against **Tz9**. Accordingly, an

alternative inhibitor containing a propyl-side chain to better mimic **T2**, **T2c** (Figure 47), was synthesised to assist in the validation of the competitive targets of **T2**.



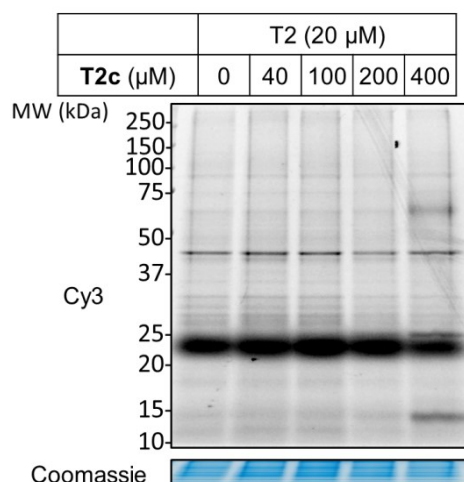
**Figure 46:** In-cell competition studies of **T2** against **Tz9**. MCF7 cells were pre-treated with **Tz9** or DMSO for 30 min, incubated with probe for 30 min, lysed, and analysed by in-gel fluorescence.

**T2c** was synthesised in an analogous manner to **T2**; addition of *n*-propanol to 3-fluoro-4-nitrobenzoic acid (**14**) to form intermediate **23** was achieved in 76 % yield in the presence of NaHMDS in THF (Figure 47). Subsequent treatment with thionyl chloride, and reaction of the resulting acyl chloride with intermediate **10** (Figure 41) in the presence of triethyl amine in DCM afforded **T2c** in 58 % yield after silica gel chromatography.



**Figure 47:** Synthetic development of **T2c**, an alternative inhibitor for **T2**. (i) Propanol, NaHMDS, THF, RT, 16 h, 76 %; (ii) SOCl<sub>2</sub>, THF, 75 °C, 2h, then **10**, NEt<sub>3</sub>, DCM reflux, 16 h, 58 %.

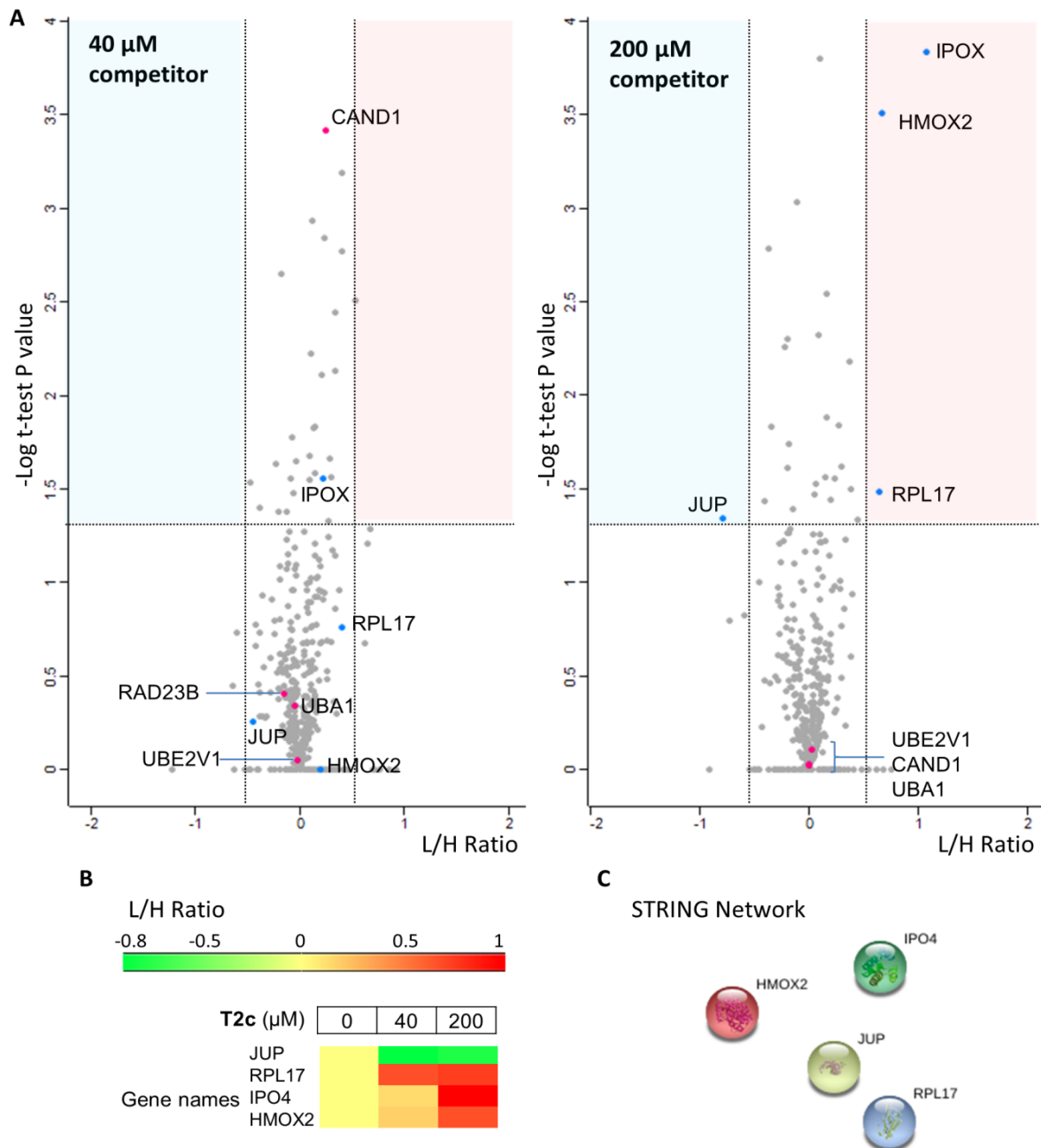
Once in hand, competitive labelling of **T2** against **T2c** was assessed by in-gel fluorescence. Cells were incubated with an excess of **T2c** for 30 minutes, followed by **T2** incubation at 20  $\mu$ M for 30 minutes and further processing as described in Section 2.3. Observed competition remained low, however, with only a slight decrease in **T2** labelling with 10-fold (200  $\mu$ M) competition by **T2c**.



**Figure 48: In-cell competition studies of T2 against T2c. MCF7 cells were pre-treated with T2c or DMSO for 30 min, incubated with probe for 30 min, lysed, and analysed by in-gel fluorescence.**

T2 labelling at 10-fold and 2-fold (40  $\mu$ M) inhibitor excess was further examined in triplicate and quantified by spike-in SILAC. MCF7 cells were incubated in triplicate with varying concentrations of T2c (0, 40 and 200  $\mu$ M) for 30 minutes at 37  $^{\circ}$ C before incubation with T2 (20  $\mu$ M) for 30 minutes followed by cell lysis. Separately, 'heavy' R10K8-labelled MCF7 cells were treated with T2 (20  $\mu$ M) for 30 minutes at 37  $^{\circ}$ C before cell lysis to form the 'spike'. Following protein concentration determination, a fixed amount of spiked lysate was added to the normal lysate in a 1:2 ratio before being ligated to Az-TB, enriched, reduced, alkylated, trypsin-digested, desalted, separated, and analysed by LC-MS/MS. The raw data was further processed using MaxQuant and Perseus software as previously described in Section 2.4.1. The results are summarised in Figure 49.

On analysis, 412 proteins were identified with valid ratio of ratios. As expected from LFQ in-cell analysis, CAND1, UBA1, UBE2V1 and TRIM28 were identified. However, USP5 and USP14 were not identified in this experiment. Though disappointing, these DUBs exhibited low LFQ intensities and as such it is not unlikely that these identifications are lost in the complex analysis of a SILAC experiment. Furthermore, a lower concentration T2 was used.



**Figure 49: Spike-in SILAC competition analysis of T2.** Samples were prepared in triplicate and spike added in a 1:2 ratio. (A) Volcano plots of T2 (20  $\mu$ M) labelling in competition with indicated concentrations T2c. Dashed lines represent statistical (vertical) and biological (horizontal) significance thresholds. Proteins in the shaded regions meet both criteria. Proteins which show significant competition under either condition (blue) and all Ub machinery (Pink, labelled) are highlighted. (B) Heat map of the 4 significantly competed targets, with L/H ratio at each inhibitor concentration indicated. (C) STRING network evaluation. No evidenced interactions between network proteins were identified.

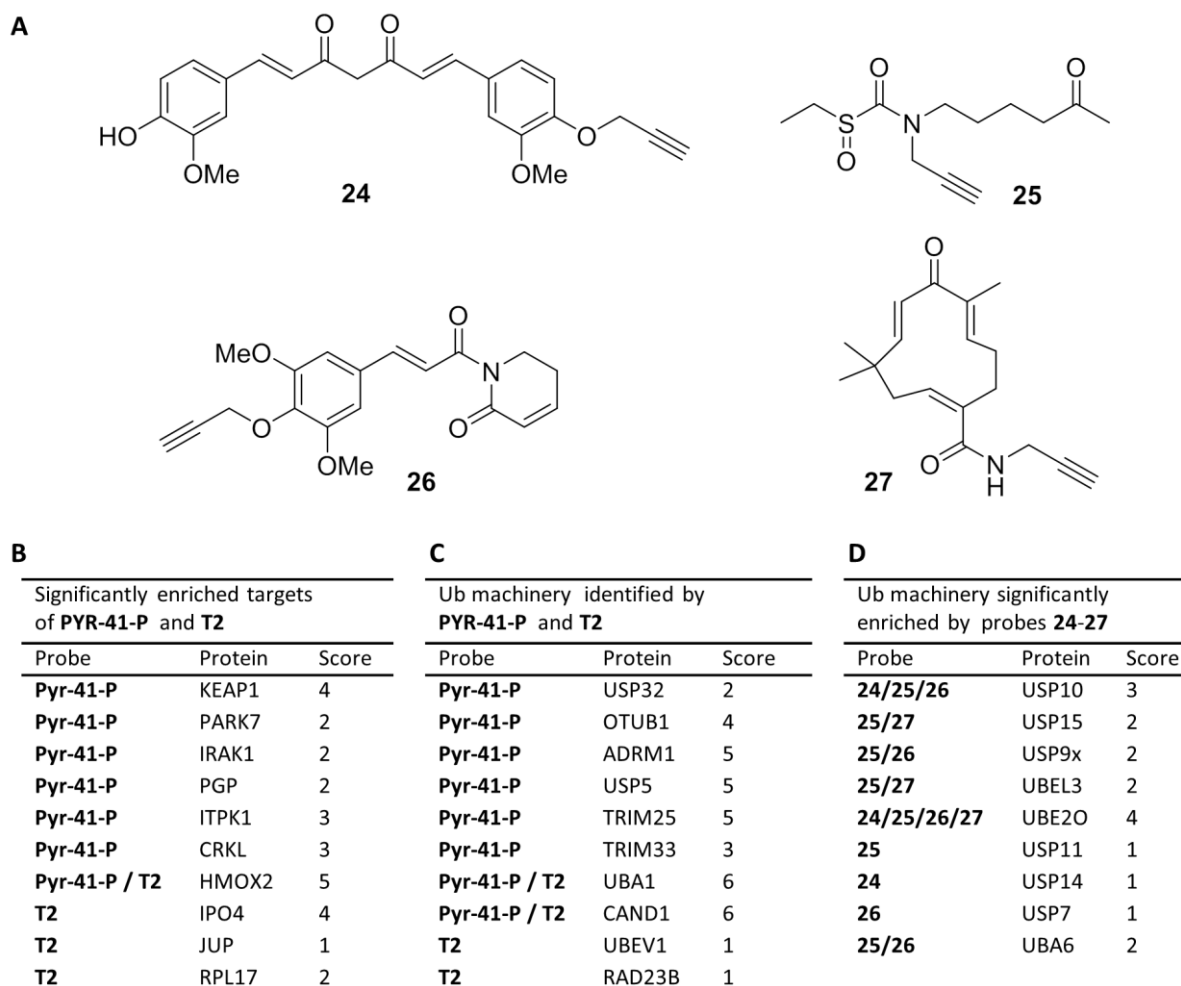
Applying a threshold ratio change of greater than 1.5 fold, 21 proteins were categorised as 'biologically significant' shifters. As with **PYR-41-P**, further analysis by the student T-test ( $P$ -value < 0.05) identified a small subset of these proteins; Importin-4 (IPO4), HMOX2, 60S ribosomal protein L17 (RPL17) and Junction plakoglobin (JUP) that demonstrated statistical significant shifts, none of which represent ubiquitination machinery. Interestingly the majority of significant shifts are the

result of induced, rather than competed labelling of these proteins on competition. It should be noted that, in keeping with in-gel fluorescence analysis, minimal competition is observed as illustrated by the narrow protein distribution in the volcano plot (Figure 49A).

Of the 4 significantly shifted targets, JUP was the only protein competitively labelled. JUP is an architectural plaque protein necessary from cell adhesion, mutations of which have been associated with cardiomyopathy.<sup>285</sup> The other three labelled targets were significantly induced on competition. IPO4 demonstrated the greatest ratio change with competition, which can be seen most clearly with the heat map representation (Figure 49B). IPO4 is a nuclear transport receptor for nuclear localization signals (NLS) in molecular substrates.<sup>286</sup> HMOX2, which has been previously described in Section 3.2.3.2, showed the next greatest ratio. Finally RPL17, as the name suggests, is a 60s ribosomal protein and unlikely to be a meaningful target. STRING analysis of the 4 hits revealed no associated interactions between the proteins identified (Figure 49C). By this analysis, competitive labelling of **T2** cannot be effectively assessed by competition with **Tz9** or **T2c**. Clearly, further work is required to ultimately determine the competitive targets of **T2**. Due to inconclusive competition data, DDR analysis is not reported.

### 3.3.3 Cross-probe comparison

On direct comparison of **PYR-41-P** and **T2** cellular targets, neither significantly (based on competition) label ubiquitination machinery. **PYR-41-P** shows typical electrophilic characteristics, including labelling of redox enzymes, whilst **T2** demonstrated limited electrophilic features. It was of interest to consider whether or not the labelled targets were a direct effect of the probe scaffold, or more generally the result of an electrophilic interaction. Consequently, significantly labelled targets identified by **PYR-41-P** and **T2** were compared to the competition data collected for other electrophilic probe molecules (**24-27**, Figure 50A) of varying scaffolds synthesized in our group.<sup>287, 288</sup> Each significant competition target of **PYR-41-P** and **T2** was scored to reflect its labelling frequency across the panel of six probes. A protein which is labelled by all the probes has a score 6; the higher the score, the less scaffold specific the interaction.



**Figure 50:**(A) Structures of the alternative electrophilic probes used in this study. (B) Comparison of the competitive targets of **PYR-41-P** and **T2** to competitive labelling of these targets by probes **24-27**. (C) Comparison of Ub machinery identified by **PYR-41-P** and **T2** to identification of these targets by probes **24-27**. (D) A selection of the Ub-machinery which are competitive targets of probes **24-27**, but not identified by **PYR-41-P** or **T2**. Competitive targets are defined by competition against the probe's parent inhibitor. The score represents the frequency (1-6) of the protein when compared across the six probes (6=highest frequency).

By this method, HMOX2 was found to be frequently labelled across the probe panel, confirming as expected that this is a non-specific electrophilic target (Figure 50B). Similarly, KEAP1 and IPO4 interacted with multiple probe types. Other targets, such as PARK7 and JUP, were less commonly labelled and therefore more likely to be due to a scaffold-specific interaction. However, whilst this comparison assists the identification of globally reactive targets, it should be noted that these probes were applied in different cell lines of interest, thus preventing the classification of unique targets. All ubiquitination machinery identified by the probe panel, regardless of the significance of probe labelling based on competition, were also compared (Figure 50C). The majority of ubiquitination machinery labelled by **PYR-41-P** and **T2** had a high ( $\geq 4$ ) labelling frequency, with CAND1 and UBA1 being labelled with all the probes, suggesting that probe labelling of these targets is non-specific. Some ubiquitination machinery did, however, appear to be less general targets.

USP32, which is labelled by **PYR-41-P**, showed a low labelling frequency across the panel as did UBEV1 and RAD23b, which are both labelled by **T2**.

Finally the ability of probes **24-27** to competitively label ubiquitination machinery not competed or identified by either **PYR-41-P** or **T2** was analysed, to determine whether these alternative scaffolds offer any improved ability to label ubiquitination machinery. It was found that several ubiquitination machinery were labelled with varying frequency by the panel, including multiple USP - family DUBs. USP10 was highly labelled across the panel, whilst other USPs was less frequently identified. A selection of identified targets is shown in Figure 50D. This suggests that these ubiquitination machinery are not 'global' targets of electrophilic probes.

### 3.4 Conclusions

The work in this Chapter focuses on the design, synthesis, and analysis of ABPs derived from small molecule inhibitors of ubiquitination machinery. Two inhibitor scaffolds were studied, leading to the testing of four related ABPs. **PYR-41-P** showed initial promise, labelling several ubiquitination enzymes by in-lysate application. Unfortunately, this probe engagement was lost on application in cell. In general **PYR-41-P** demonstrated poor in cell labelling efficiency, which was somewhat improved by omitting cystine from the growth media. Furthermore, competition studies with NEM lead to the hypothesis that the probe may be quenched by cellular free thiols. Of the three T series probes, **T2** showed the greatest labelling ability, with several ubiquitination machinery labelled in cells. In general, labelling was not affected by reducing conditions, suggesting that the majority of **T2** targets were not labelled via Cys residues. Quantitative proteomic analysis of **PYR-41-P** and **T2** in competition against their parent inhibitor or inhibitor analogue was disappointing, with **PYR-41-P** not competing any Ub machinery and inconclusive **T2** competition data. **PYR-41-P** did, however, label a number of electrophilic targets, such as KEAP1 and HMOX2. The confidence in these findings is, however, restricted to the non-stringent statistical thresholds applied. Finally, comparison between proteins identified by **PYR-41-P** and **T2** and those identified with alternative electrophilic probes **24-27** was used to identify generic targets of electrophilic probes and to postulate proteins whose probe interaction is more scaffold specific.

Though this work has been unsuccessful in identifying a suitable probe molecule to study ubiquitination machinery in cells, the issues discussed in here illustrate some of the difficulties faced by researchers during probe development, and highlight the importance of identifying suitable inhibitor scaffolds to act as starting points in probe design. With this in mind, Chapter 4 discusses a fragment screening approach to identify novel starting points for E3 ligase inhibitor / probe design.

## **Chapter 4    Identifying potential scaffolds for E3 inhibitors: HOIP as a case study**

---

Fragment based screening has been established as a valid drug discovery approach for enzyme targets.<sup>289</sup> The approach provides a means of identifying novel starting points for inhibitor, or probe, compounds. This Chapter reports the results of a fragment library screen against the minimal catalytic core of HOIP, a RBR E3 ligase. Hits were primarily identified using Differential Scanning Fluorimetry, with further validation by WaterLOGSY NMR spectroscopy. Subsequent characterisation of a number of hits by Micro-Scale Thermophoresis and efforts towards X-ray crystal structure determination are also described.



## 4.1 Introduction

To further our ability to study ubiquitination enzymes it is important to expand the number of probe scaffolds available in the researcher's toolbox. In contrast to Chapter 2 and 3, which focussed on the development of literature inhibitors into probe molecules, the work in this Chapter focuses on the *de novo* identification of molecular starting points, or fragments, (<300 Da) for further development into inhibitor and probe compounds of ubiquitination machinery. HOIP, a RBR E3 ligase was studied using a fragment based strategy, which is further described below.

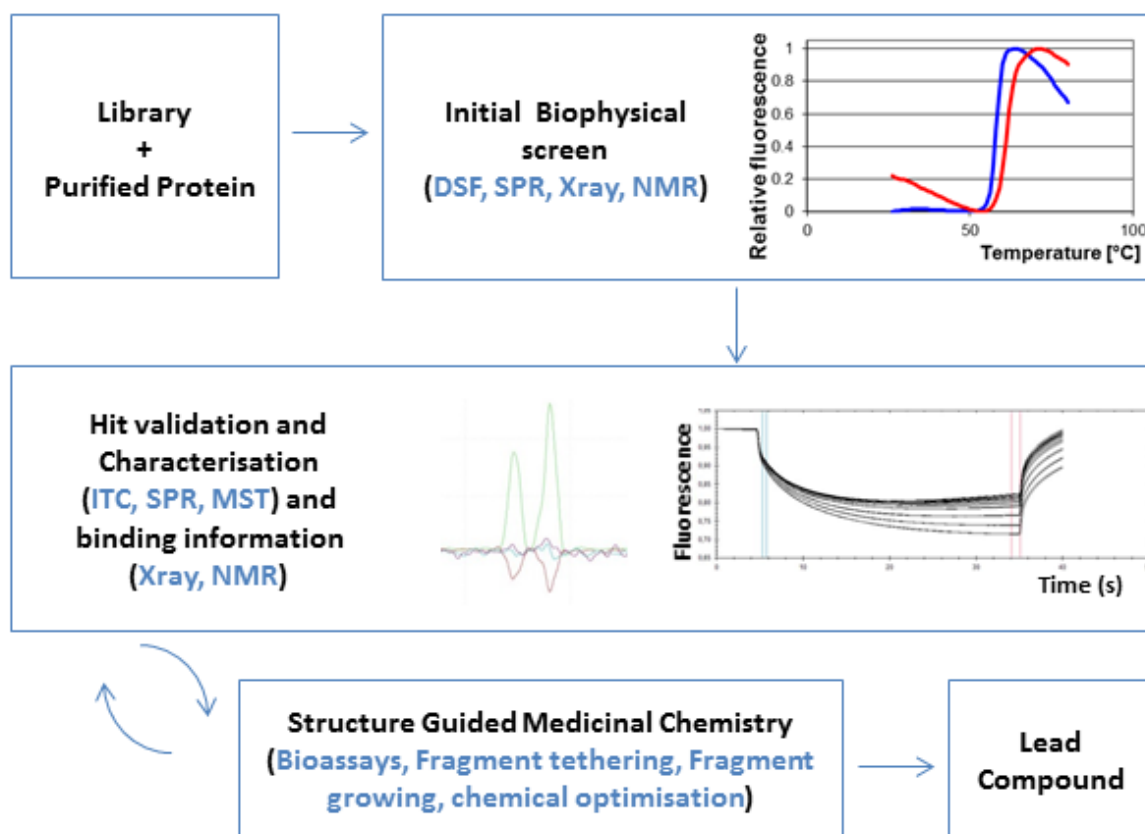
### 4.1.1 Fragment-based drug discovery (FBDD)

The workflow utilised in a typical fragment-based screen is shown in Figure 51. Briefly, a purified recombinant protein of interest is screened against a library of molecular fragments using a high throughput biophysical assay. Any hits identified from this primary screen are further validated and characterised using complementary techniques. Knowledge of how a fragment binds, enables hit optimisation through structure guided medicinal chemistry, leading to the development of lead compound(s) for drug discovery.

In contrast to High Throughput Screens (HTS), where libraries of drug-like molecules (~500 MW) are screened to identify high affinity (nM- $\mu$ M) interactions, fragment-based studies identify weaker interactions ( $\mu$ M-mM) with lower mass (<300 MW) compound libraries.<sup>290</sup> Although fragment interactions are not as potent as those of larger molecules, it is possible to achieve greater ligand efficiency (defined as the binding energy per heavy atom of a ligand to its binding partner), thus providing an initial high quality starting point for further drug development. If structural data is to hand, fragments can be developed by growing into other protein pockets to enhance their potency and/or selectivity. Software development which enables 'automated' fragment growing computationally, such as that developed by Novartis, facilitates optimisation.<sup>291</sup> An alternative approach involves linking together fragments which bind to different protein pockets. Such 'Fragment growing' and 'fragment linking' methods have been reported with some success for fragment development against other targets in the literature, both in academic and industrial settings.<sup>289, 290, 292</sup> Indeed, the first fragment based drug, Vemurafenib, was approved by the FDA in 2011 as a treatment for metastatic melanoma.

Despite this precedent for success, fragment development does present challenges. Whilst traditional HTSs employ very large ( $10^6$ ) compound libraries, fragment libraries are much smaller ( $10^3$ ), but can cover a relatively large area of chemical space. Library design is therefore essential in

order to conduct a diverse high quality screen. Libraries must consist of high purity compound samples, with good solubility at high concentration. The target itself also dictates the experimental formats available for a large scale screen, since the quantity of target that can be produced, its purity, and solubility need to be taken into consideration. Similarly, its ability to crystallise and compatibility with NMR methods are important criteria to meet to successfully characterise potential fragment hits.



**Figure 51: Schematic representation of the general FBDD workflow. A protein target is screened against a fragment library, and identified hits further validated and characterised using complementary biophysical methods. Iterative cycles of structure-guided hit optimisation, validation, and characterisation, result in the identification of novel lead compounds for further drug development.**

The ability to successfully screen an enzyme is dependent on the availability of a robust assay. Due to the low affinity of the interactions being detected, false positive rates are often high in FBDD. Consequently, hits are identified through a variety of complementary biophysical and structural techniques. Used in combination, the acquired data enables triaging of potential hits and should lead to a reduction in false positives. The biophysical and structural techniques employed in fragment-based screens are briefly discussed in the following Section.

## 4.1.2 Biophysical and structural techniques

### 4.1.2.1 Differential scanning fluorimetry (DSF)

DSF is a method of measuring the temperature at which a protein unfolds, or 'melts' ( $T_m$ ) when subjected to a temperature gradient.<sup>293</sup> A reporter dye binds to hydrophobic regions that become exposed on unfolding, leading to an increased fluorescence signal. Fragment binding should stabilise the protein, leading to an increase in the unfolding temperature. The change in  $T_m$  ( $\Delta T_m$ ) induced is used to rank the fragments as potential hits.

It is a fast, cheap, and experimentally easy method of conducting a large high throughput primary screen whilst consuming low quantities of target and fragment. However, analysis can be difficult for proteins consisting of multiple domains that present multiple unfolding events. The reporter dye may also interfere with some fragments. Due to this lack of reliability, careful analysis of DSF data is required, as well as the use of complementary techniques to detect false positives.

### 4.1.2.2 Surface plasmon resonance (SPR)

SPR is a method of measuring protein-fragment interaction by measuring the changes in refractive index of a surface displaying immobilised protein as fragment is injected across the surface. Immobilised protein in a relevant conformation is required and the technique is more technically challenging than others, however, only small amounts of material are required and, unlike DSF, it is possible to directly calculate dissociation constants ( $K_d$ ), although the enthalpy of unfolding is required.

### 4.1.2.3 Ligand-detected NMR spectroscopy

There are several 1D NMR techniques that enable ligand-detected binding events to label-free protein. A classical CPMG (Car-Purcell-Meiboom-Gill) experiment relies on changes in tumbling rate (and therefore nuclear relaxation rate) of a fragment on its binding to protein. Slower tumbling rates on binding lead to faster T2 nuclear relaxation in comparison to the free, unbound form.<sup>289</sup> Such changes in T2 relaxation in turn translate into observable peak broadening or removal in the NMR spectra. Saturation Transfer Difference (STD) is another technique which transfers magnetisation from the protein to the fragment via the Nuclear Overhauser effect (NOE) only when the fragment is bound.<sup>294</sup> WaterLOGSY is a further variation on STD, whereby magnetisation transfer occurs between the fragment and water molecules in the protein's solvation shell when the fragment is bound.<sup>295</sup> This transfer of magnetism is observed as an inverted resonance signal in the NMR spectra. All the above techniques require optimisation for the target of interest, high amounts of protein, and provide a medium throughput screening platform. As they rely on changes in tumbling rates,

best results are seen for large globular proteins (>30 kDa); the techniques are less applicable for small or dynamic proteins.

#### 4.1.2.4 Protein-detected NMR spectroscopy

Low throughput protein-detected NMR techniques require large quantities of isotopically ( $^{15}\text{N}$ ) labelled material, and an assigned 2D spectrum. Though this method enables binding site mapping, the necessity of *a priori* structural information makes this technique less generally applicable to therapeutic targets.<sup>289</sup>

#### 4.1.2.5 Isothermal titration calorimetry (ITC)

ITC measures the energy required to maintain a constant temperature as a small molecule is titrated into a protein sample with respect to a control. It is the 'gold standard' characterisation method for protein-small molecule interactions and provides a full set of thermodynamic parameters ( $\Delta G = -RT\ln K_a = \Delta H - T\Delta S$ ). However, due to the low affinity of fragment interactions, data collection can present technical challenges, as the  $\Delta H$  may not be enough to detect. Large quantities of protein are required, as well as high fragment concentration and solubility. Matching DMSO levels in the protein and fragment solutions is important during sample preparation, as small concentration differences can create data artefacts. Further complications may arise if the fragment binds to multiple sites. The method is low throughput, and therefore not suitable as a primary detection method. None the less, ITC is routinely employed in fragment applications as a characterisation method once optimal conditions has been defined for a target protein, due to the importance of thermodynamic parameters in the downstream development of inhibitors molecules.<sup>289</sup>

#### 4.1.2.6 Micro-scale thermophoresis (MST)

Micro-Scale Thermophoresis (MST) is a recently developed method of determining the state of a protein (e.g. bound vs unbound) by measuring changes in molecular mobility across micro-scale temperature gradients.<sup>296</sup> Temperature gradients are induced by an IR-laser, and thermophoresis is detected by measuring the fluorescent distribution of labelled proteins. A molecule's thermophoretic signal is affected by changes to its surface area, hydration shell and effective charge of the protein, therefore, a binding event results in a change of thermophoretic signal. It has a fast measurement time and simple sample preparation to provide binding affinity data. As measurements are recorded in capillaries, sample consumption is low. In some cases the fluorescent signal observed before a temperature gradient is induced is dependent on fragment concentration, providing an additional measure of binding. MST is particularly well suited for fragment binding studies due to its high sensitivity range (nM-mM). However, it does require fluorescently labelled

protein, which may interfere with fragment binding. Due to its infancy, the theory of thermophoresis is still under debate and MST is yet to be fully validated as a technique; various binding profiles are possible and require careful analysis. That aside, although currently not as prevalent as ITC, MST does offer a much desired fast and low consuming method of studying low affinity interactions.

#### 4.1.2.7 X-ray crystallography

Due to the atomic detail that it provides, X-ray crystallography remains an unparalleled method for structure determination.<sup>289</sup> As well as identifying binding sites (even those of very weak binders), it can detect changes in protein conformation on binding, providing a wealth of detail to inform downstream inhibitor development. Generally, protein crystallisation requires moderate amounts of purified macromolecule, and optimised conditions that produce crystals of sufficient quality for X-ray diffraction data collection.<sup>297</sup> Experimental electron density enables the generation of a 3D structure. For the purposes of FBDD, protein crystals need to be either soaked or co-crystallised with the fragment of interest in order to solve the bound structures. Accordingly, proteins amenable to FBDD should yield reproducible crystals that can withstand high DMSO and fragment concentrations. Generally a resolution of 2.5 Å or greater is required to accurately map the fragment into the electron density. Though X-ray diffraction is a powerful technique, and essential for further fragment development, its low throughput nature makes it better suited as a characterisation technique rather than a means of initial screening for most laboratories.

The suitability of each technique for each stage of fragment discovery is summarised in Table 1 .

	Initial Screening	Validation	Characterisation	Protein consumption	Throughput
DSF				+	+++
SPR				+	++
NMR (ligand)				++	++
NMR (Protein)				+++	+
ITC				+++	+
MST				+	+++
X-ray				++	+

**Table 1: 'Traffic light' summary of the suitability of biophysical methodologies at each stage of FBDD. Protein consumption and throughput considerations are included (+ = low; ++ = medium, +++ = high).**

#### 4.1.3 Screening E3 ligases

The drug discovery effort towards Ub machinery and the resulting inhibitors has been previously introduced in Chapter 1. Though both HECT and RBR E3 ligase families possess active sites, and as

such are somewhat conventional drug targets, they have received surprisingly little attention to date. The application of FBDD to HOIP, an RBR, is discussed in this Chapter.

#### **4.1.4 HOIP**

HOIP is the catalytic component of the LUBAC E3 ligase complex.<sup>24</sup> LUBAC is a particularly interesting E3 ligase: it is the only known linear chain E3 ligase, and the only macromolecular E3 ligase. To date, there are no known inhibitors of HOIP, or for any other E3 ligases belonging to its RBR E3 ligase subfamily. HOIP plays a key role in inflammation and innate immunity, through NF- $\kappa$ -B activation, and is an antiapoptotic regulator.<sup>298</sup> Interestingly, Mackay *et al.* have found that HOIP depletion sensitises cancer cells to cisplatin, leading to the hypothesis that inhibition of HOIP in combination with chemotherapeutic treatments would enhance the efficacy of platinum-derived chemotherapeutics.<sup>113</sup>

HOIP employs a catalytic cysteine in its Ring2 domain to transfer a donor Ub onto an acceptor Ub during linear chain synthesis. The minimal catalytic core (P37) required to enable this synthesis has been determined by Stieglitz *et al.*<sup>298</sup> X-ray crystal structural analysis of P37 in complex with a donor and acceptor Ub during linear chain formation reveals a hydrophobic pocket interacting with the donor Ub C-terminal tail adjacent to the active site, presenting an interaction which could be targeted by a small molecule. As the unbound X-ray structure of P37 was also determined, P37 represented a valid target for a fragment-based drug discovery screen.

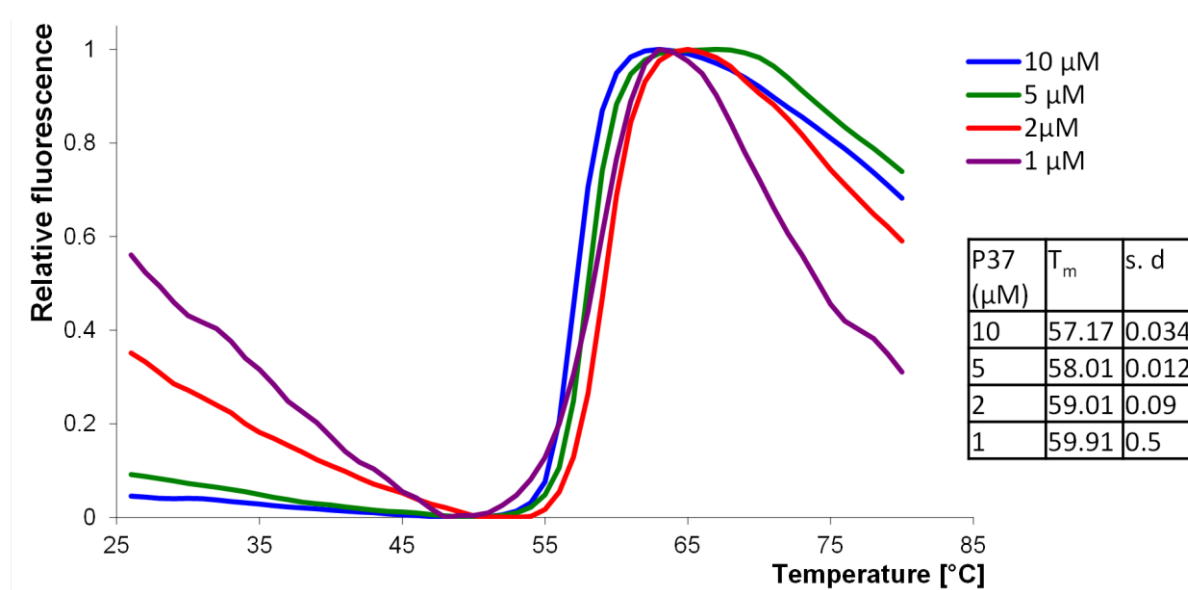
The study was conducted as detailed below in collaboration with Katrin Rittinger (Crick Institute, Mill Hill). P37 had previously been produced in large quantities and in high purity in Katrin Rittinger's laboratory. Consisting of a single protein domain, P37 was well suited to DSF studies; therefore, this was employed as a preliminary screening platform.

## **4.2 DSF screening of HOIP catalytic core**

### **4.2.1 Optimisation of DSF assay for P37**

First, the suitability of DSF as a FBDD platform for P37 was determined. P37 was provided as a solution in 50 mM HEPES, 150 mM NaCl, 1 mM DTT at pH 7.5. As HEPES buffers are generally compatible with the DSF assay this buffer was used for initial screen tests. Sypro<sup>®</sup> Orange, a common dye for DSF applications, was also used; its relatively high excitation wavelength of 492 nm minimises the likelihood of the small molecule screen interfering with the dye's optical properties.

A series of P37 concentrations were tested in the DSF assay to aid selection of a suitable concentration for the assay (Figure 52). Briefly, varying concentrations of P37 were incubated in the presence of SYPRO® Orange dye and 2 % DMSO, then fluorescence measurements were recorded over a 25-95 °C gradient, using a real time PCR machine. Readings were performed in triplicate. The raw data was analysed using an Excel workbook customised for DSF analysis, which was developed and made publically available by Dr Frank Niesen (Structural Genomics Consortium, UK <ftp:ftp.sgc.ox.ac.uk/pup/biophysics>). The workbook was used to create normalised graphs to represent the data, and  $T_m$  values were calculated by fitting the data to the Boltzmann equation using GraphPad Prism 5 software.

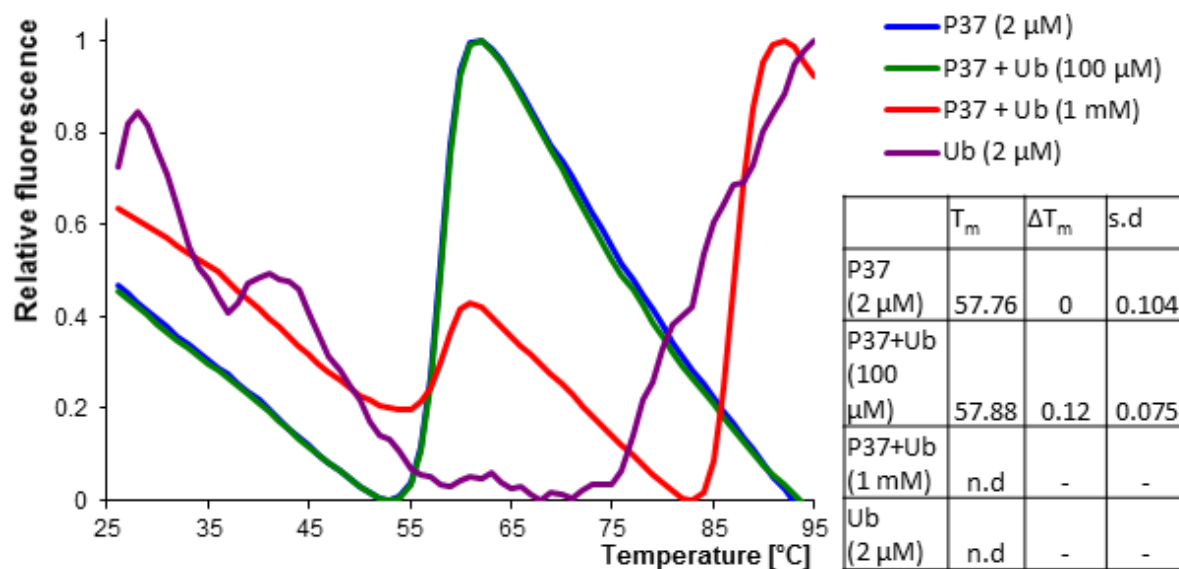


**Figure 52: Initial protein concentration optimisation for DSF assay.** P37 in 50 mM HEPES, 150 mM NaCl, 1 mM DTT, pH 7.5, 2 % DMSO was incubated with 10x SYPRO® Orange dye and the fluorescence was recorded over a 25-95 °C gradient.  $T_m$  values were calculated from the mid-point of the melting curve using GraphPad Prism 5 software. Curves shown are representatives of triplicate readings, the mean and standard deviation of which are indicated.

In all cases P37 demonstrated a sharp unfolding transition at a high  $T_m$  (> 55 °C), suggesting that P37 is unfolding as a single domain transition, and is stable enough in isolation to conduct a DSF assay. 2 µM was selected as a suitable concentration for conducting the fragment based screen, as it was the lowest concentration tested that presented minimal background fluorescence. The reproducibility of the reading was high, with a standard deviation (s.d) below 0.1 °C, making it a high quality negative control.

As no small molecule P37 binders are known, it was proposed that binding of P37 to Ub could be used as a positive binding control. The use of a protein binder has its own complications, however, due to its own unfolding profile contributing to the fluorescence reading in the DSF assay. P37 was

incubated as previously described, with the addition of Ub (100  $\mu$ M or 1 mM), and  $T_m$  curves measured (Figure 53). Whilst 100  $\mu$ M of Ub made little impact on  $T_m$ , an increased concentration of 1 mM significantly changed the fluorescence profile. Accordingly, Ub was not carried through as a positive control for the screen. Instead, the screen was conducted using the highly reproducible negative control.



**Figure 53: Determining the suitability of Ub as a positive control. P37 and Ub were co-incubated with SYPRO® Orange dye and the fluorescence recorded. P37 and Ub alone readouts are included as controls**

#### 4.2.2 Initial screen of P37

Work proceeded to screen a fragment library against P37 using the optimised DSF assay conditions. The Maybridge Ro<sub>3</sub> 2500 Diversity Fragment Library (Thermo Fisher Scientific) was chosen, due to its high structural diversity and rule-of-three compliance as outlined below:

- ✓ MW  $\leq$  300 Da (mean value of 181.1 Da)
- ✓ clogP  $\leq$  3.0 (mean value of 1.6 Da)
- ✓  $\leq$  3 Hydrogen bond donors / acceptors (mean values of 0.9 and 2.4 respectively)
- ✓  $\leq$  3 rotatable bonds (mean value of 1.0)

Fragment quality ( $\geq$  95 % purity) and DMSO solubility (200 mM) were also assured by the supplier.

Briefly, P37 was incubated in the presence of SYPRO® Orange dye with 2 mM fragment or equivalent DMSO control, then fluorescence measurements recorded over a 25-95 °C gradient, and  $T_m$  values



recorded as previously described. The error of the assay was determined by measuring the standard deviation of the DMSO control. An assay plate was deemed to be of suitable quality if the standard deviation of the negative controls was  $\leq 0.1^\circ\text{C}$ . The total error of the assay across all 29 plates measured was greater, with the standard deviation of the negative control standing at  $0.3^\circ\text{C}$ . This difference is accounted for by an observed lowering of the control average  $T_m$  over time, which can be correlated to the use of a new Sypro<sup>®</sup> Orange batch from the supplier (due to the scale of the screen). Therefore, fragments were evaluated relative to their assay plate standard deviation. A change in average  $T_m$  ( $\Delta T_m$ )  $> 2$  s.d was considered to be a possible hit fragment.

The results of the initial DSF screen are summarised in Table 2. Fragment screen hit rates are often high (typically 3-10 %). As a positive hit rate ( $> 2$  s.d) of 5.6 % is mid-range for an initial fragment-based screen, a more stringent cut off was implemented ( $> 4$  s.d) to define initial hits. This decreased the number of defined hits to 67 (2.9 %) for further analysis. Of these, 41 fragments induced a  $\Delta T_m > 0.5^\circ\text{C}$ . It was observed that a high percentage of fragments induced a negative shift. It is unknown if these fragments would benefit from further analysis. It has been suggested that the identification of destabilising factors could aid buffer design for crystallisation purposes.<sup>299</sup> Although potentially of interest, focus remained on positive shifters alone for the purpose of screening P37.

Fragment classification	Applied threshold	Number of fragments	Percentage
+ve shift	$\Delta T_m > 2$ s.d	131	5.6 %
	$\Delta T_m > 4$ s.d	67	2.9 %
	$\Delta T_m > 5^\circ\text{C}$	41	1.8 %
-ve shift	$\Delta T_m < 2$ s.d	935	40.3 %
	$\Delta T_m < 4$ s.d	530	22.9 %
	$\Delta T_m < 5^\circ\text{C}$	265	11.15 %
Non-shift	$2$ s.d $\geq \Delta T_m \geq -2$ s.d	1168	50.4 %
	$4$ s.d $\geq \Delta T_m \geq -4$ s.d	1637	70.5 %
	$5^\circ\text{C} \geq \Delta T_m \geq -5^\circ\text{C}$	1928	83.1 %
Ambiguous melting curve	Invalid Boltzmann fit	86	3.7 %

**Table 2: Overview of results from DSF screen of Maybridge Ro<sub>3</sub> 2500 Diversity Fragment Library.**

In order to streamline the number of hits taken forward whilst retaining chemical diversity, the 67 hits were subjected to a structure similarity search across all members of the fragment library using ChemBioFinder (Perkin Elmer). In this way structurally similar hits were categorised into groups. Non-hit fragments were also grouped into these categories: on inspection of their  $\Delta T_m$  this enabled the prediction of potential SAR data for fragment binding to P37. Some fragment categories were highly populated, such as those containing a sulphonyl imidazole, thiourea, or aromatic sulphone

motif. Hits not falling into a group were considered as independent categories. After assessment, 6/15 categories were discarded on the basis of:

- ✗ Likely cross reactivity of chemical groups present (eg. hydrazines, Michael acceptors)
- ✗ The predicted SARs implied that a hit was a one off or false positive. (i.e. the majority of the category are defined as non-hits).
- ✗ Taking predicted SARs into account, the major motif was not suitable for fragment growth. (i.e. strong reliance on additional groups beyond the major motif, which prevented any variation to the fragment structure whilst maintaining a positive  $\Delta T_m$ ).

The melting curve of each hit was then manually inspected and hits with irregular curves (eg. bowled shaped, high background) discarded. Of the remaining eight categories one representative hit, and in some cases a further analogue, were taken forward for further validation, as detailed in Table 3. Whether a hit had been identified in other proteins screens conducted by our research group against the Ro3 library was also checked. This was found to be the case for fragment **11A5**, and several hits in the thiourea category. Thioureas are renowned unspecific protein binders, and it was surprising to find that so many thioureas (86 fragments, 10 of which afforded a positive  $\Delta T_m$  with P37) had been included during library design. Though not useful as specific binders for P37, **11A5** and one thiourea fragment, **27C4**, were retained for comparison to other hits in future validation experiments.

Fragments **18H9** and **26E11** gave the largest thermals shift of +3.3 and 1.6 °C respectively. **18H9** represents one of ten sulphonyl imidazole containing fragments present in the Ro3 library, only one of which did not induce a positive  $\Delta T_m$ . This strongly suggests that this motif is indeed binding to P37. **26E11** is a representative of the aromatic sulphone containing fragment category, which also includes **11A5**. Interestingly, of the 19 aromatic sulfone containing fragments in the screen, only 6 induced a positive shift. These SARs could be valuable for future fragment development. The remaining fragments (**27A7**, **24H2**, **23B4**, **8C8**, **18D4** and **27C4**) all afforded weaker shifts in the 0.8-0.6 °C range. It is difficult to draw any meaningful SAR data for these fragments, possible due to the low  $\Delta T_m$  values measured. Melting curves for the fragments and analogues carried forward are shown in Figure 54 and Figure 55.

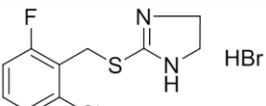
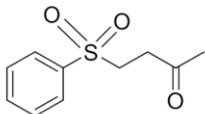
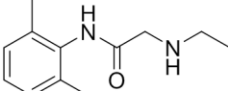
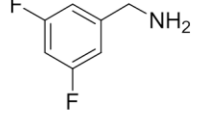
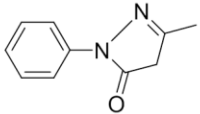
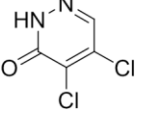
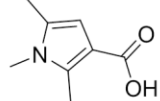
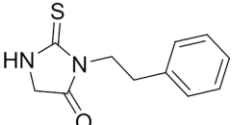
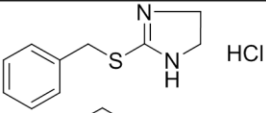
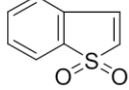
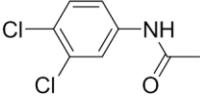
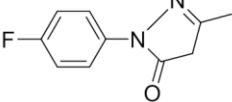
A				
Fragment ID (plate ID)		$\Delta T_m$ (°C)	Chemical Structure	MW (Da)
CD05341 (18H9)		+3.3		244.70
FM00330 (26E11)		+1.6		212.27
SEW03864 (27A7)		+0.8		206.29
AC39479 (24H2)		+0.7		143.14
NRB05254 (23B4)		+0.7		174.20
AC42667 (8C8)		+0.6		164.99
CC08601 (18D4)		+0.6		153.18
RF03152 (27C4)		+0.6		220.30
B				
Fragment ID (Plate ID)	Analogue of:	$\Delta T_m$ (°C)	Chemical Structure	MW (Da)
DP01047 (21A2)	(18H9)	+1.7		228.75
MWP0004 (11A5)	(26E11)	+1.7		166.20
RDR01104 (27A4)	(27A7)	+1.0		204.06
RH01710 (23G3)	(23B4)	+0.5		192.19

Table 3: Fragment ID, plate ID,  $\Delta T_m$ , Chemical structure, molecular weight of the (A) fragments and (B) analogues selected for further validation.

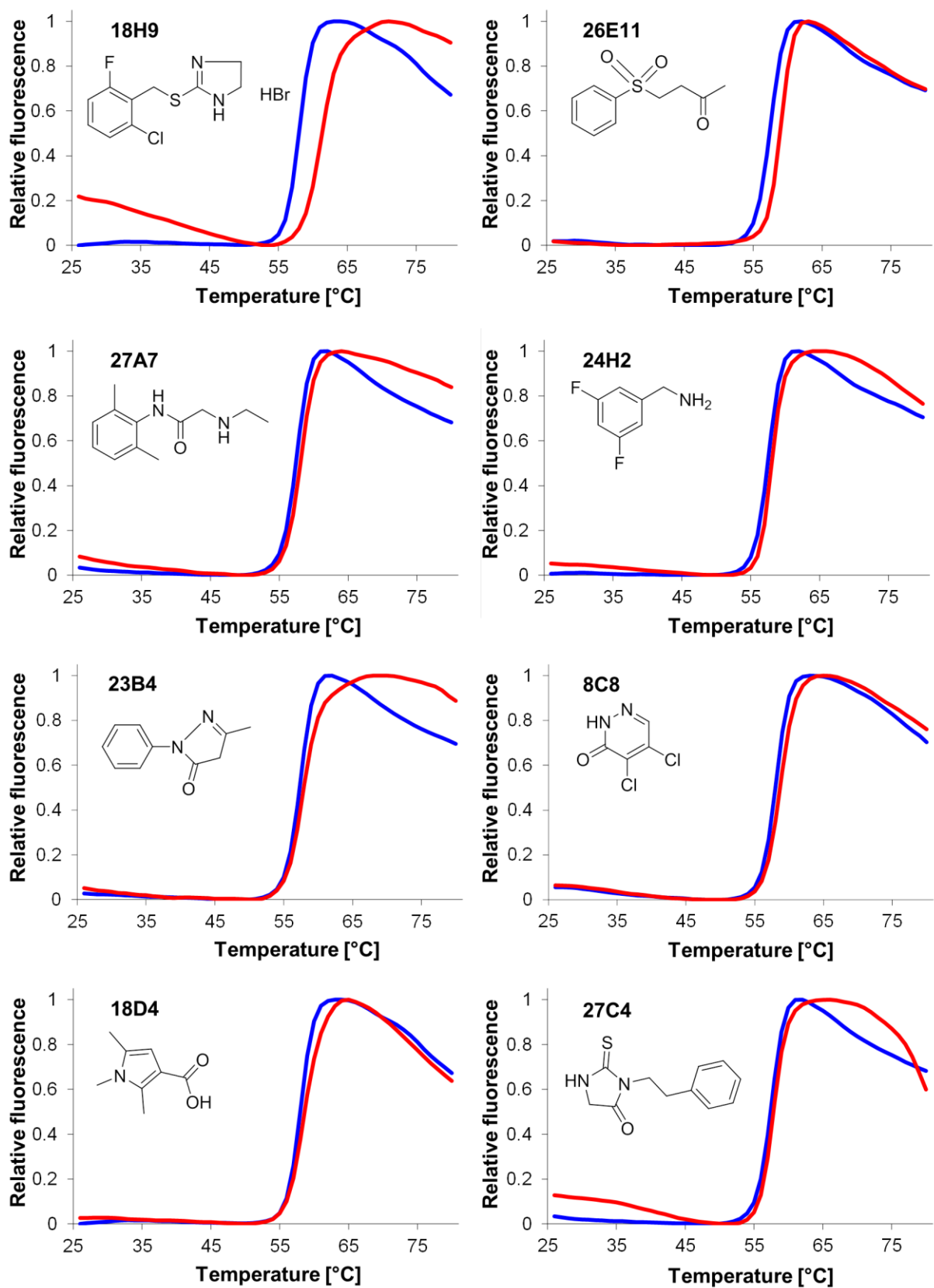


Figure 54: Melting curves for P37 (2  $\mu$ M) in the presence 2 mM of indicated fragments (red) or DMSO control (blue).

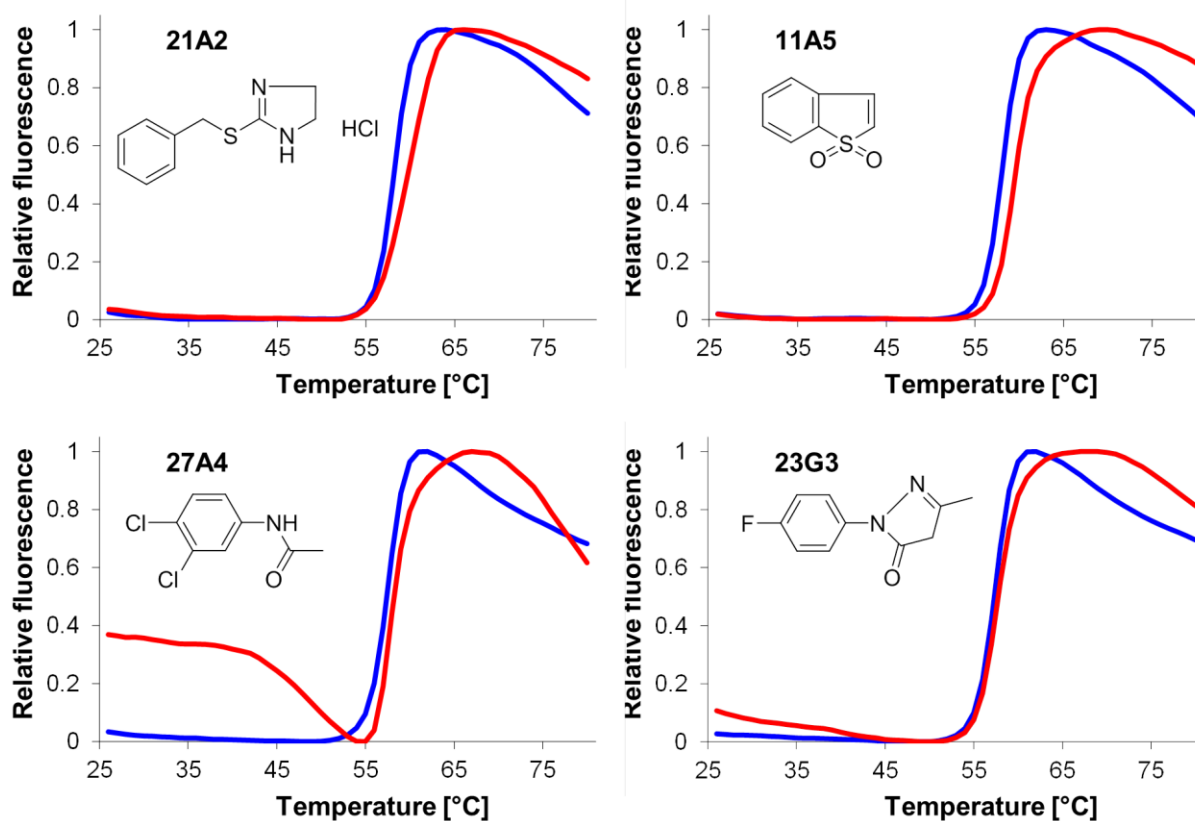


Figure 55: Melting curves for P37 (2 μM) in the presence 2 mM of indicated fragment analogues (red) or DMSO control (blue). Representative melting curves for P37 (2 μM).

### 4.3 Concentration dependence studies by DSF

It was of interest to study the effect of fragment concentration on binding. Generally,  $\Delta T_m$  should reasonably correlate with binding affinity and fragment concentration.<sup>300</sup> The DSF assay was repeated in duplicate at varying concentrations between 0.5-2.0 mM for indicated hits and analogues in order to further validate fragment binding. These 'titration' experiments illuminated three main types of binding behaviour: linear, saturated and inverted (Figure 56).

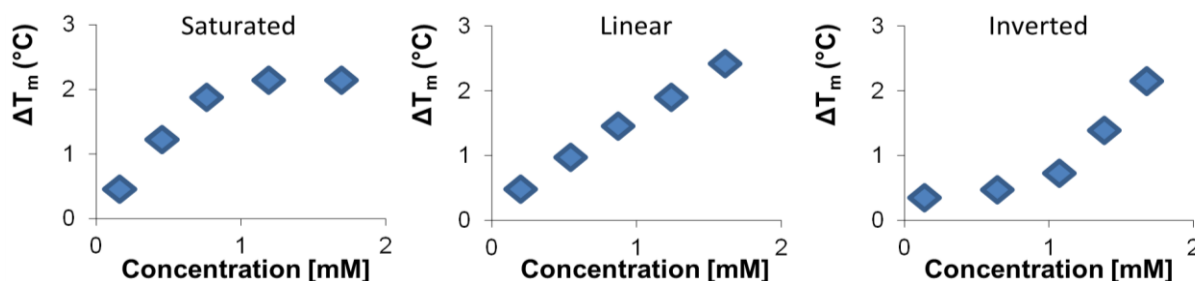
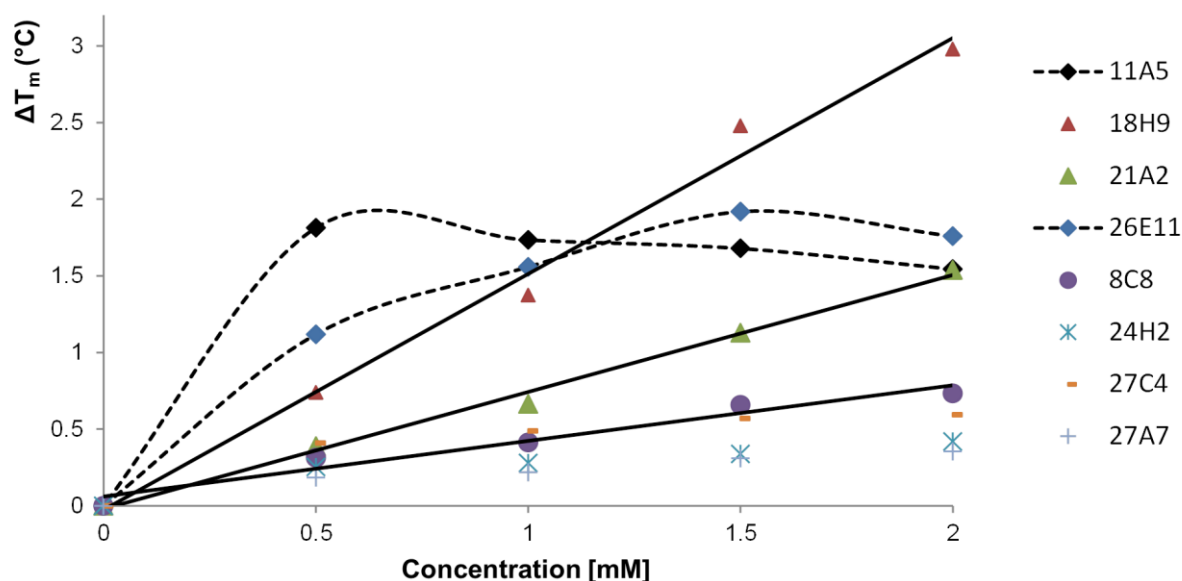


Figure 56: Cartoon representations of three alternative concentration dependent fragment behaviours during DSF experiment: saturated, linear and inverted.

The effect of fragment concentration on  $\Delta T_m$  is shown in Figure 57. **26E11** and its analogue **11A5** both demonstrated saturated concentration dependence across this concentration range. In

contrast, **18H9** and its analogue **21A2** demonstrated linear concentration dependences. Of the remaining four fragments tested, three (**27C4**, **27A7**, **24H2**) exhibited an insignificant effect on  $\Delta T_m$  with varying concentration, as determined by a 2 standard deviation cut-off with respect to the negative control. **8C8** displayed a significant but weak linear dependence. It is postulated that the concentration dependence of **8C8** is attributed to an ‘inverted’ binding behaviour, whereby stabilisation significantly increases at higher fragment concentration.



**Figure 57: Ligand concentration dependence of  $\Delta T_m$  by DSF for selected fragments.**

The relationship between ligand concentration and  $\Delta T_m$  has been studied both theoretically and experimentally.<sup>301, 302</sup> Assuming a 1:1 binding model and a single unfolding transition, they should correlate in a roughly linear fashion. However, a saturated relationship is also possible if ligand solubility is limited, or if the ligand increasingly binds to denatured protein. Equally, it has been suggested that minimal concentration dependence or inverted relationships are observed when the concentration range is much lower than the  $K_d$  value of the fragment.<sup>300</sup>

Although titration experiments add to the body of evidence for hit identification, alternative validation methods are necessary to determine true fragment binders, as discussed in the following sections.

## 4.4 WaterLOGSY validation

Ligand-detected waterLOGSY NMR spectroscopy was chosen as a preliminary means of fragment characterisation, as it required modest amounts of protein and material and was of moderate throughput. The method had also been previously used with success in the laboratory on an alternative protein target. Briefly,  $^1\text{H}$  spectra were acquired for each fragment at 1 mM using the same buffer conditions employed for DSF analysis. A waterLOGSY experiment was then run for each fragment alone, before addition of P37 and further waterLOGSY acquisition. Peak inversion on protein addition was indicative of fragment binding. Initially, P37 was added at a final concentration of 10  $\mu\text{M}$ , and increased to 40  $\mu\text{M}$  if spectral shifts were inconclusive.

The results of waterLOGSY validation of fragments are summarised in Figure 58. In some cases, analogous fragments from within a fragment category were also validated to provide further information. **18H9** and **27C4** were validated as binders, with at least one proton signal being inverted on protein addition. Analysis of other fragments was less conclusive. A decrease in 'trough depth' with increasing protein concentration was observed for **8C8**, **21A2**, **26E11** and **27A4**. Side-shifting of peaks on addition of protein was difficult to characterise; this was observed for **18D4** (data not shown) and partially for **11A5**. No inversion was observed for **23B4** or its analogue **23G3** (data not shown), nor **27A7**. No waterLOGSY was obtained for **24H2** due to weak proton signal.

The titration experiment and waterLOGSY data together suggested that **18H9** was a valid hit. It was somewhat surprising that **27C4** was validated by NMR after the weak concentration dependence observed by DSF. This, coupled with the known promiscuous binding of thioureas to proteins, led to the exclusion of **27C4** from further study. Conversely, after encouraging titration results, **26E11** and its analogue **11A5** performed poorly in waterLOGSY experiments. Data for **8C8** and **24H2** were also inconclusive; consequently the four fragments **18H9**, **26E11**, **8C8** and **24H2** were taken forward for further analysis.

The inconclusive nature of the waterLOGSY data could be in part due to the small size of P37; 25 kDa lies below the recommended threshold of 30 kDa for robust ligand detected NMR analysis.<sup>289</sup> The suitability of waterLOGSY also depends on how well hydrated the binding site is and how transient the binding event is; both of these factors influence the capacity for magnetisation transfer. STD and CPMG experiments offer valid alternatives. STD is beneficial for hydrophobic binding sites, whilst CPMG is useful for studying tighter binding interactions. However, due to time constraints and the amount of material available neither of these alternative experiments have been attempted to date.

Although protein-detected NMR spectroscopy would have provided an alternative validation method, insufficient quantities of <sup>15</sup>N labelled P37 was available at the time of screening, and a characterised correlation spectra was not available. ITC was also discarded as an alternative method, due to the amount of protein required. SPR, although consuming low amounts of protein and ligand, would have required significant optimisation. Therefore, in the interest of time and making use of available expertise and equipment within the collaborative groups, the four selected fragments, **18H9**, **26E11**, **8C8** and **24H2**, were further validated and characterised by MST.



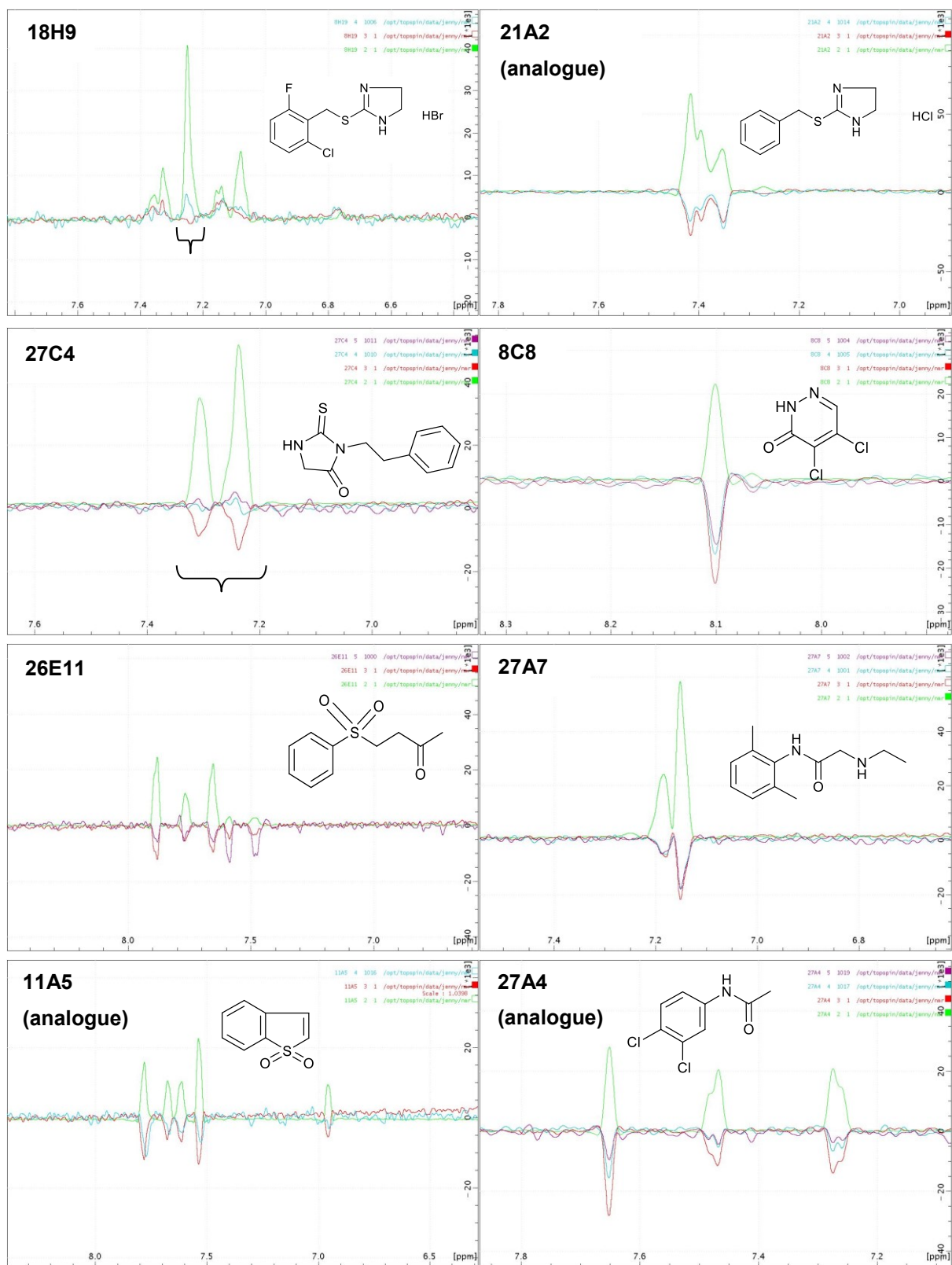
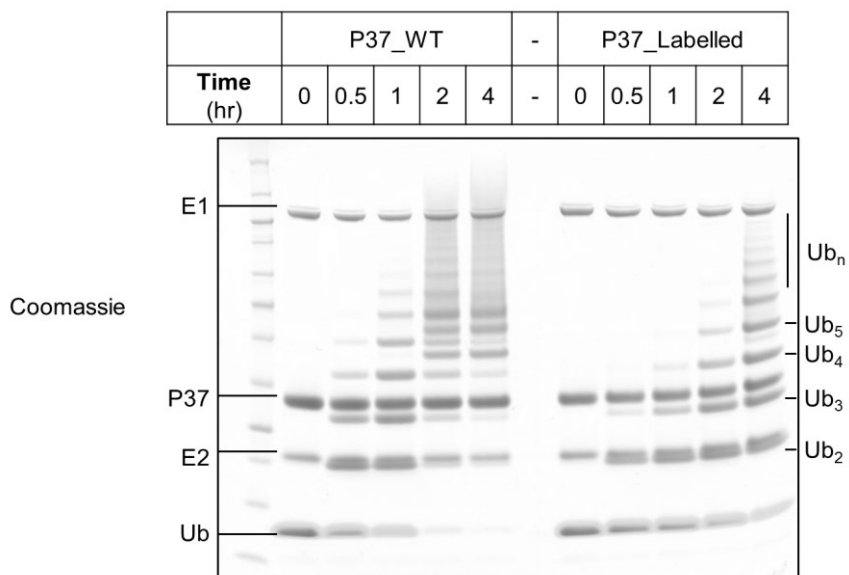


Figure 58: WaterLOGSY analysis, with spectral overlay of 1D fragment alone (green), waterLOGSY fragment alone (red), fragment plus protein at 10  $\mu$ M (cyan) and in some cases 40  $\mu$ M (purple). Inverted resonances are indicated by an open bracket.

## 4.5 Fragment characterisation by MST

All MST experimentation was undertaken with the help of Dr Steve Martin at the Crick Institute (Mill Hill). In order to measure  $K_d$  values by MST, P37 was first fluorescently labelled with a Cy5 fluorophore using a Monolith NT protein labelling kit. The kit utilises an *N*-hydroxysuccinimide (NHS) ester of the fluorophore, which is amine reactive with an optimum reaction pH between 8.3-8.5. Due to the presence of 12 lysine residues and an isoelectric point (IP) of 8.21, P37 is not an ideal protein as there are multiple amines that can be labelled. Consequently, the labelling reaction was undertaken below the recommended pH, at pH 7.5 in order to favour N-terminal reactivity. The reaction was left to proceed in the dark for 45 minutes, before removal of excess unreacted fluorophore by gel filtration. A labelling efficiency of 33 % was determined spectroscopically by measuring absorbance at 650 nm and 280 nm, with measured concentrations of labelled and unlabelled protein approximately 2  $\mu$ M and 6  $\mu$ M respectively.

An ubiquitination assay was performed by Katrin Rittinger's group in order to check the effect of labelling on protein activity. Labelled and wild type (WT) P37 were separately incubated with recombinant UBA1 (E1), ATP, Ubch7 (E2) and Ub for 4 hours at 25 °C. Samples were taken at set time intervals, the reaction quenched by addition of sample loading buffer, and then proteins were separated by SDS-PAGE and visualised using Coomassie staining (Figure 59).



**Figure 59: Determining the effect of fluorophore addition on p37 catalytic activity.** P37 was incubated with UBA1 (E1), ATP, Ubch7 (E2) and Ub at 25 °C. Time course samples were separated by SDS-page and Coomassie stained. WT= wild type. Ub<sub>2-n</sub> indicates linear chain stoichiometry.

Time dependent formation of linear Ub chains was observed as a series of higher MW bands. Band pattern analysis indicated that the labelled protein had a reduced enzymatic activity in comparison to wild type P37. This could mean that labelling interferes with binding of the fragments and as such the MST data cannot be used with high confidence. However, some catalytic activity does remain, with labelled protein showing a similar band activity after 2 hours to wild type after 1 hour. Consequently, the labelled P37 was used to measure fragment binding affinities by MST, but these measurements should be regarded as preliminary data requiring further validation using an alternative characterisation technique, such as ITC, with wild type P37.

In order to determine binding affinities, 100 nM of labelled protein in 50 mM HEPES, 100 mM NaCl, 0.5 mM TCEP, 0.05 % Tween, pH 7.5 was incubated for 5 minutes with a 5 mM-75 nM fragment concentration series at a constant DMSO concentration of 2 %. The samples were then transferred into capillaries and MST measurements were recorded using a Monolith NT.115, reading fluorescence at 670 nm. Both fluorescence and thermophoresis measurements were recorded and used to fit the raw data to non-linear binding curves. The obtained binding affinities of **18H9**, **26E11**, **8C8** and **24H2** are summarised in Table 4. Representative graphs are presented in Figure 60 and 61.

Fragment	Binding affinity ( $K_d$ )		Time dependence? (Y/N)
	Fluorescence determined (mM)	Thermophoresis determined (mM)	
<b>18H9</b>	1.08	0.92	Y
<b>26E11</b>	4.00	4.75	N
<b>8C8</b>	~45 mM	n.d	N
<b>24H2</b>	~29 mM	n.d	N

**Table 4: Summary of binding affinities determined by MST, comparing values recorded by both initial fluorescence and thermophoretic methods.**

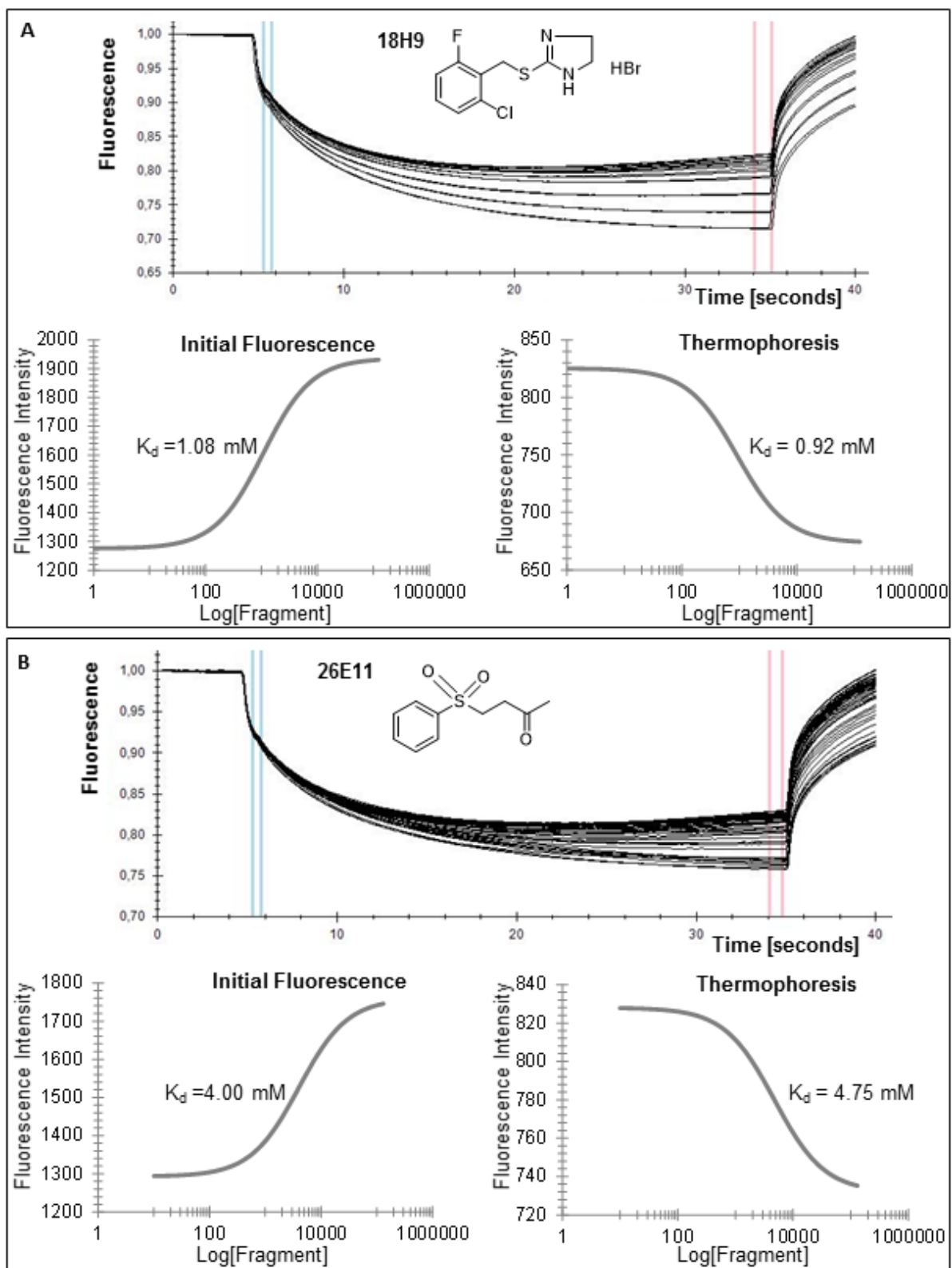


Figure 60: MST fluorescence (670 nm) profiles for fragments (A) 18H9 and (B) 26E11 binding to P37 (100 nM). A 2 mM-75 nM fragment concentration range was measured at 2 % DMSO. On introduction of an IR-laser induced temperature gradient, time dependent thermophoretic signal is measured (blue and red lines).  $K_d$  curve fittings of initial fluorescence and thermophoretic data are included.

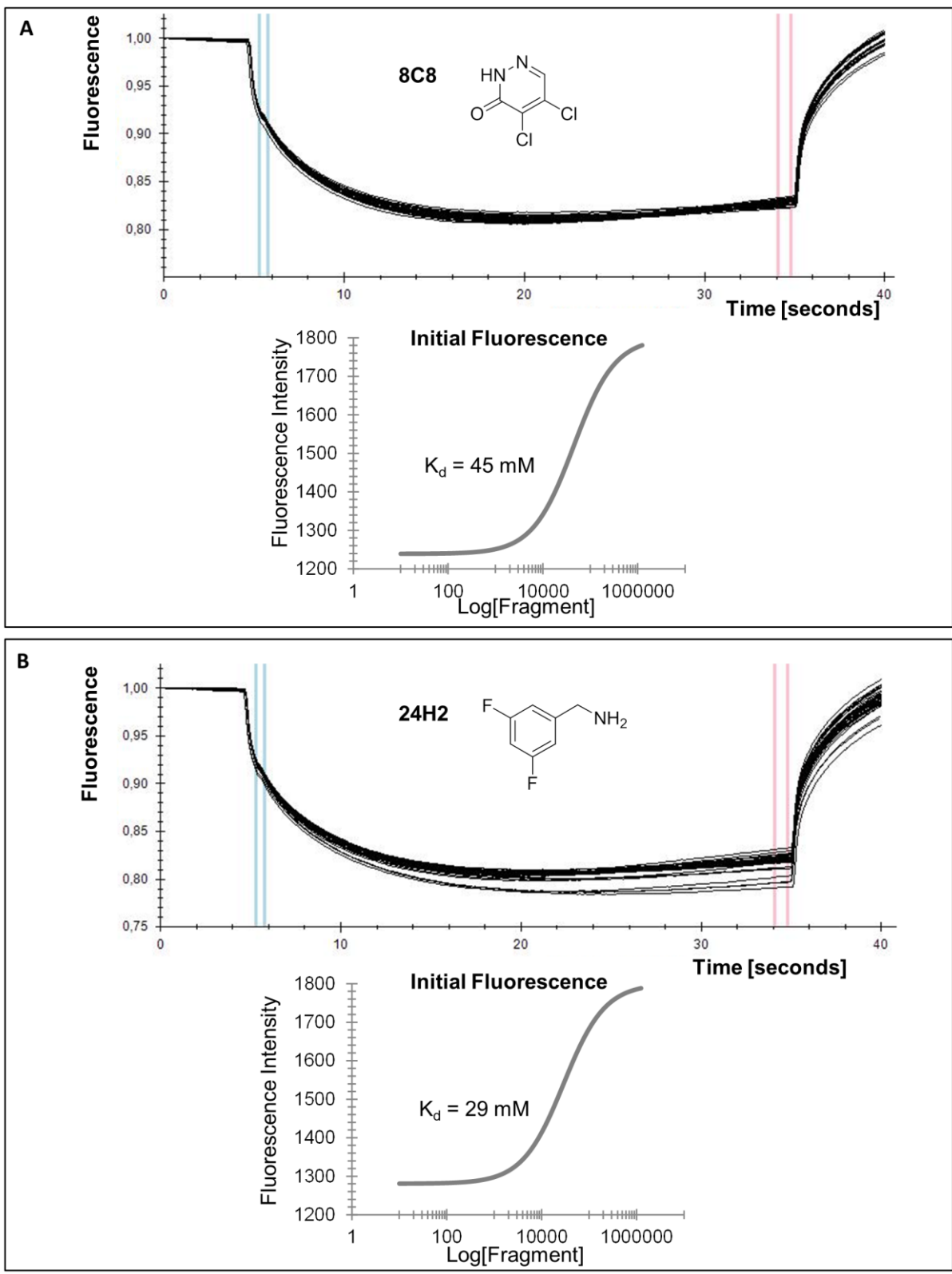
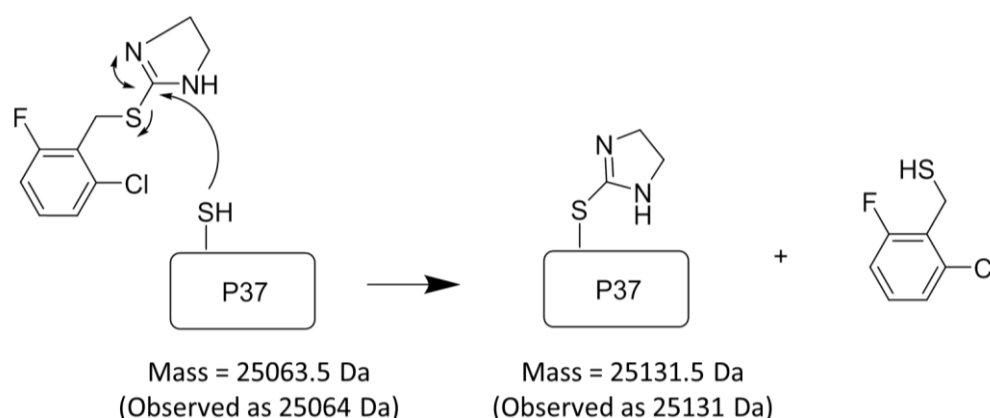


Figure 61: MST fluorescence (670 nm) profiles for fragments (A) 8C8 and (B) 24H2 binding to P37 (100 nM). A 5 mM-75 nM fragment concentration range was measured at 2 % DMSO. On introduction of an IR-laser induced temperature gradient, time dependent thermophoretic signal is measured (blue and red lines).  $K_d$  curve fittings of initial fluorescence data are included.

On initial testing, **18H9** showed insignificant binding to labelled P37. However, after incubation overnight at 4 °C the fragment demonstrated a good fit to the binding curve, affording a  $K_d$  of approximately 1 mM (Figure 60A).  $K_d$  values calculated by fitting both the fluorescent and thermophoresis data were in agreement. It was hypothesised that the observed time dependence was the result of a slow irreversible binding event. This hypothesis was confirmed by mass spectrometry, which revealed a mass change of +68 Da following overnight incubation of WT P37 with **18H9** (Figure 62). This mass change can be attributed to the addition of an imidazole ring, through the reaction of the catalytic cysteine with the sulphonyl imidazole containing fragment.



**Figure 62: Postulated mechanism for slow irreversible binding of fragment 18H9 to P37. For MS spectra please refer to Appendix Figure 1.**

**26E11** had an observed  $K_d$  of approx. 4.3 mM, averaged from both the fluorescent and the thermophoresis data, with no time dependence observed (Figure 60B). **24H2** and **8C8** had very large  $K_d$  values, calculated from initial fluorescence, of approximately 30 mM and 45 mM respectively which were not time dependent (Figure 61). As these values were much greater than the 5 mM maximum concentration used for MST dilution series, the  $K_d$  values were estimated by fixing the upper intensity limit to 1800: the maximal fluorescence reached by **18H8**, which reached saturation. Due to available ligand stock concentrations and protein aggregation at higher percentage DMSO, higher concentration readings were not attempted, but could be attempted as future work with new ligand stocks of higher concentration.

The  $K_d$  values recorded by MST are in agreement with the qualitative trend observed in the initial DSF screen. Together, this suggests that although measurements recorded using labelled P37 are not optimal, they do still provide insight into fragment binding with P37. Although large, the measured  $K_d$  values of **18H9** and **26E11** lie within an accessible region for fragment discovery<sup>303</sup> and should be characterised further.

## 4.6 X-ray crystallography

Fragments **18H9** and **26E11** were taken forward by Katrin Rittinger for X-ray crystallography, and co-crystallisation experiments attempted with P37. The ability of P37 to withstand DMSO was tested and found to be sufficient to conduct such trials, with structural integrity maintained in the presence of up to 10 % DMSO. However crystals with fragments bound were not observed. P37 has a very fragile crystal structure due to the presence of flexible loops, which is supported by a helical base platform.<sup>298</sup> It was suggested that a binding fragment would improve the structure's stability and thus the sturdiness of the crystal structure: P37 crystallises more robustly when in complex with Ub. Therefore co-crystallisation experiments were also attempted with this structure. However, no fragment binding was observed.

It is difficult to draw conclusions from this result. It is possible that the fragments do not bind, or do not bind tightly enough, under these conditions. If **18H9** binds to the catalytic cysteine as postulated, however, is it possibly that it is excluded by the presence of active site binding Ub.

## 4.7 Conclusions

The work in this Chapter focuses on a fragment based drug discovery strategy against the RBR E3 ligase HOIP. A library of 2320 fragments was screened against P37, the minimal catalytic core of HOIP. P37 alone displayed a high thermal stability. The hit rate was acceptable when a stringent 4 s.d. threshold was implemented. Hits that satisfied these requirements were further grouped by structural similarity, leading to eight representative compounds and four further analogues being taken forward for validation. On investigating the effect of fragment concentration on  $\Delta T_m$ , several fragment behaviours were observed. WaterLOGSY NMR spectroscopy was used as an alternative technique to validate fragment hits; however, most of the hit compounds did not induce a shift, or led to inconclusive findings. There are several possible explanations for this observation: i) the hits are false positives; ii) the interactions are too weak to detect; iii) the protein is not suitable for analysis by waterLOGSY. Further characterisation of a selection of fragment hits by MTS revealed at least two low mM potency hits, one of which (**18H9**) was a covalent binder. It is possible that this interaction is occurring with the active site cysteine.

It is generally agreed that high resolution structural data is essential to enable characterisation of fragment binding modes and thus the rational development of fragments into lead compounds. To date, however, attempts to crystallise P37 with fragments **18H8** and **26E11** have been unfruitful.

Targeting the catalytic core represents one strategy of RBR ligase inhibition. However, the mechanism of these ligases is reliant on the formation of other transient complexes, which could also be targeted. In Chapter 5, model systems to study these alternative mechanisms of RBR inhibition are discussed.



## **Chapter 5      Crystallographic models to facilitate the structural study of E3 ligases.**

---

Structural understanding of enzymatic reactions underpins our ability, and ultimately influences our strategy, to disrupt protein function. Protein complexes formed during enzymatic events are often difficult to study due to their transient nature. The development of more stable model systems which mimic such complexes is one method of furthering our understanding. This Chapter reports the development of a covalent Ub-E2-E3 ligase complex through the use of a tri-maleimide trap. Trap design and synthesis, optimisation of Ub mono-incorporation strategies, and efforts towards trapping the Ub-Ubch7-HHARI complex are described.

## 5.1 Introduction

### 5.1.1 Structural study of transient complexes

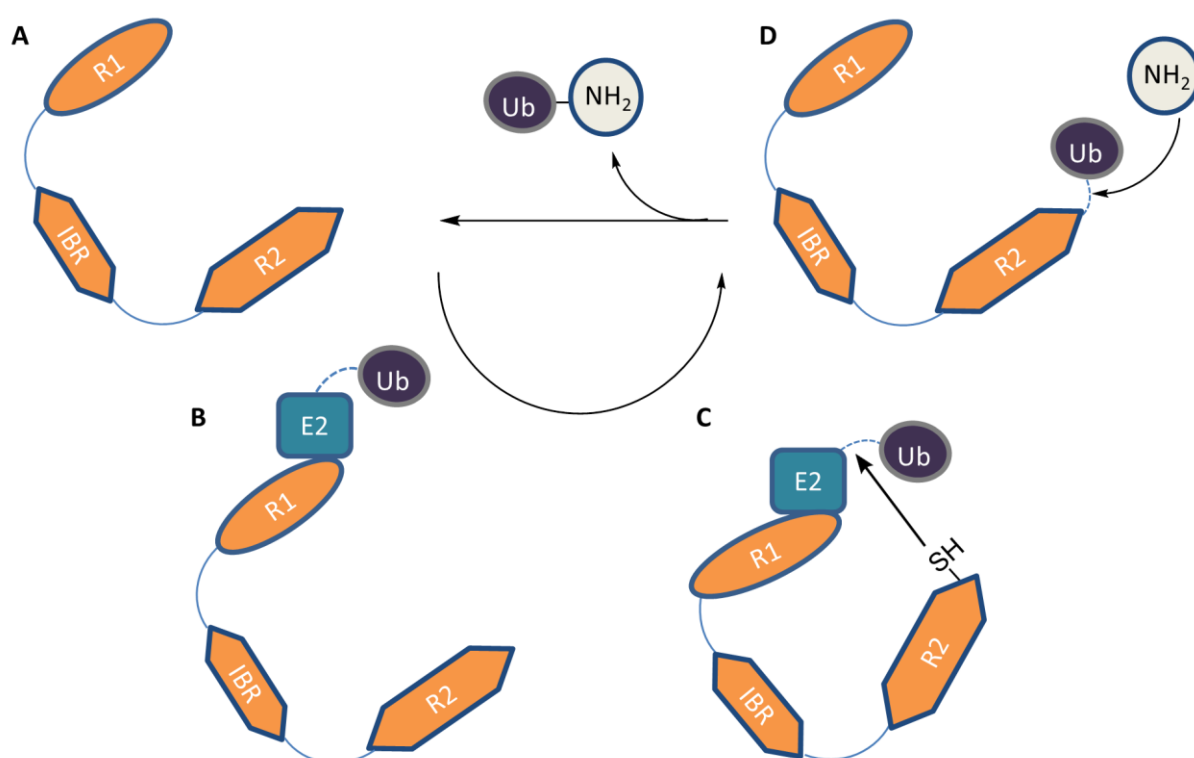
X-ray crystallography is a key structural tool used to study the interaction of proteins and the mechanism of enzymatic reactions.<sup>304</sup> Thereby it enables the identification of targetable pockets and grooves which are related to an enzyme's function. Further downstream, as attempted in Chapter 4, it allows characterisation of the binding mode of interacting compounds and facilitates structure driven optimisation of drug design.<sup>297, 305</sup> As Ub transfer events frequently involve transient complex formation it can be difficult to gain an understanding of the interactions that are occurring. Consequently, various strategies have been adopted to gain structural insights into these events.

One such method is to stabilise the transient intermediates involved in a complex. Plechanovova *et al.* studied the complex formed during Ub transfer from E2 to E3 by mutating the E2 catalytic Cys to a Lys residue, thereby converting the E2-Ub from a transient thioester to a more stable amide intermediate.<sup>37</sup> This method introduced only a small perturbation from the native complex, and enabled the visualisation of E2 to E3 Ub transfer, through an E2 'priming' mechanism for the first time. An alternative method of gaining insight into transient complexes is to covalently trap all interacting constituents. Although clearly a perturbation on the native complex, such models can provide novel insights into weakly explored interactions. Kamadurai *et al.* have previously reported the covalent trapping of the WW3 domain of the E3 HECT ligase Rsp5, with a peptidyl substrate mimic during Ub transfer.<sup>46</sup> As HECT ligases employ catalytic Cys residues to transfer Ub, it was possible to trap these nucleophiles using an electrophilic maleimide group. The use of a substrate fragment, and further removal of any non-catalytic cysteine by mutation, further simplified the model. None the less, new information could be drawn from such a model, and provided that the enzyme employs a catalytic Cys, the same methodology can in principle be applied to other enzyme families.

### 5.1.2 Studying RBR ligase complexes

As previously mentioned in Chapter 1, RBRs consist of two RING regions, RING1 and RING2, separated by a IBR region (Figure 63A).<sup>52</sup> The RBR family includes ligases of clinical interest such as parkin, whose dysfunction is linked to the early onset of Parkinson's disease<sup>306</sup>, and the aforementioned HOIP. The X-ray crystallographic studies conducted on HOIP by Stieglitz *et al.* identified a targetable pocket during transfer of Ub onto substrate (in the case of HOIP, an acceptor Ub), and thus supported its suitability for the fragment-based screen described in Chapter 4.

However, this only represents one possible point of disruption of RBR ligase function (Figure 63D). An alternative point to disrupt is the initial transfer of Ub onto the RING2 catalytic cysteine (Figure 63C). Though E2-Ub-RING1 docking has been previously reported for an RBR (Figure 63B), the complex formed during Ub transfer to RING2 (Figure 63C) is extremely transient and as such is poorly studied to date.<sup>307</sup> The development of a model system to provide structural insights into this latter complex would benefit two-fold: i) in aiding the mechanistic understanding of RBR ligase function and ii) in determining the tractability of this complex for future small molecule screens and ultimately drug design. Accordingly, a model system was developed as detailed below in collaboration with Katrin Rittinger (Crick Institute, Mill Hill).



**Figure 63: The RBR ligase mechanism and candidate points of disruption. (A)** RBRs consist of three major protein domains: RING1 (R1), RING2 (R2) and IBR. **(B)** Incoming Ub-E2 conjugate binds to R1. **(C)** Intermediary Ub transfer onto the catalytic cysteine of R2. **(D)** Subsequent ligation of Ub onto substrate. Theoretically, structural understanding of complexes B, C or D could be used to rationalise drug discovery approaches.

#### 5.1.2.1 HHARI

HHARI is an RBR ligase. It interacts with the E2 Ubch7 to form a stable complex and is known to facilitate ubiquitination (and hence regulation) of translation initiation factor EIF4E2, suggesting that it may play a role in protein translation.<sup>308</sup> Although both the Apo and E2-Ring1 bound structures of HHARI have been solved, to date the conformational changes that occur during Ub transfer from Ubch7 to HHARI-RING2, prior to transfer to substrate, remain unknown.<sup>307</sup>

Due to its relevance as an RBR ligase, and availability in Katrin Rittinger's laboratory, HHARI was chosen to test our capability to covalently trap an RBR-E2-Ub complex (Figure 63C). The following sections describe the development of a covalent trap to conduct these structural studies and its subsequent optimisation and evaluation.

## 5.2 Trap development

### 5.2.1 Trap design

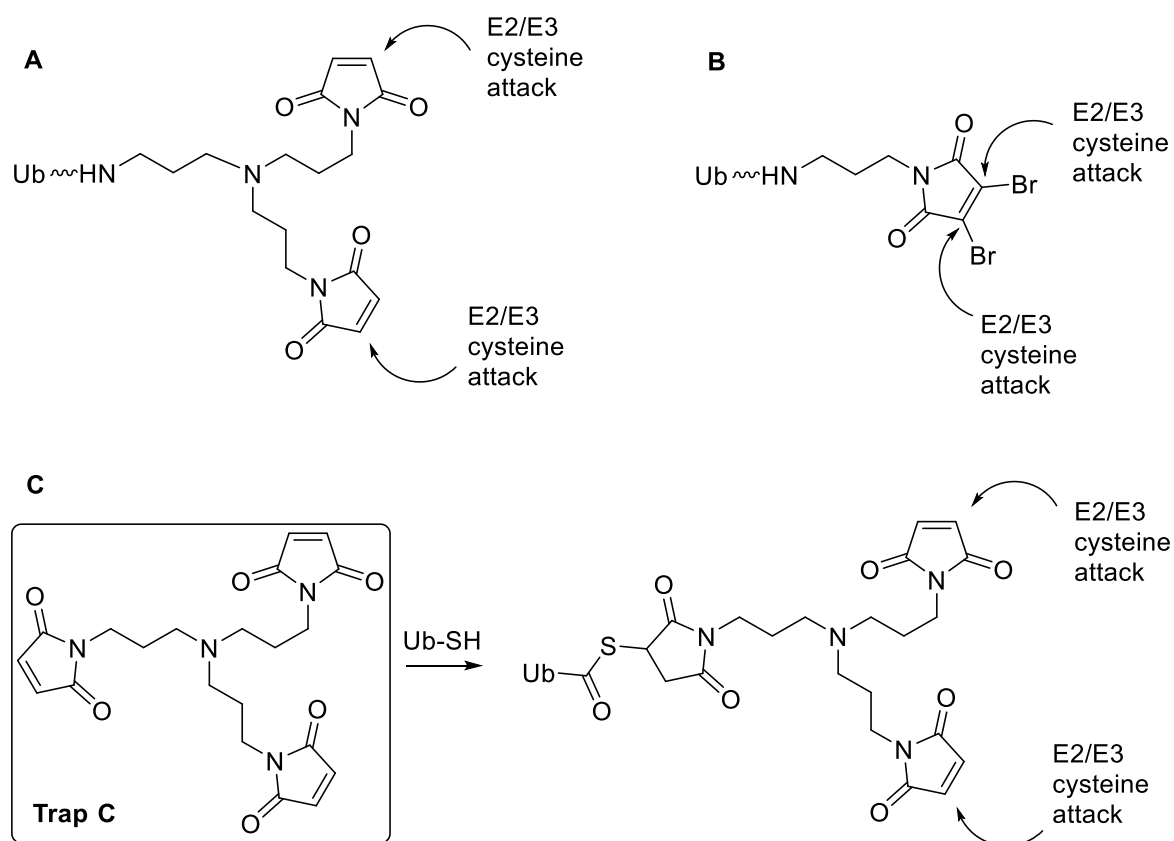
Due to its literature precedent for cysteine trapping, maleimide chemistry was chosen as a starting point for trap design.<sup>46</sup> Two alternative approaches were postulated: Mono- and Dual-maleimide trapping (Figure 64).

Complex isolation through a mono-trapping approach (Figure 64A), whereby each ligase is trapped on a separate maleimide, provides flexibility to enable sufficient movement of the complex components to adopt native conformations within the trapped complex. Alternatively a dual-trapping strategy (Figure 64B) allows covalent attachment of both ligases to the same moiety through the use of bromo-maleimides. Although dual trapping is potentially a powerful strategy, the retention of the double bond makes reversibility of trap binding possible. The trap design also provides minimal flexibility for ligase binders. On this basis it was decided that a mono-trapping approach would initially be employed. A three carbon linker was chosen, as this was deemed sufficiently flexible and was commercially available.

In order to simplify the enzyme trapping from a 4- to a 3-component reaction, it was decided that Ub should be preloaded onto the trap before introduction of any additional enzymes. Furthermore, it was hypothesised that pre-loading Ub, which is recognised by both E2 and E3 ligases, may assist subsequent enzyme binding in an orientation relevant to their function.

Classically, C-terminal modification of Ub is achieved by nucleophilic addition to Ub-thioester intermediates prepared through an intein-based chemical ligation method (introduced in Section 1.2.5.2).<sup>178</sup> This method is, however, associated with poor yields; negating its use was thus beneficial when considering trap design. Sortase tagging offered a potential method of Ub-trap ligation.<sup>309</sup> This method, however, requires the presence of a C-terminal LPXTG motif as a sortase recognition sequence, thereby adding a further four amino acids to the active site region of the trapped complex. It was decided that this level of modification would bring undesirable complexity to a model whose ultimate purpose is for structure elucidation. Ultimately it was decided that it was

easier to incorporate Ub into the trap using a C-terminal cysteine containing mutant (UbG76C). As with the ligases, this nucleophilic Ub mutant could be added to a maleimide moiety, leading to the design of a symmetrical tri-maleimide trap, **Trap C** (Figure 64C). It was postulated that mono-Ub addition could be achieved through controlled stoichiometric addition of Ub to the trap, prior to incubation with ligase enzymes.



**Figure 64: Postulated trap designs. (A) Di-maleimide trap, using a mono-trapping approach. (B) Bromo-maleimide trap, using a dual trapping approach. (C) Tri-maleimide trap, using a mono-trapping approach. Trap C was chosen for this study.**

### 5.2.2 Trap synthesis

The synthetic route of **Trap C** is outlined in Figure 65. Following a literature reported method, intermediate **29** was formed in 68 % yield by the addition of ethyl chloroformate to maleimide (**28**) at 5 °C in the presence of *N*-methylmorpholine.<sup>310</sup> Subsequently **Trap C** was assembled in 44 % yield by the addition of 3 equivalents of **29** to tris(3-aminopropyl)amine in the presence of NaHCO<sub>3</sub>. Sufficient purity for use in biochemical assays was achieved by silica chromatography.

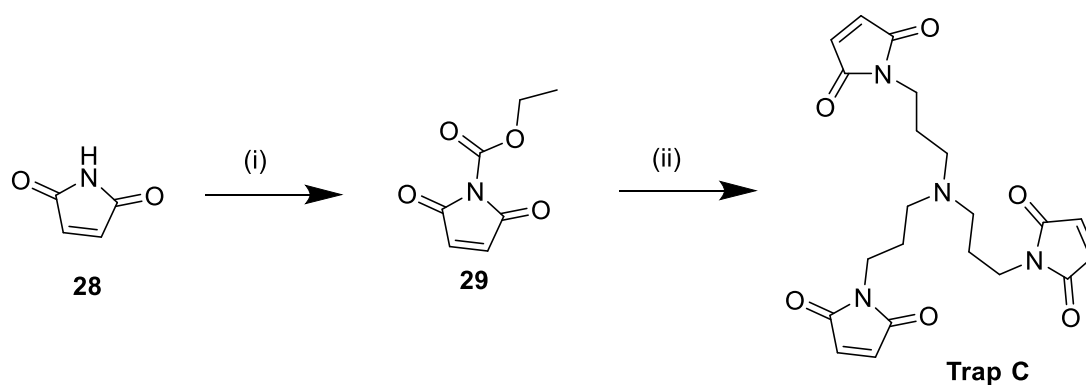


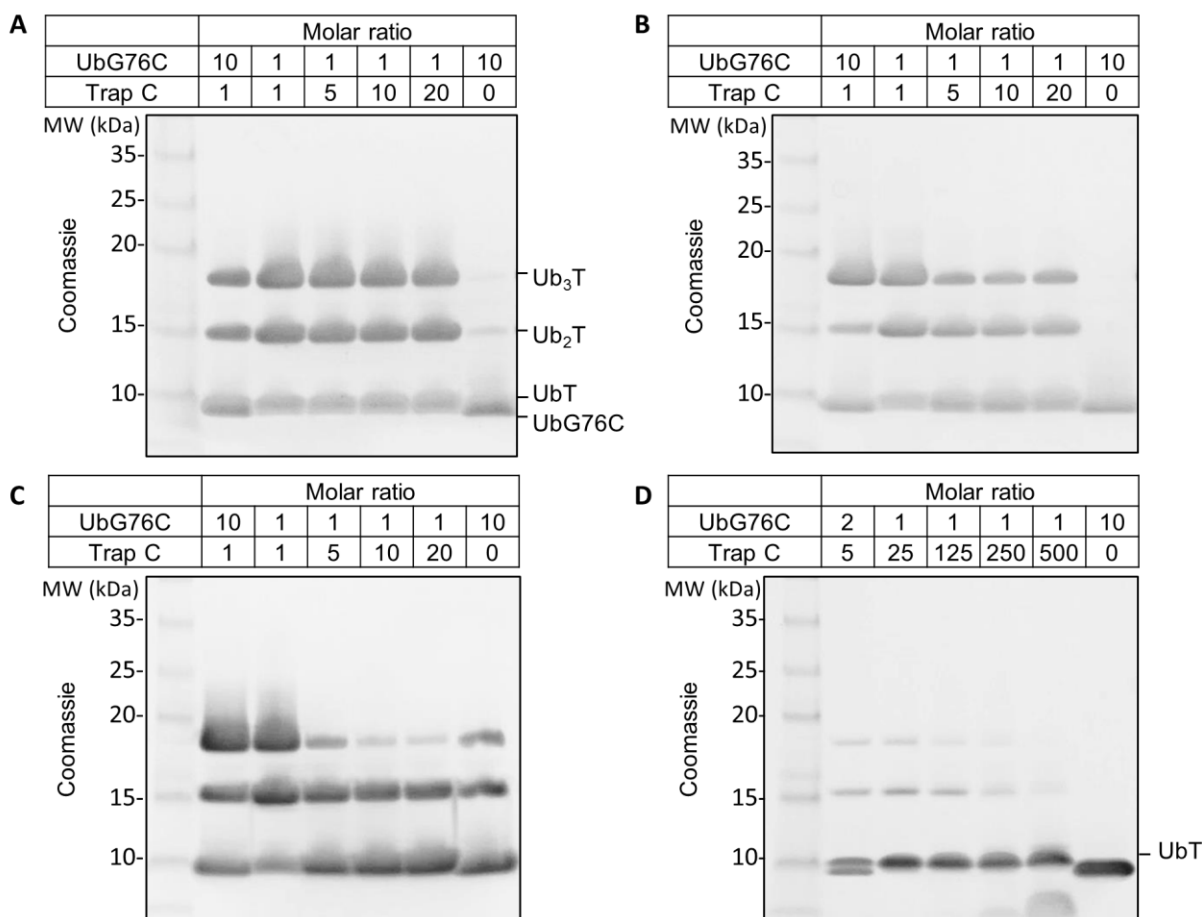
Figure 65. Tri-maleimide trap synthetic development: i) ethyl chloroformate, NMM, EtOAc, RT, 1 h, 68 %; ii) tris(3-aminopropyl)amine, NaHCO<sub>3</sub>, THF/H<sub>2</sub>O, 0 °C, 4 h, 44 %.

### 5.3 Towards a trapped Ub-Ubch7-HHARI complex

#### 5.3.1 Mono-Ub trap formation

With **Trap C** in hand, the feasibility of mono-protein addition to a tri-functional trap was determined with the help of Dr Nick Brown at the Crick Institute (Mill Hill). Initially UbG76C was added to **Trap C** in tris-HCl buffer (pH 8.5) and incubated in varying ratios for 10 minutes at room temperature before quenching by adding DTT and subsequent separation by SDS-PAGE (Figure 66A). UbG76C is clearly highly reactive with the trap with mono- (Ub-T), di- (Ub<sub>2</sub>-T) and tri- (Ub<sub>3</sub>-T) Ub trap complexes observed under these conditions, including with the highest molar excess of trap. The band pattern observed remained the same under both reducing and non-reducing SDS-PAGE running conditions (data not shown), indicating that the observed bands are due to covalent interaction with the trap.

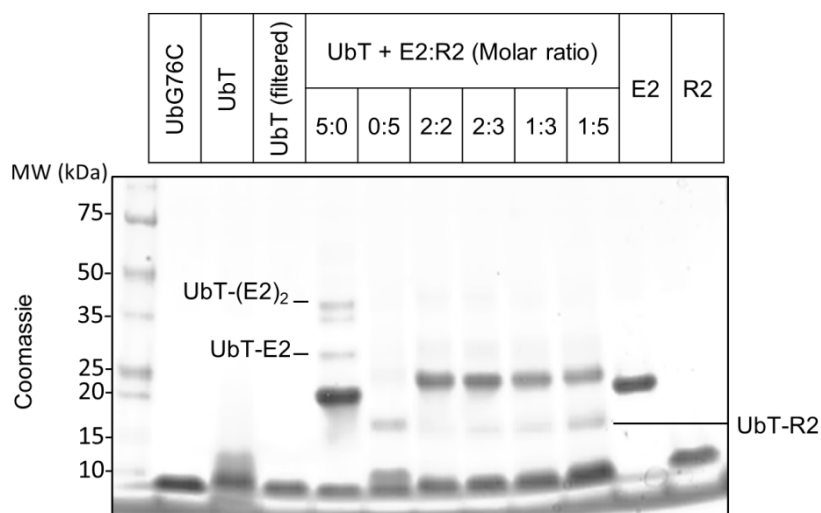
Changing the order of addition of reagents, such that **Trap C** was added to UbG76C, somewhat reduced the formation of (Ub<sub>3</sub>-T) (Figure 66B). It is hypothesised that this order of addition results in a high initial local concentration of trap relative to UbcG76C, which would favour lower order addition of UbG76C. Running the reaction in MBS buffer at a reduced pH of 6.5 further decreased the addition of multiple Ub species (Figure 66C). This improvement is most likely the result of reduced cross-reactivity of amine containing amino acids such as Lys at reduced pH due to protonation. Optimal conditions for mono-addition were ultimately achieved by performing the reaction on ice for 3 minutes with 500 fold trap excess (Figure 66D). The effect of reaction dilution was also examined, but was found to be less influential on the reaction than temperature (data not shown). Once generated, mono Ub-trap was successfully separated from excess **Trap C** by gel filtration.



**Figure 66: Optimisation of mono-Ub trap formation.** UBG76C (8.5 kDa) and Trap C were incubated at indicated molar ratios, quenched with DTT, separated by SDS-PAGE and analysed by Coomassie staining. (A) pH 8.5, Ub-to-trap addition, RT, 10 min. (B) pH 8.5, trap-to Ub addition, RT, 10 min. (C) pH 6.5, trap-to Ub addition, RT, 10 min. (D) pH 6.5, trap-to Ub addition, 0 °C, 3 min.

### 5.3.2 Trapping a minimal E3 domain

Having optimised Ub-T formation, ligase additions were attempted. Initially, addition of the isolated RING2 domain of HHARI (R2), with its E2 partner Ubch7, was attempted. R2 was chosen, as it only possesses one cysteine residue, providing a simpler model to determine the feasibility of ligase trapping. A pre-mixed solution of Ubch7 and R2 was added to filtered Ub-T on ice for 3 minutes before quenching by adding DTT and subsequent separation by SDS-PAGE (Figure 67). Additions of different ratios of Ubch7 and R2 were attempted. Separate control samples for excess Ubch7 and R2 with Ub-T were included to aid identification of bands.



**Figure 67: Preliminary Ub-T trap investigations with Ubch7 and R2. UBG76C and Trap C were incubated in a 1:500 molar ratio on ice for 3 min and filtered (7 kDa MWCO) before incubation with Ubch7 (E2) and R2 at indicated molar ratios. Samples were quenched with DTT, separated by SDS-PAGE and analysed by Coomassie staining. E2 (17.9 kDa) and R2 (7.3 kDa) control samples are included.**

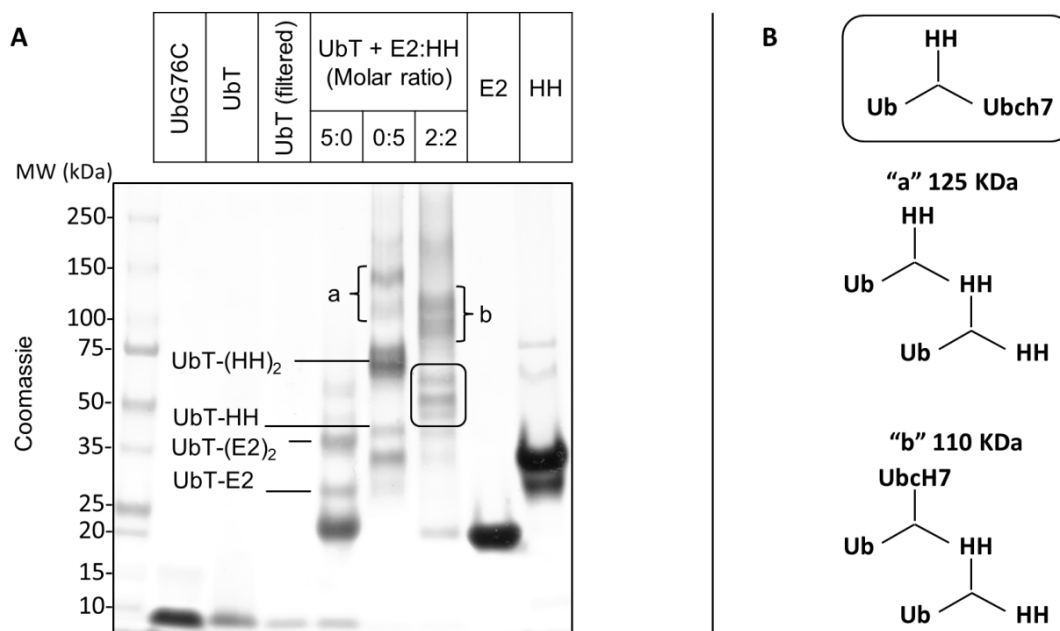
It was observed that, unlike UbG76C, neither Ubch7 nor R2 are particularly reactive with Ub-T when incubated in isolation. Some Ub-T-Ubch7 and Ub-T-(Ubch7)<sub>2</sub> formation is observed, but most of the E2 enzyme remains unreacted. Similarly, a low level of Ub-T-R2 formation is observed. On co-incubation, a band shift in the region of 25 kDa is observed. A complex of this mass could correspond to Ub-T-R2 (23 kDa), however, strikingly Ub-T levels remain unchanged. This band therefore does not seem to correspond to the formation of an Ub-T-ligase complex. Neither is it likely to be a direct E2-R2 heterodimer (~25 kDa), as this was not observed in a further control gel under either reducing or non-reducing conditions (Appendix Figure 2). An increased molar ratio of R2 led to a more intense Ub-T-R2 band, whilst a decreased molar ratio of Ubch7 led to a modest decrease in the intensity of the ‘product’ band. This finding suggests that the 25 kDa could simply be a gel running effect on Ubch7. Although all incubations were repeated in various ratios and orders of enzyme addition, the observed band pattern did not change; this is not surprising, considering the previously observed low reactivity of the proteins with the trap, and further suggests that the newly observed is Ub-T independent. In any case, this model appears insufficient due to a weak reactivity of both ligase components with Ub-T.

### 5.3.3 Trapping a larger HHARI fragment

As R2 was found to be poorly reactive with Ub-T, an alternative HHARI fragment, HH (HH94-397 (W379A)), was tested in combination with Ubch7 to scope the feasibility of ligase trapping. As HH and Ubch7 were known to form a stable complex in solution measurable by both Octet and ITC (Nick Brown, Crick Institute, direct communication), it was hypothesised that co-incubation would also



circumvent issues due to the low activity of Ubch7 in isolation with Ub-T. A pre-mixed 1:1 solution of Ubch7 and HH were added to Ub-T on ice for 3 minutes before quenching with DTT and separation by SDS-PAGE (Figure 68A). Separate control samples for excess Ubch7 and HH with Ub-T were included to aid identification of bands.



**Figure 68. Preliminary trap investigations with Ubch7 and HH. (A)** UBG76C and Trap C were incubated in a 1:500 molar ratio on ice for 3 min and filtered (7 kDa MWCO) before incubation with Ubch7 (E2) and HH at indicated molar ratios. Samples were quenched with DTT, separated by SDS-PAGE and analysed by Coomassie staining. E2 and HH (35.6 kDa) control samples are included. **(B)** Postulated structures attributed to bands in indicated higher MW regions.

Encouragingly, in isolation HH proved to be much more reactive than R2 towards Ub-T; minimal Ub-T remains and both Ub-T-HH and Ub-T-(HH)<sub>2</sub> formation is observed. Additionally, higher mass complexes were observed. It is hypothesised that these correspond to cross reactivity of HH, which possesses three non-catalytic cysteine residues (Figure 68B). After co-incubation, minimal HH remains and a new band that corresponds to the formation of a putative Ub-T-Ubch7-HH complex (63 kDa) appears. Some Ub-T-HH, minimal UbT-HH<sub>2</sub> and excess Ubch7 also remain. The persistence of isolated Ubch7 in the presence of Ub-T-HH indicates that not all HH and Ubch7 form a stable complex. It is postulated that the higher mass bands correspond to species resulting from cross reactivity of non-catalytic cysteines. These bands persist under reducing and non-reducing conditions (data not shown). In conclusion, although the purity of product formation needs to be addressed, maleimide traps are a viable method of isolating a RBR-E2-Ub transient complex.

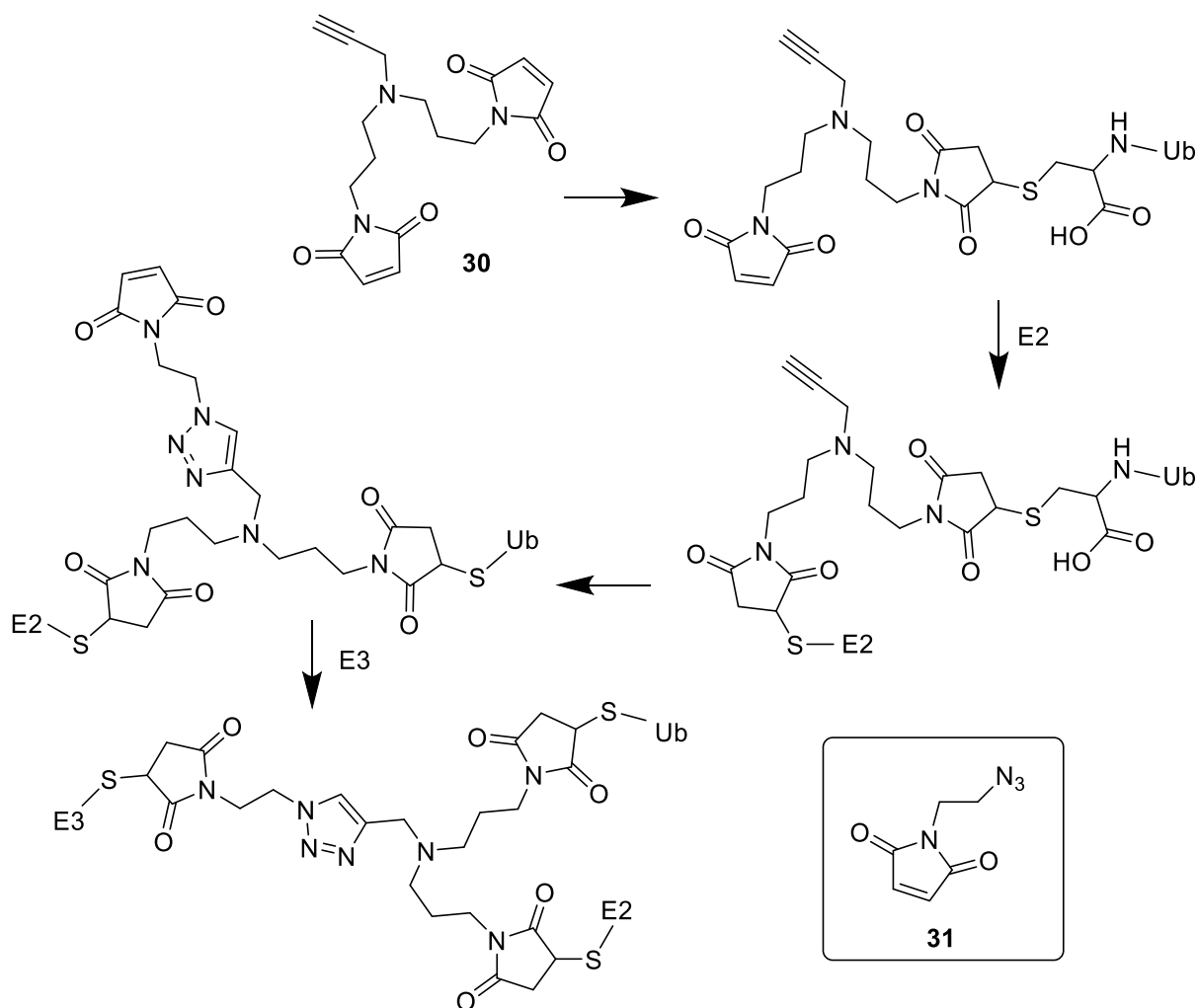
## 5.4 Conclusions / future work

A tri-maleimide trap was successfully synthesised using standard synthetic methods. Reaction conditions were optimised to enable selective formation and purification of a mono-ubiquitinated trap (Ub-T). Primary tests which employed the minimal E3 domain, R2, and Ubch7 indicated that the reaction efficiency with Ub-T is low in the case of both proteins. A larger fragment of HHARI, HH, did prove to be more reactive towards Ub-T in isolation and when added in conjunction with Ubch7. Although a putative band that proves the formation of Ub-T-Ubch7-HH was observed under these conditions, additional higher mass bands could be detected. One explanation for these bands is the cross-reactivity of additional cysteine residues present in Ubch7 and HH leading to multi-trap species.

Though these findings provide a promising start, the model complex needs to be formed exclusively so that it can be purified with adequate purity for crystallisation purposes. In order to achieve this objective the following issues need to be addressed:

*i) Additional reaction products need to be minimised.* The use of cysteine free HH and Ubch7 mutant should significantly decrease the amount of side products. A cysteine free Ubch7 mutant has been successfully developed by Katrin Rittinger's group; however, the preparation of a full mutant of HH has been proved troublesome to date. Alternatively, trap chemistry could be adjusted to provide increased control for sequential addition of complex components. The use of clickable entities could be a viable route to allow the controlled addition of maleimide groups to the trap before each protein addition, as postulated in Figure 69. An alkyne containing maleimide trap (**30**) has been previously reported in the literature, and could be used in an analogous fashion to **Trap C** to trap Ub and a ligase.<sup>46</sup> Subsequent CuAAC ligation to an azide containing maleimide (**31**), would result in the introduction of the final maleimide group thereby enabling the controlled addition of another ligase.

*ii) Amount of remaining substrate needs to be minimised.* Residual substrate should be reduced by ensuring that the UBCH7-HH complex is formed quantitatively. As HH and Ubch7 are known to form a stable complex in solution, this should reduce the reaction to secondary order. An excess of Ub-T could be also be used with respect to the incoming enzyme complex as this presents a large mass difference (~55 kDa) which can be separated during purification. However, this is only applicable if cysteine free mutants become available.



**Figure 69: Postulated application of clickable maleimide trap entities, to enable controlled addition of ligases.**

As soon as these purity issues are addressed, the complex can be isolated for the first time followed by first crystallisation attempts.

## Chapter 6 Conclusions and future work

---

The main aim of this PhD project was to develop chemical tools to assist the study of ubiquitination machinery. This Chapter highlights the key outcomes of this work and discusses the broader implications for our understanding of ubiquitination machinery and their study. Future directions are also proposed.

### 6.1 ABPs to study ubiquitination machinery

#### 6.1.1 Conclusions

ABPs derived from both peptide and small molecule based inhibitors were designed and successfully synthesised (Chapter 2 and Chapter 3).

Peptide derived ABPs showed initial promise as pan Ub machinery probes in HeLa cell lysate. Both **Pg-12mer-VME** and **Pg-12-mer-VS** were promiscuous labellers, with little difference observed between the different electrophilic traps tested. **Pg-12mer-VME** was much more reactive than its Ub-derived ABP analogue, and was capable of labelling a number of E3 ligases in MCF7 cell lysate. However, on quantitative proteomic analysis of **Pg-12mer-VME**, few Ub enzymes were identified as competitive targets of this ABP. Following the development of a DDR model in MCF7 cells, application of **Pg-12mer-VME** in this model was inconclusive; though some variation in labelling was observed, none of the targets had been previously identified as competitive probe targets.

Although initial application of peptide derived probes in-cell revealed that the probes were cell permeable and that the observed labelling could be out competed by parent inhibitor, on proteomic analysis no ubiquitination machinery was observed. Further in-cell fluorescent studies with TAMRA-labelled analogues suggest that the 12-mer peptide sequence broadly localises across the cell, whilst alternative peptide sequences, such as 15-mer, localise to distinct regions. Though the addition of NLS or cell penetrating peptide sequences could provide tailored ABP cellular localisation, such additions proved detrimental to probe labelling capability as observed by in-gel fluorescence.

Small molecule derived ABPs were based on two different inhibitors: **PYR-41** and **Tz9**. Like the peptide derived probes, **PYR-41-P** was a promiscuous labeller which showed initial promise as a pan Ub machinery probe in HeLa cell lysate. However, its cellular applications as an Ub probe again

proved difficult: although some Ub machinery was labelled in MCF7 cells, none were identified as competitive targets on quantitative proteomic analysis of this ABP.

**T2** was determined to be the strongest labelling **Tz9** probe analogue, and was taken forward to further study. However, although on reaction with cysteine residues **T2** is expected to form thioester bonds, relative probe labelling observed by SDS-PAGE was not affected by reducing conditions, suggesting that **T2** is mainly not labelling via cysteine residues. LC-MS/MS analysis of **T2** labelling in cell lysate was not determined, but some Ub machinery was identified on its application in-cell. However unlike the other ABPs evaluated, the reported target of **Tz9**; RAD6B, was not identified by **T2**. Quantitative proteomic analysis was compounded by the lack of competitive labelling observed in the presence of **Tz9**, or alternative inhibitor analogue **T2c**. Further comparison to alternative electrophilic probes studied by members of the Tate group allowed the identification of generic targets of electrophilic probes. In this way the differing electrophilic profiles of **PYR-41-P** and **T2** were evaluated.

Literature reported Ub-derived probes enable the robust identification of multiple DUBs in cell lysates. Alternative Ub-derived ABPs, such as the Ub-E2 conjugate ABP probes reported by Pao *et al.* have recently been shown to effectively label RBR ligases in lysate models.<sup>184</sup> Collectively, our work has demonstrated probe labelling beyond the DUBs and RBR ligases currently assessable with these probes. However, all the probes studied are promiscuous, and have proven difficult to evaluate robustly due to a lack of significant competition observed for this ubiquitination machinery against the parent inhibitors. The observed discrepancy between in lysate and in cell labelling has further implications for the in-lysate study of ubiquitination enzymes; it would be interesting to see what the labelling capability of Ub-derived probes is in cell. In addition, it is clear that the general activity of electrophilic species in cell needs to be taken into consideration when designing ABP competition experiments. This follows from the observed quenching of **PYR-41-P** cellular labelling when treated in DMEM, and the postulated shielding of probe activity following NEM incubation.

Taken together, our findings suggest that promiscuous probes may not be generally suited for this type of study, but that the pursuit of cellular ABPs for ubiquitination machinery remains relevant, due to the disparate labelling observed in cell and in lysate. Generic electrophilic probes, such as IA-alkyne<sup>193, 194</sup>, do not strongly enrich for Ub machinery, and the promiscuous ABPs reported in this study do not offer much improvement on this for cellular quantitative proteomic studies. Indeed, the Ub machinery identified by **T2** and **PYR-41-P** were frequently labelled by other electrophilic Tate

group probes. Therefore, the design of more selective probes, which retain their activity in cell would be more beneficial.

## 6.1.2 Future work

### 6.1.2.1 General considerations

- Explore alternative inhibitor scaffolds, which may lead to selective, cell permeable Ub-ABPs. Several alternative literature inhibitors exist, which offer alternate (reported) reactivity and selectivity. In particular, the small molecule Heclin could be pursued.<sup>227</sup> Hectlin was reported to broadly inhibit HECT E3 ligases in cells. As His-tagged HECT ligases were expressed in HEK293 for this study, however, the additional targets of Heclin are unknown. Bay 11-7821 could also be pursued.<sup>214</sup> *In vitro* studies with recombinant ubiquitination enzymes revealed that this small molecule inhibits 24 E2 ligases. If this inhibition translates into cells, Bay 11-7821 may provide a starting point for an E2 ABP probe. Without comprehensive target identification, however, the promiscuous nature of such probes remains elusive.
- Increase confidence in identified ABP targets. This can be achieved through the application of a trypsin-cleavable capture reagent, which enables both the detection of modified peptides, and the identification of the modified amino acid residues. In this way, labelling of catalytic Cys can be determined. This will be particularly useful for the analysis of **T2** labelling, which appears to significantly label non-Cys residues. Additional validation of ABP targets identified by LC-MS/MS can be undertaken by western blot analysis.
- Attempt alternative, 'label-free' methods of probing inhibitor scaffolds. As competition against the parent inhibitor was not always possible or significant, CETSA could be employed as an alternative method of probing the targets **Tz9** and **PYR-41**.
- Explore the networks generated from the competitive targets of each probe in further detail. This analysis could lead to testing the probes in alternative cell lines of interest, or in alternative biological models.
- Complementary to ABP profiling, Ub profiling using a K- $\epsilon$ -GG antibody enrichment strategy would enable the identification of the downstream effects of the parent inhibitors, which could be linked back to ubiquitination machinery, and further validated by western blot analysis.

### 6.1.2.2 Probe-specific considerations

- Although the impact of additional NLS or cell penetrating peptide sequences on probe labelling in cell was studied and found to negatively impact on probe labelling, their cellular localisation could still be elucidated. If these peptides provide an interesting profile, the efficiency of CuAAC for these probes could be analysed by MS, and if necessary alternative bioorthogonal labelling strategies attempted.
- Although no ubiquitination machinery was identified by **Pg-12-mer-VME** in cell, this could be further confirmed by maximising the coverage of the LC-MS/MS analysis. This could be achieved by cellular fractionation into cytosolic, nuclear and membrane components following probe treatment, and separate LC-MS/MS analysis of these fractions. Additional fractionation of these samples into multiple LC-MS/MS runs would increase the coverage of low abundance hits.
- Alternative, less promiscuous peptide probes could be designed. This could be achieved with an alternative peptide sequence. For example, labelling by the 15-mer peptide sequence, or longer peptides could be explored by LC-MS/MS. Alternatively, peptide scaffolds, which mimic the 3D architecture of Ub around the flexible C-terminal tail and therefore should be more selective to ubiquitination machinery, could be designed. However, cell permeability should remain a consideration. In addition, the use of alternative electrophilic traps could be explored. It is postulated that reducing trap reactivity would lead to less promiscuous probes; however this presents a difficult balancing act, between reduced promiscuity and retained reactivity with E2 and E3 ligase enzymes.
- Explore the NEM inductive effects observed for **PYR-41-P** by in-gel fluorescence by LC-MS/MS analysis. In addition, explore the NEM competitive probe targets for **T2**.

## 6.2 FBDD and structural models for RBR E3 ligases

### 6.2.1 Conclusions

A commercial library of fragments was screened by DSF against P37, the minimal catalytic core of the RBR ligase HOIP (Chapter 4). Several hit fragments were identified in this primary screen, displaying linear, saturated, or inverted concentration dependences. A series of hits were subjected to validation by waterLOGSY NMR spectroscopy, and further characterisation by MST. In this way, two low mM confirmed hits: **26E11** and **18H9** were identified for further study, with **18H9** determined to be a covalent binder. X-ray crystallography of fragment bound P37 was attempted, however, crystallisation attempts have proved unfruitful to date, preventing the further development of these fragments. Although the structural characterisation of these fragments remains to be seen, these fragments remain putative starting points for HOIP inhibitor design, and could be applied in the further development of selective HOIP, or pan RBR probe design.

Though this work has successfully identified two fragments which bind to P37, it has also highlighted some issues with the FBDD methodology employed. Although the primary hit rate following the DSF screen was reasonable for a fragment screen, the majority of tested hits were not confirmed by waterLOGSY NMR. However, subsequent characterisation of one such fragment, **26E11**, by MST revealed a mM range hit fragment. This suggests that other hits should not be unduly discarded before further testing by alternative validation methods.

In addition, a tri-maleimide trap was designed and synthesised to enable the covalent trapping, and ultimately structural analysis, of transient RBR E3 ligase complexes. Structural insight into such complexes furthers our mechanistic understanding and enables the 'druggability' of these complexes to be evaluated. Conditions to enable mono-ubiquitination of the trap were optimised, and the putative formation of a trapped Ub-E2-E3 complex observed. This work represents the starting point for further development of an Ub-E2-E3 complex structural model. The high reactivity of maleimide was tempered by lowering pH conditions to favour Cys additions. However, cleanly forming complexes with proteins possessing multiple Cys residues proved problematic and limited the tractability of this approach for crystallographic studies.

Since starting this work, The E2-R2-Ub complex of HOIP has been solved by Lechtenberg *et al.*<sup>47</sup> the authors utilised a Cys to Lys mutation to provide a stable E2-Ub conjugate which then crystallised with HOIP in the presence of free ubiquitin. This presents the first structural information for



an active RBR-E2-Ub complex. Once optimised, crystallisation of our model complex will provide further, complementary information for the study of RBR ligase catalysis.

## 6.2.2 Future work

### 6.2.2.1 FBDD

- Validate fragments using alternative methods, such as STD NMR spectroscopy, to increase the confidence of hit fragments.
- Further characterise fragments by ITC, which does not require a labelled protein analogue, to complement the current characterisation data collected by MST. In addition, conducting the linear ubiquitination assay employed by Katrin Rittinger's laboratory in the presence and absence of fragment, will determine the biochemical effect of the fragment hits on P37 activity.
- Optimise crystallisation conditions for P37 to enable the structural characterisation of hit fragments. Alternatively, if the <sup>15</sup>N 2D NMR of P37 is assigned, binding sites could be determined by protein detected NMR.
- Test the hit fragments on alternative RBR ligases, to determine their selectivity, and potential scope for the development of an RBR selective probe compounds from these fragments.
- Elucidate the binding site of fragment **18H9** with P37, by MALDI MS analysis of tryptic peptides. If truly catalytic cysteine binding, the application of fragment **18H9** as an alternative E3 ligase reactive electrophilic trap could be explored. The kinetics of **18H9** are slow however; this would require further study and optimisation, and the selectivity of the interaction remains to be seen.

### 6.2.2.2 Structural models

- Conditions to enable the exclusive formation of the desired Ub-E2-E3 complex, with minimal by-products and residual starting materials require further optimisation. An alternative approach would be the controlled, step-wise addition of each ligase. Both of these approaches as discussed fully in Chapter 5.

# Chapter 7 Materials and Methods

---

## 7.1 Chemical synthesis

### 7.1.1 General methods

All solvents and reagents were purchased from Sigma-Aldrich, Alfa Aesar, or VWR unless otherwise stated, and used without further purification. Moisture sensitive reactions were performed in oven dried flasks, under a nitrogen or argon atmosphere. Anhydrous solvents were dispensed using Pure Solv™ solvent drying towers (Innovative Technology Inc.) Brine refers to a saturated solution of sodium chloride.

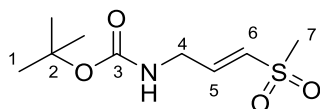
Analytical thin layer chromatography was carried out using Merck Si<sub>60</sub>, F<sub>254</sub> chromatography sheets. Spots were visualised by UV light or through use of an appropriate stain (iodine, ninhydrin, or potassium permanganate). Flash column chromatography was carried out manually on Merck 60 Å silica gel, eluting with solvents as supplied under a positive air pressure, or run on a Biotage Isolera™ One flash purification system using a wet-loading Biotage SNAP cartridge.

High resolution Mass spectra were acquired by the Imperial Mass Spectrometry service with *m/z* values reported in Daltons. Analytical and preparative LC-MS experiments were performed using a Waters HPLC system, consisting of a 2767 autosampler, 515 pump, 2998 photodiode array detector, and a 3100 electrospray ionisation (ESI) mass spectrometer. The system was equipped with Waters X Select C18 columns, running linear gradients of MeOH / H<sub>2</sub>O (0.1 % formic acid). Flow rates of 1.2 mL/min and 20 mL/min were used in analytical and preparative mode respectively. Water was removed using a Christ alpha 2-4 LD lyophiliser.

<sup>1</sup>H and <sup>13</sup>C NMR spectra were recorded on a Bruker Av-400 (400 Hz) or Av-500 (500 Hz) instrument at room temperature (RT) using deuterated solvents as a reference for internal deuterium lock. Chemical shift data is given as  $\delta_{C/H}$  in units of parts per million (ppm) relative to tetramethylsilane (TMS), where  $\delta$  (TMS) = 0.00 ppm. The multiplicity of each signal is indicated by: s = singlet; bs = broad singlet; d = doublet; t = triplet; q = quartet; m = multiplet. Coupling constants (*J*), calculated using MestReNova® NMR software, are quoted in Hz and recorded to the nearest 0.1 Hz. Spectra are included in the Appendices for reference.

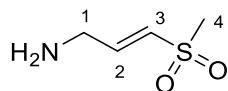
Microwave reactions were undertaken in a Biotage Initiator.

### 7.1.2 *Tert*-butyl[(2*E*)-3-(methylsulfonyl)prop-2-en-1-yl]carbamate (**2**)



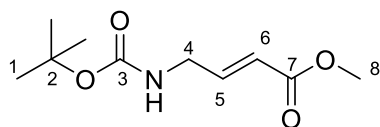
The following procedure was adapted from a previously reported method.<sup>177</sup> To a suspension of anhydrous  $K_2CO_3$  (285 mg, 2.06 mmol, 1.5 eq) in anhydrous THF (10 mL), diethyl [(methylsulfonyl) methyl]phosphonate (475 mg, 2.06 mmol, 1.5 eq) was added. The mixture was stirred for 1 hr at RT before a solution of *N*-Boc-2-aminoacetaldehyde (219 mg, 3.13 mmol, 1.0 eq) in anhydrous THF (2 mL) was added. The reaction was stirred overnight at RT, then quenched with  $H_2O$  (10 mL). THF was removed *in vacuo* and the crude material extracted with DCM (15 mL). The organic layer was washed with 2 % HCl (15 mL), 10 %  $NaHCO_3$  (15 mL), dried over  $MgSO_4$  and concentrated *in vacuo*. Subsequent purification by silica gel chromatography (hexane.: EtOAc, 1:1) afforded **2** (170 mg, 53 %) as a pale yellow oil:  $R_f$  0.2 (hexane / EtOAc, 1:1);  $^1H$  NMR (400 MHz,  $CDCl_3$ )  $\delta_H$  6.94 (dt,  $J = 15.1$ , 4.3 Hz, 1H,  $C_5H$ ), 6.53 (dt,  $J = 15.1$ , 2.0 Hz, 1H,  $C_6H$ ), 4.91 – 4.86 (bs, 1H, NH), 4.08-3.85 (m, 2H,  $C_4H$ ), 2.97 (s, 3H,  $C_7H$ ), 1.48 (s, 9H,  $C_1H$ );  $^{13}C$  NMR (101 MHz,  $CDCl_3$ )  $\delta_C$  155.5 ( $C_3$ ), 145.2 ( $C_5$ ), 129.6 ( $C_6$ ), 80.3 ( $C_2$ ), 42.9 ( $C_7$ ), 40.7 ( $C_4$ ), 28.3 ( $C_1$ ); LC-MS  $m/z$  ( $ES^+$ ) 258 [ $M+Na$ ] $^+$ , 80 %; 234 [ $M-H$ ] $^-$ , 100 %).

### 7.1.3 (2*E*)-3-(methylsulfonyl)prop-2-en-1-amine trifluoroacetate (**VS**)



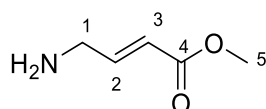
This compound was prepared following a previously reported method.<sup>177</sup> To a reaction flask charged with **2** (23.5 mg, 0.10 mmol 1.0 eq) a 33 % solution of TFA in DCM (4 mL) was added. The reaction was stirred for 1 hr at RT and then concentrated *in vacuo*. The crude residue was resuspended and co-evaporated with anhydrous toluene ( $3 \times 4$  mL) to afford **VS** (28.5 mg, quant.) as a pale yellow oil. The product was reacted immediately without further purification  $^1H$  NMR (400 MHz, MeOD)  $\delta_H$  7.05 – 6.82 (m, 2H,  $C_2H$  /  $C_3H$ ), 3.84 (dd,  $J = 5.2$ , 1.2 Hz, 2H,  $C_1H$ ), 3.02 (s, 3H,  $C_4H$ );  $^{13}C$  NMR (101 MHz, MeOD)  $\delta_C$  137.3 ( $C_2$ ), 133.8 ( $C_3$ ), 41.1 ( $C_4$ ), 38.7 ( $C_1$ ); LC-MS  $m/z$  ( $ES^+$ ) 136 [ $M+H$ ] $^+$ , 100 %; HRMS found 153.0692 [ $M+NH_4$ ] $^+$ , expected 153.0698.

### 7.1.4 *tert*-butoxy[(*2E*)-4-methoxy-4-oxobut-2-en-1-yl]oxoammonium (**3**)



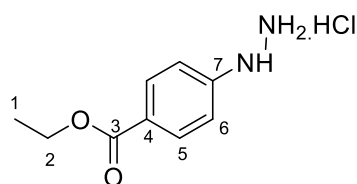
The following procedure was adapted from a previously reported method.<sup>177</sup> To a suspension of anhydrous  $K_2CO_3$  (651 mg, 4.71 mmol, 1.5 eq) in anhydrous THF (10 mL), methyl diethylphosphonoacetate (863  $\mu$ L, 4.71 mmol, 1.5 eq) was added. The mixture was stirred for 1 hr at RT before a solution of *N*-Boc-2-aminoacetaldehyde (500 mg, 3.13 mmol, 1.0 eq) in anhydrous THF (5 mL) was added dropwise. The reaction was stirred overnight at RT, then quenched with  $H_2O$  (10 mL). THF was removed *in vacuo* and the crude material was extracted with DCM (15 mL). The organic layer was washed with 2 % HCl (15 mL), 10 %  $NaHCO_3$  (15 mL), dried over  $MgSO_4$  and concentrated *in vacuo*. Subsequent purification by silica gel chromatography (hexane : EtOAc, 4:1) afforded **3** (453 mg, 66.5 %) as a colourless oil:  $R_f$  0.4 (hexane / EtOAc 3:1);  $^1H$  NMR (400 MHz,  $CDCl_3$ )  $\delta_H$  6.94 (dt,  $J = 15.7, 4.8$  Hz, 1H,  $C_5H$ ), 5.97 (dt,  $J = 15.7, 1.9$  Hz, 1H,  $C_6H$ ), 4.74 (s, 1H, NH), 4.04 – 3.87 (m, 2H,  $C_4H_2$ ), 3.76 (s, 3H,  $C_8H_3$ ), 1.47 (s, 9H,  $C_1H_3$ );  $^{13}C$  NMR (101 MHz,  $CDCl_3$ )  $\delta_C$  166.6 ( $C_7$ ), 154.1 ( $C_3$ ), 145.1 ( $C_5$ ), 120.9 ( $C_6$ ), 80.0 ( $C_2$ ), 51.7 ( $C_8$ ), 41.3 ( $C_4$ ), 28.4 ( $C_1$ ); LC-MS  $m/z$  ( $ES^+$ ) 238 [ $M+Na$ ], 80 %, 216 ( $[M+H]^+$ , 50 %).

### 7.1.5 (*2E*)-4-methoxy-4-oxobut-2-en-1-aminium tosylate (VME)



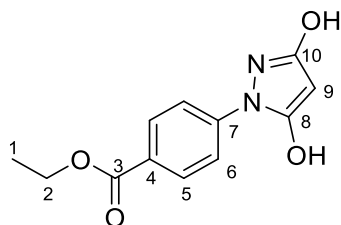
This compound was prepared following a previously reported method.<sup>178</sup> *p*-Toluenesulfonic acid was dehydrated following a literature procedure.<sup>311</sup> A solution of *tert*-butoxy[(*2E*)-4-methoxy-4-oxobut-2-en-1-yl]oxoammonium **3** (21.5 mg, 0.1 mmol, 1.0 eq) and anhydrous *p*-toluenesulfonic acid (19.0 mg, 0.10 mmol, 1.0 eq) in anhydrous  $Et_2O$  (2 mL) was stirred overnight at RT. Concentration *in vacuo* afforded **VME** (28.4 mg, quant.) as a colourless oil. The product was reacted immediately without further purification;  $^1H$  NMR (400 MHz, MeOD)  $\delta_H$  7.75 – 7.66 (m, 4H, TsOH), 7.24 (d,  $J = 8.0$  Hz, 4H, TsOH), 6.97 – 6.86 (m, 1H,  $C_2H$ ), 6.16 (dt,  $J = 15.9, 1.7$  Hz, 1H,  $C_3H$ ), 3.80-4.74 (m, 5H  $C_5H_3 / C_1H_2$ ), 2.37 (s, 6H, TsOH);  $^{13}C$  NMR (101 MHz, MeOD)  $\delta_C$  167.1 ( $C_4$ ), 143.4(TsOH), 141.8 ( $C_2$ ), 139.8 (TsOH), 129.8 (TsOH), 126.9 (TsOH), 125.9 ( $C_3$ ) 52.4 ( $C_5$ ), 40.9 ( $C_1$ ), 21.3 (TsOH); LC-MS  $m/z$  ( $ES^+$ ) 116 [ $M+H$ ] $^+$ , 100 %; HRMS found 116.0704 [ $M+H$ ] $^+$ , expected 116.0712.

### 7.1.6 Ethyl 4-hydrazinylbenzoate hydrochloride (7)



This compound was prepared following a previously reported method.<sup>268</sup> To a suspension of 4-hydrazinobenzoic acid (1.00 g, 6.57 mmol, 1.0 eq) in absolute ethanol (30 mL) thionyl chloride (884 mg, 539  $\mu$ L, 7.43 mmol, 1.1 eq) was added dropwise at 0 °C. The mixture was allowed to warm to RT, before being refluxed for 16 hr. The reaction was then cooled and diluted with Et<sub>2</sub>O (30 mL). The resulting precipitate was isolated by suction filtration, washed with Et<sub>2</sub>O and dried *in vacuo* to afford **7** (1.04 g, 73 %) as a lustrous colourless solid: R<sub>f</sub> 0.4 (hexane / EtOAc 3:1); <sup>1</sup>H NMR (400 MHz, MeOD)  $\delta_{\text{H}}$  8.03 – 7.90 (d, *J* = 8.8 Hz, 2H, C<sub>5</sub>H), 7.05 – 6.88 (d, *J* = 8.8 Hz, 2H, C<sub>6</sub>H), 4.30 (q, *J* = 7.1 Hz, 2H, C<sub>2</sub>H<sub>2</sub>), 1.33 (t, *J* = 7.1 Hz, 3H, C<sub>1</sub>H<sub>3</sub>); <sup>13</sup>C NMR (101 MHz, MeOD)  $\delta_{\text{C}}$  167.6 (C<sub>3</sub>), 150.5 (C<sub>7</sub>), 132.2 (C<sub>5</sub>), 125.1(C<sub>4</sub>), 114.2 (C<sub>6</sub>), 62.0 (C<sub>2</sub>), 14.6 (C<sub>1</sub>); LC-MS *m/z* (ES<sup>+</sup>) 181 ([M+H]<sup>+</sup>, 75 %).

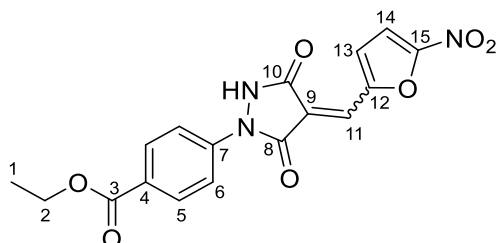
### 7.1.7 Ethyl 4-(3,5-dioxypyrazolidin-1-yl)benzoate (5)



The following procedure was adapted from a previously reported method.<sup>269</sup> **7** (1.00 g, 4.63 mmol, 1.0 eq) was stirred for 15 min at RT in a solution of 10 % Na<sub>2</sub>CO<sub>3</sub> (10 mL) and DCM (20 mL). The organic layer was separated and the aqueous layer extracted with DCM (3  $\times$  15 mL). The combined organic layers were dried over MgSO<sub>4</sub>, filtered and concentrated *in vacuo*. The resulting residue was dissolved in anhydrous DCM (10 mL) and DMAP (1.02 g, 8.96 mmol, 2.0 eq) was added. Malonyl chloride (436  $\mu$ L, 4.48 mmol, 1.0 eq) was added in one portion and the reaction stirred for 1h at RT before being quenched with saturated NH<sub>4</sub>Cl solution (10 mL). The aqueous layer was extracted with DCM (3  $\times$  10 mL), the combined organic fraction was washed with 1 M HCl (3  $\times$  40 mL) and then dried over MgSO<sub>4</sub>, filtered and concentrated *in vacuo* to afford **5** (1.07 g, 94 %) as a yellow solid: R<sub>f</sub> 0.1 (EtOAc/AcOH, 99:1); <sup>1</sup>H NMR (400 MHz, DMSO-d<sub>6</sub>)  $\delta_{\text{H}}$  8.05 – 6.75 (m, 4H, C<sub>5</sub>H / C<sub>6</sub>H), 4.31 (q, *J* = 7.3 Hz, 2H, C<sub>2</sub>H<sub>2</sub>), 3.69 (s, 1H, C<sub>9</sub>H), 1.33 (t, *J* = 7.3 Hz, 3H, C<sub>1</sub>H<sub>3</sub>); <sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>)  $\delta_{\text{C}}$  165.8

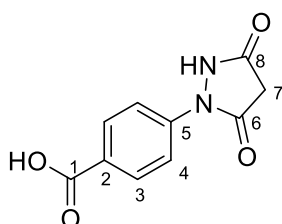
(C<sub>3</sub>), 130.8 (C<sub>5</sub>), 130.1, 127.1, 117.6 (C<sub>6</sub>), 112.9, 106.4, 61.2 (C<sub>2</sub>), 60.7, 39.9 (C<sub>9</sub>), 37.8, 14.3 (C<sub>1</sub>); *m/z* (ES-TOF) 247 ([M-H]<sup>-</sup>, 100 %), 495 ([2M-H]<sup>-</sup>, 100 %), 249 ([M+H]<sup>+</sup>, 60 %).

### 7.1.8 4-((4*E*)-4-[(5-nitro-2-furyl)methylidene]ethyl-3,5-dioxypyrazolidin-1-yl)benzoate (PYR-41)



The following procedure was adapted from a previously reported method.<sup>267</sup> A solution of **5** (20.0 mg, 0.08 mmol, 1.0 eq) and 5-nitro-2-furaldehyde (17.0 mg, 0.08 mmol, 1.0 eq) in MeOH (1 mL) was refluxed under microwave conditions for 20 min, after which formation of a dark red precipitate was observed. The precipitate was collected by centrifugation (3000 g), washed with Et<sub>2</sub>O (2 × 2 mL), DCM (2 × 2 mL) and dried *in vacuo* to afford **PYR-41** (27.0 mg, 91 %) as a dark red solid: <sup>1</sup>H NMR (400 MHz, DMSO-*d*<sub>6</sub>) δ<sub>H</sub> 8.43 (d, *J* = 4.1 Hz, 0.5H, C<sub>11</sub>H), 8.25 (bs, 0.5H, C<sub>11</sub>H), 8.05-7.78 (m, 5H, C<sub>5</sub>H, C<sub>6</sub>H, C<sub>13</sub>H / C<sub>14</sub>H), 7.54 (d, *J* = 10.5 Hz, 1H, C<sub>13</sub>H / C<sub>14</sub>H), 4.25 (q, *J* = 7.1 Hz, 2H, C<sub>2</sub>H<sub>2</sub>), 1.27 (t, *J* = 7.1 Hz, 3H, C<sub>1</sub>H<sub>3</sub>); <sup>13</sup>C NMR (101 MHz, DMSO-*d*<sub>6</sub>) δ<sub>C</sub> 172.1, 165.6 (C<sub>3</sub>), 154.1 (C<sub>8</sub>/C<sub>10</sub>), 153.6 (C<sub>8</sub>/C<sub>10</sub>), 150.6, 130.8 (C<sub>5</sub>/C<sub>13</sub>/ C<sub>14</sub>), 130.7 (C<sub>5</sub>/C<sub>13</sub>/ C<sub>14</sub>), 125.3 (C<sub>11</sub>), 122.2, 117.7 (C<sub>13</sub>/C<sub>14</sub>), 115.0 (C<sub>6</sub>), 61.1 (C<sub>2</sub>), 14.7 (C<sub>1</sub>); LC-MS *m/z* (ES<sup>-</sup>) 370 ([M-H]<sup>-</sup>, 100 %); HRMS *m/z* (ES-ToF) found 370.0681 [M-H]<sup>-</sup>, expected 370.0675.

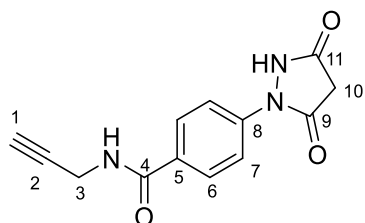
### 7.1.9 4-(3,5-dioxypyrazolidin-1-yl)benzoic acid (**9**)



**5** (0.08 g, 0.32 mmol, 1.0 eq) was stirred in a solution of 1 M NaOH (2 mL) and MeOH (1 mL) for 1 hr at RT. MeOH was removed under reduced pressure and the remaining aqueous solution adjusted to pH 3 using 1 M HCl. The mixture was extracted with EtOAc (3 × 5 mL), dried over MgSO<sub>4</sub>, filtered, and concentrated *in vacuo* to afford **9** (61.5 mg, 86 %) as a pale yellow solid: <sup>1</sup>H NMR (400 MHz, Acetone-*d*<sub>6</sub>) δ<sub>H</sub> 8.06 (d, *J* = 8.9 Hz, 2H, C<sub>3</sub>H), 7.86 (d, *J* = 8.9 Hz, 2H, C<sub>4</sub>H), 3.58 (s, 2H, C<sub>7</sub>H<sub>2</sub>); <sup>13</sup>C NMR

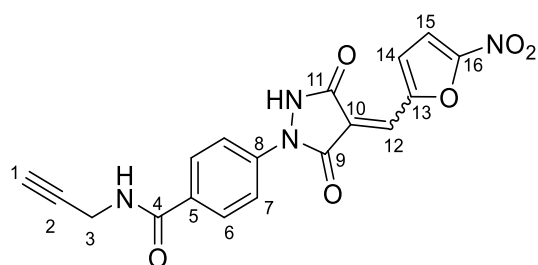
(101 MHz, DMSO- $d_6$ )  $\delta_c$  167.8(C<sub>1</sub>), 167.2 (C<sub>6</sub>/ C<sub>8</sub>), 131.3, 130.8 (C<sub>3</sub>), 117.6 (C<sub>4</sub>), 111.3, 38.3 (C<sub>7</sub>); LC-MS  $m/z$  (ES<sup>+</sup>) 221 ([M+H]<sup>+</sup>, 80 %); HRMS  $m/z$  (ES-ToF) found 221.0562 [M+H]<sup>+</sup>, expected 221.0562.

#### 7.1.10 4-(3,5-dioxypyrazolidin-1-yl)-*N*-(prop-2-yn-1-yl)benzamide (4)



To a solution of **9** (0.20 g, 0.90 mmol, 1.0 eq) in anhydrous DMF (1 mL) EDCI hydrochloride (0.192 g, 1.00 mmol, 1.1 eq) and DIPEA (175  $\mu$ L, 1.00 mmol, 1.1 eq) were added. The solution was stirred for 10 min at RT before propargyl amine (0.064 g, 1.00 mmol, 1.1 eq) was added. The reaction was stirred for 16 hr at RT, before the DMF was removed at 60 °C using a Genevac<sup>®</sup>. The resulting crude residue was dissolved in 28 % MeOH/H<sub>2</sub>O and purified by preparative LC-MS (5-98 % MeOH/H<sub>2</sub>O (0.1 % formic acid)) to afford **4** (21.2 mg, 9.1 %) as a colourless solid:  $R_f$  0.2 (MeOH/DCM 1:9); <sup>1</sup>H NMR (400 MHz, DMSO- $d_6$ )  $\delta_H$  9.03 – 8.80 (m, 1H, NH), 8.11 – 7.63 (m, 4H, C<sub>6</sub>H / C<sub>7</sub>H), 4.06 (dd,  $J$  = 5.5, 2.8 Hz, 2H, C<sub>3</sub>H), 3.66 (s, 2H, C<sub>10</sub>H), 3.14 (t,  $J$  = 2.8 Hz, 1H, C<sub>1</sub>H); <sup>13</sup>C NMR (101 MHz, DMSO- $d_6$ )  $\delta_c$  165.6 (C<sub>4</sub>), 129.5, 128.7 (C<sub>6</sub>), 118.3, 117.8 (C<sub>7</sub>), 104.1, 81.8 (C<sub>2</sub>), 73.4 (C<sub>1</sub>), 38.3 (C<sub>10</sub>), 28.9 (C<sub>3</sub>); LC-MS  $m/z$  (ES<sup>+</sup>) 258 ([M+H]<sup>+</sup>, 70 %), 256 ([M-H]<sup>-</sup>, 70 %); HRMS  $m/z$  (ES-ToF) found 258.0887 [M+H]<sup>+</sup>, expected 258.0879.

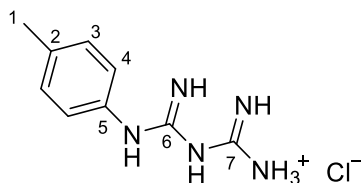
#### 7.1.11 4-{{(4*E*)-4-[(5-nitrofuran-2-yl)methylidene]-3,5-dioxypyrazolidin-1-yl}}-*N*-(prop-2-yn-1-yl)benzamide (PYR-41-P)



A solution of **4** (5.3 mg, 0.021 mmol, 1.0 eq) and 5-nitro-2-furaldehyde (3 mg, 0.021 mmol, 1.0 eq) in MeOH (1 mL) was refluxed under microwave conditions for 20 min, on which formation of a deep purple precipitate was observed. The precipitate was collected by centrifugation (3000 g), washed with Et<sub>2</sub>O (2  $\times$  2 mL), DCM (2  $\times$  2 mL) and dried *in vacuo* to afford **PYR-41-P** (5.5 mg, 69 %) as a deep

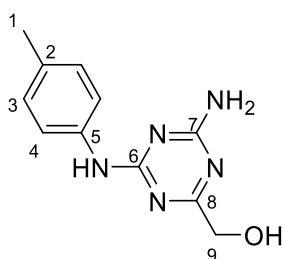
purple solid:  $^1\text{H}$  NMR (400 MHz, DMSO- $d_6$ )  $\delta_{\text{H}}$  8.95 (t,  $J$  = 5.6 Hz, 1H, NH), 8.50 (d,  $J$  = 4.1 Hz, 1H,  $\text{C}_{12}\text{H}$ ), 8.19 – 7.45 (m, 6H,  $\text{C}_6\text{H}/\text{C}_7\text{H}/\text{C}_{14}\text{H}/\text{C}_{15}\text{H}$ ), 4.05 (dd,  $J$  = 5.6, 2.5 Hz, 2H,  $\text{C}_3\text{H}$ ), 3.14 (t,  $J$  = 2.5 Hz, 1H,  $\text{C}_1\text{H}$ );  $^{13}\text{C}$  NMR (101 MHz, DMSO- $d_6$ )  $\delta_{\text{C}}$  196.5, 165.1, 153.1, 150.2, 146.4, 128.4 ( $\text{C}_6/\text{C}_{15}$ ), 128.3 ( $\text{C}_6/\text{C}_{15}$ ), 124.8 ( $\text{C}_{12}$ ), 121.9, 117.8, 117.3, 114.6 ( $\text{C}_{14}$ ), 113.1 ( $\text{C}_7$ ), 81.4 ( $\text{C}_2$ ), 73.0( $\text{C}_1$ ), 28.5 ( $\text{C}_3$ ); LC-MS  $m/z$  ( $\text{ES}^+$ ) 381 ( $[\text{M}+\text{H}]^+$ , 80 %); HRMS  $m/z$  ( $\text{ES-ToF}$ ) found 379.0671  $[\text{M}-\text{H}]^-$ , expected 379.0679.

### 7.1.12 *N*-(4-methylphenyl)imidodicarbonimidic diamide (**11**)



This compound was prepared following a previously reported method.<sup>211</sup> A mixture of 4-methylaniline (457 mg, 4.27 mmol, 1.0 eq) and dicyandiamide (359 mg, 4.27 mmol, 1.0 eq) was dissolved in aqueous HCl (3 M, 1.4 mL) and heated at 90 °C overnight. The reaction was allowed to cool, and the resulting solid isolated by suction filtration. The solid was washed with THF (5 mL), hexane (5 mL) and acetone (5 mL) before being dried *in vacuo* to afford **11** (752 mg, 92 %) as a colourless HCl salt:  $^1\text{H}$  NMR (400 MHz, DMSO- $d_6$ )  $\delta_{\text{H}}$  9.54 (s, 1H, NH), 7.22 (m, 6H,  $\text{C}_3\text{H}$ , 4  $\times$  NH), 7.10 (m, 2H,  $\text{C}_4\text{H}$ ), 7.01 (s, 2H, NH), 2.25 (s, 3H,  $\text{C}_1\text{H}_3$ );  $^{13}\text{C}$  NMR (101 MHz, DMSO- $d_6$ )  $\delta_{\text{C}}$  161.42 ( $\text{C}_6/\text{C}_7$ ), 155.99 ( $\text{C}_6/\text{C}_7$ ), 136.42 ( $\text{C}_2/\text{C}_5$ ), 132.94 ( $\text{C}_2/\text{C}_5$ ), 129.52 ( $\text{C}_4$ ), 121.62 ( $\text{C}_3$ ), 20.89 ( $\text{C}_1$ ); LC-MS  $m/z$  ( $\text{ES}^+$ ) 192 ( $[\text{M}+\text{H}]^+$ , 100 %).

### 7.1.13 {4-amino-6-[(4-methylphenyl)amino]-1,3,5-triazin-2-yl}methanol (**10**)

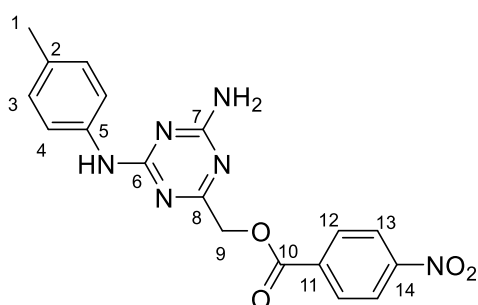


This compound was prepared following a previously reported method.<sup>211</sup> A fresh solution of 37 % NaOMe in MeOH was prepared directly before use by dissolving sodium (160 mg) in anhydrous MeOH (1 mL) under an Ar filled condenser. To a solution of **11** (500 mg, 2.20 mmol, 1.0 eq) in anhydrous MeOH (2.5 mL), NaOMe (350  $\mu\text{L}$ , 37 % w/v in MeOH, 2.29 mmol, 1.04 eq) was added. The reaction was further diluted with MeOH (20 mL) and stirred for 3 hours at RT. Ethyl glycolate (252



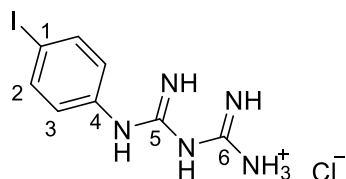
$\mu\text{L}$ , 2.65 mmol, 1.2 eq) was then added dropwise and the reaction was refluxed overnight. After cooling, the crude product was concentrated in *vacuo* and subsequent purification by silica gel chromatography (hexane: EtOAc, 2:8) afforded **10** (90 mg, 39 %) as a colourless solid:  $R_f$  0.3 (hexane / EtOAc, 2:8);  $^1\text{H}$  NMR (400 MHz, MeOD)  $\delta_H$  7.67 – 7.51 (m, 2H,  $\text{C}_3\text{H}$ ), 7.26 – 7.11 (m, 2H,  $\text{C}_4\text{H}$ ), 4.37 (s, 2H,  $\text{C}_9\text{H}_2$ ), 2.32 (s, 3H,  $\text{C}_1\text{H}_3$ );  $^{13}\text{C}$  NMR (101 MHz, MeOD)  $\delta_C$  176.6 ( $\text{C}_6/\text{C}_7/\text{C}_8$ ), 166.8 ( $\text{C}_6/\text{C}_7/\text{C}_8$ ), 164.2 ( $\text{C}_6/\text{C}_7/\text{C}_8$ ), 136.3 ( $\text{C}_2/\text{C}_5$ ), 132.5 ( $\text{C}_2/\text{C}_5$ ), 128.7 ( $\text{C}_4$ ), 120.7 ( $\text{C}_3$ ), 63.2 ( $\text{C}_9$ ), 19.5 ( $\text{C}_1$ ); LC-MS  $m/z$  ( $\text{ES}^+$ ) 232 ( $[\text{M}+\text{H}]^+$ , 100 %).

#### 7.1.14 {4-amino-6-[(4-methylphenyl)amino]-1,3,5-triazin-2-yl}methyl 4-nitrobenzoate (**Tz9**)



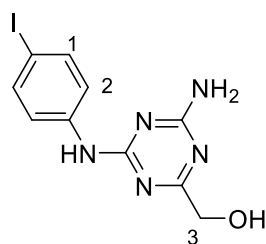
This compound was prepared following a previously reported method.<sup>211</sup> To a solution of **10** (13 mg, 0.056 mmol, 1.0 eq) and triethylamine (17  $\mu\text{L}$ , 0.124 mmol, 2.2 eq) in anhydrous DCM (5 ml), 4-nitrobenzoyl chloride (10.4 mg, 0.056 mmol, 1 eq) was added and the reaction was refluxed overnight. On cooling, the crude mixture was washed with brine (2  $\times$  5 mL), dried over  $\text{NaSO}_4$ , filtered, and concentrated *in vacuo*. Subsequent purification by silica gel chromatography (hexane: EtOAc, 1:1) afforded **Tz9** (15 mg, 70 %) as a colourless solid:  $R_f$  0.3 (hexane / EtOAc, 1:1);  $^1\text{H}$  NMR (400 MHz,  $\text{DMSO-d}_6$ )  $\delta_H$  9.40 (s, 1H, NH), 8.47 – 8.37 (m, 2H,  $\text{C}_{13}\text{H}$ ), 8.34 – 8.20 (m, 2H,  $\text{C}_{12}\text{H}$ ), 7.70-7.30 (m, 2H,  $\text{C}_3\text{H}$ ), 7.25- 6.55 (m, 4H,  $\text{C}_4\text{H} / \text{NH}_2$ ), 5.14 (s, 2H,  $\text{C}_9\text{H}_2$ ), 2.16 (s, 3H,  $\text{C}_1\text{H}_3$ );  $^{13}\text{C}$  NMR (101 MHz,  $\text{DMSO-d}_6$ )  $\delta_C$  172.5 ( $\text{C}_6/\text{C}_7/\text{C}_8/\text{C}_{10}$ ), 167.0 ( $\text{C}_6/\text{C}_7/\text{C}_8/\text{C}_{10}$ ), 164.5 ( $\text{C}_6/\text{C}_7/\text{C}_8/\text{C}_{10}$ ), 164.4 ( $\text{C}_6/\text{C}_7/\text{C}_8/\text{C}_{10}$ ), 150.8, 137.4 ( $\text{C}_2/\text{C}_5$ ), 135.4 ( $\text{C}_2/\text{C}_5$ ), 131.4 ( $\text{C}_{12}$ ), 129.1 ( $\text{C}_4$ ), 124.4 ( $\text{C}_{13}$ ), 120.6 ( $\text{C}_3$ ), 66.0 ( $\text{C}_9$ ), 20.8 ( $\text{C}_1$ ); LC-MS  $m/z$  ( $\text{ES}^+$ ) 381 ( $[\text{M}+\text{H}]^+$ , 100 %). HRMS  $m/z$  ( $\text{ES}^+$ ) found 381.1310  $[\text{M}+\text{H}]^+$ , expected 381.1311.

### 7.1.15 *N*-(4-iodophenyl)imidodicarbonimidic diamide (**20**)



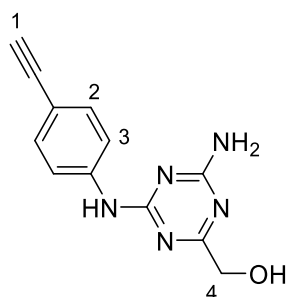
A mixture of 4-iodoaniline (935 mg, 4.27 mmol, 1.0 eq) and dicyandiamide (359 mg, 4.27 mmol, 1.0 eq) was dissolved in aqueous HCl (3 M, 1.4 mL) and heated at 90 °C overnight. The reaction was allowed to cool, and the resulting precipitate isolated by suction filtration. The solid was washed with THF (5 mL), hexane (5 mL) and acetone (5 mL) before being dried *in vacuo* to afford **20** (816 mg, 63 %) as a pale purple HCl salt:  $^1\text{H}$  NMR (400 MHz, DMSO- $d_6$ )  $\delta_{\text{H}}$  9.92 (s, 1H, NH), 7.66 – 7.57 (m, 2H, C<sub>3</sub>), 7.40 (s, 4H, NH), 7.26 – 7.17 (m, 2H, C<sub>2</sub>), 7.10 (s, 2H, NH);  $^{13}\text{C}$  NMR (101 MHz, DMSO- $d_6$ )  $\delta_{\text{C}}$  161.8 (C<sub>5</sub> / C<sub>6</sub>), 155.2 (C<sub>5</sub> / C<sub>6</sub>), 139.2 (C<sub>4</sub>), 137.7 (C<sub>2</sub>), 123.2 (C<sub>3</sub>), 87.2 (C<sub>1</sub>); LC-MS  $m/z$  (ES<sup>+</sup>) 304 ([M+H]<sup>+</sup>, 100 %). HRMS  $m/z$  (ES<sup>+</sup>) found 304.0063 [M+H]<sup>+</sup>, expected 304.0069.

### 7.1.16 {4-amino-6-[(4-iodophenyl)amino]-1,3,5-triazin-2-yl}methanol (**21**)



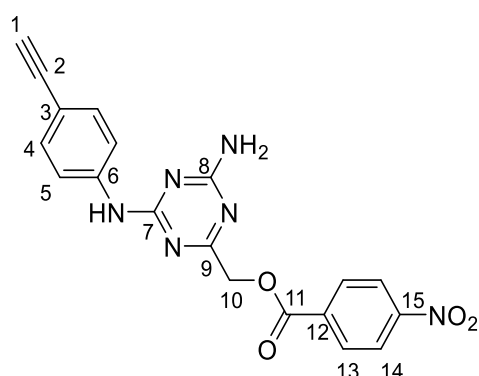
A fresh solution of 37 % NaOMe in MeOH was prepared directly before use by dissolving sodium (160 mg) in anhydrous MeOH (1 mL) under an Ar filled condenser. To a solution of **20** (200 mg, 0.66 mmol, 1.0 eq) in anhydrous MeOH (20 mL), NaOMe (100  $\mu\text{L}$ , 37 % w/v in MeOH, 0.69 mmol, 1.04 eq) was added. The reaction was further diluted with MeOH (20 mL) and stirred for 3 hours at RT. Ethyl glycolate (150  $\mu\text{L}$ , 0.79 mmol, 2.4 eq) was then added dropwise and the reaction was refluxed overnight. The resulting precipitate was washed by trituration in cold methanol (2  $\times$  5 mL) and dried *in vacuo* to afford **21** (180 mg, 80 %) as a colourless solid:  $R_f$  0.7 (MeOH / H<sub>2</sub>O, 1:1);  $^1\text{H}$  NMR (400 MHz, DMSO- $d_6$ )  $\delta_{\text{H}}$  9.65 (s, 1H, NH), 7.70 – 7.50 (m, 4H, C<sub>1</sub>, C<sub>2</sub>), 7.12 (bs, 2H, NH<sub>2</sub>), 4.21 (s, 2H, C<sub>3</sub>H<sub>2</sub>); LC-MS  $m/z$  (ES<sup>+</sup>) 344 ([M+H]<sup>+</sup>, 100 %).

### 7.1.17 {4-amino-6-[(4-ethynylphenyl)amino]-1,3,5-triazin-2-yl}methanol (**15**)



A 2 mL microwave vial was charged with **21** (50 mg, 0.14 mmol, 1 eq), bis(triphenylphosphine) palladium(II) dichloride (5 mg, 0.007 mmol, 0.05 eq) and copper iodide (1.3 mg, 0.007 mmol, 0.05 eq) before being sealed, evacuated, and filled with Ar gas. The vial was evacuated and refilled with Ar a further 2 times before DMF was added (1.5 mL), followed by triethylamine (40  $\mu$ L, 0.28 mmol, 2 eq). The reaction was stirred for 5 min, before ethynyltrimethylsilane (25  $\mu$ L, 0.18 mmol, 1.2 eq) was added and the reaction was heated at 100  $^{\circ}$ C overnight. Once cooled, the reaction was quenched with water (20 mL), and extracted with EtOAc (2  $\times$  20 mL). The combined organic layers were filtered through Celite and concentrated *in vacuo*. The resulting crude material was re-suspended in a slurry of  $K_2CO_3$  (20 mg, 0.14 mmol, 1 eq) in MeOH (5 mL) and stirred overnight at RT. Solvent removal *in vacuo* afforded **15** (26.2 mg, 63.7 %) as a colourless salt. This intermediate was carried through without further purification. Crude analysis:  $R_f$  0.6 (100 % EtOAc); LC-MS  $m/z$  ( $ES^+$ ) 242 ( $[M+H]^+$ ), 100 %).

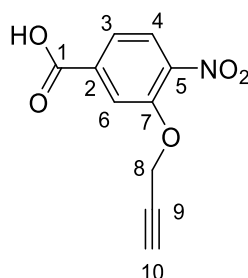
### 7.1.18 {4-amino-6-[(4-ethynylphenyl)amino]-1,3,5-triazin-2-yl}methyl 4-nitrobenzoate (**T1**)



Crude intermediate **15** was dissolved in anhydrous THF (10 mL), with triethylamine (43  $\mu$ L, 0.31 mmol, 2.2 eq). 4-nitrobenzoyl chloride (26  $\mu$ L, 0.14 mmol, 1 eq) was added, and the reaction was refluxed overnight. THF was removed *in vacuo* and the crude mixture dissolved in DCM (10 mL),

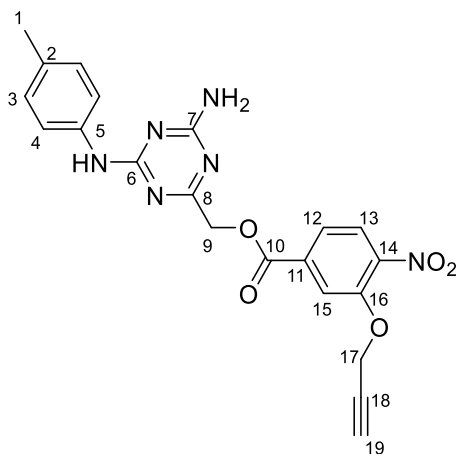
washed with brine (2 × 10 mL), dried over Na<sub>2</sub>SO<sub>4</sub>, filtered, and concentrated *in vacuo*. Subsequent purification by silica gel chromatography (hexane: EtOAc, 1:1) afforded **T1** (3.6 mg, 6.4 %) as a colourless solid: R<sub>f</sub> 0.2 (hexane / EtOAc, 1:1); <sup>1</sup>H NMR (500 MHz, DMSO-d<sub>6</sub>) δ<sub>H</sub> 9.73 (s, 1H, NH), 8.43 – 8.26 (m, 4H, C<sub>13</sub>H, C<sub>14</sub>H), 8.16 (dd, *J* = 8.4 Hz, 2H, impurity), 7.84 – 7.59 (m, 2H, C<sub>4</sub>H), 7.33–6.99 (m, 4H, NH, C<sub>5</sub>H) 6.65 (s, 1H), 5.17 (s, 2H, C<sub>10</sub>H), 4.01 (s, 1H, C<sub>1</sub>H); <sup>13</sup>C NMR (limited assignment from HSQC and HMBC, 126 MHz, DMSO-d<sub>6</sub>) δ<sub>C</sub> 172.8, 164.6, 150.9, 149.3, 135.4, 131.8 (C<sub>5</sub>), 131.5, 130.9 (C<sub>13</sub> / C<sub>14</sub>), 130.2 (impurity), 123.8 (C<sub>13</sub> / C<sub>14</sub>), 123.0 (impurity), 119.4 (C<sub>4</sub>), 79.4 (C<sub>1</sub>), 65.4 (C<sub>10</sub>); LC-MS *m/z* (ES<sup>+</sup>) 391 ([M+H]<sup>+</sup>, 100 %); HRMS *m/z* (ES-ToF) found 391.1167 [M+H]<sup>+</sup>, expected 391.1155.

### 7.1.19 4-nitro-3-(prop-2-yn-1-yloxy)benzoic acid (**12**)



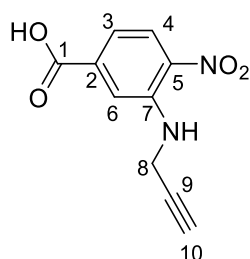
The following procedure was adapted from a previously reported method.<sup>281</sup> A mixture of propargyl alcohol (100 mg, 0.54 mmol, 1.0 eq) and NaHMDS (1.30 mL of a 1 M solution in THF, 1.30 mmol, 2.4 eq) was stirred for 10 min at RT under N<sub>2</sub>. 3-fluoro-4-nitrobenzoic acid was suspended in anhydrous THF (0.50 mL), and added drop wise to the reaction flask. The reaction was stirred overnight at RT, after which time the solvent was removed *in vacuo*. The resulting pale yellow solid was washed with 2 % HCl and dried under vacuum to afford **12** (52.8 mg, 44 %) as a pale yellow solid. The product was found to be pure by TLC analysis: R<sub>f</sub> 0.6 (EtOAc, 1 % AcOH); <sup>1</sup>H NMR (400 MHz, MeOD) δ<sub>H</sub> 7.95 (d, *J* = 1.5 Hz, 1H, C<sub>6</sub>H), 7.82 (d, *J* = 8.3 Hz, 1H, C<sub>4</sub>H), 7.70 (dd, *J* = 8.3, 1.5 Hz, 1H, C<sub>3</sub>H), 4.97 (d, *J* = 2.4 Hz, 2H, C<sub>8</sub>H<sub>2</sub>), 3.09 (t, *J* = 2.4 Hz, 1H, C<sub>10</sub>H); <sup>13</sup>C NMR (101 MHz, MeOD) δ<sub>C</sub> 167.2 (C<sub>1</sub>), 149.8 (C<sub>2</sub> / C<sub>5</sub> / C<sub>7</sub>), 142.8 (C<sub>2</sub> / C<sub>5</sub> / C<sub>7</sub>), 137.0 (C<sub>2</sub> / C<sub>5</sub> / C<sub>7</sub>), 124.5 (C<sub>3</sub>), 122.0 (C<sub>6</sub>), 116.1 (C<sub>4</sub>), 77.1 (C<sub>10</sub>), 76.9 (C<sub>9</sub>), 56.8 (C<sub>8</sub>); LC-MS *m/z* (ES<sup>+</sup>) 220 ([M+H]<sup>+</sup>, 100 %); HRMS *m/z* (ES<sup>-</sup>) Found: 220.0254, expected 220.0246.

### 7.1.20 {4-amino-6-[(4-methylphenyl)amino]-1,3,5-triazin-2-yl)methyl 4-nitro-3-(prop-2-yn-1-yloxy)benzoate (T2)



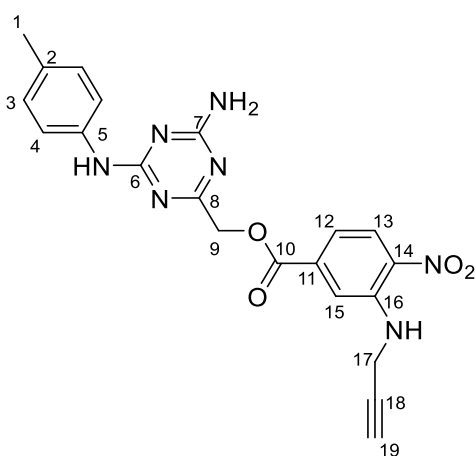
To a suspension of **12** (21.0 mg, 0.10 mmol, 1.1 eq) in anhydrous toluene (10 mL) under N<sub>2</sub>, thionyl chloride (50 μL, 0.73 mmol, 8.5 eq) was added. The reaction mixture was heated to 75 °C under N<sub>2</sub> for 2 hr, after which time the solvent was removed *in vacuo*. The crude material was then resuspended in anhydrous DCM (5 mL), **10** (20 mg 0.09 mmol, 1.1 eq) and triethylamine (24 μL, 0.17, 2.0 eq) was added, and the reaction refluxed overnight, resulting in the formation of a precipitate. The solvent was removed *in vacuo* and the crude mixture was dissolved in EtOAc (10 mL) before being washed with 2 % HCl (2x 10 mL), brine (1 × 10 mL), dried over MgSO<sub>4</sub>, filtered, and concentrated *in vacuo*. Subsequent purification by silica gel chromatography (hexane: EtOAc, 1:1) afforded **T2** (30.9 mg, 74.9 %) as an off white solid: R<sub>f</sub> 0.2 (hexane / EtOAc, 1:1); <sup>1</sup>H NMR (400 MHz, DMSO-d<sub>6</sub>) δ<sub>H</sub> 9.39 (s, 1H, NH), 8.07 (d, *J* = 8.3 Hz, 1H, C<sub>13</sub>H), 7.96 (d, *J* = 1.6 Hz, 1H, C<sub>15</sub>H), 7.82 (dd, *J* = 8.3, 1.6 Hz, 1H, C<sub>12</sub>H), 7.74-7.25 (m, 2H, C<sub>3</sub>H), 7.26-6.46 (m, 4H, C<sub>4</sub>H, NH), 5.38 – 4.96 (m, 4H, C<sub>9</sub>H<sub>2</sub>, C<sub>17</sub>H<sub>2</sub>), 3.71 (bs, 1H, C<sub>19</sub>H), 2.17 (bs, 3H, C<sub>1</sub>); <sup>13</sup>C NMR (101 MHz, DMSO-d<sub>6</sub>) δ<sub>C</sub> 172.0 (C<sub>10</sub>), 166.5 (C<sub>6</sub>/C<sub>7</sub>/C<sub>8</sub>), 163.9 (C<sub>6</sub>/C<sub>7</sub>/C<sub>8</sub>), 153.4, 151.7, 149.3, 142.9, 136.9, 134.0, 131.1, 128.6, 125.2 (C<sub>15</sub>), 122.6 (C<sub>12</sub>), 120.2 (C<sub>3</sub> /C<sub>4</sub>), 116.2 (C<sub>13</sub>), 79.9 (C<sub>19</sub>), 77.7 (C<sub>9</sub>), 57.3 (C<sub>17</sub>), 20.3 (C<sub>1</sub>); LC-MS *m/z* (ES<sup>+</sup>) 435 ([M+H]<sup>+</sup>, 100 %); HRMS *m/z* (ES<sup>+</sup>) found 435.1413 [M+H]<sup>+</sup>, expected 435.1417.

### 7.1.21 4-nitro-3-(prop-2-yn-1-ylamino)benzoic acid (**13**)



The following procedure was adapted from a previously reported method.<sup>280</sup> A solution of 3-fluoro-4-nitro-benzoic acid (100 mg, 0.54 mmol, 1 eq) and propargylamine (2 mL of a 40 % aqueous solution, 13.5 mmol, 25 eq) was heated at 85 °C overnight. The reaction mixture was cooled to RT, acidified with aqueous 2 % HCl (dropwise) and extracted with EtOAc (3 × 10 mL). The combined organic fraction was washed with 2 % HCl (3 × 10 mL), dried over NaSO<sub>4</sub>, filtered, and concentrated *in vacuo* to afford **13** (86.2 mg, 72.6 %) as an orange solid. The product was found to be pure by TLC analysis: R<sub>f</sub> 0.7 (EtOAc, 1 % AcOH); <sup>1</sup>H NMR (400 MHz, MeOD) δ<sub>H</sub> 8.25 (d, *J* = 8.8 Hz, 1H, C<sub>4</sub>H), 7.81 (d, *J* = 1.7 Hz, 1H, C<sub>6</sub>H), 7.35 (dd, *J* = 8.8, 1.7 Hz, 1H, C<sub>3</sub>H), 4.26 (d, *J* = 2.5 Hz, 2H, C<sub>8</sub>H<sub>2</sub>), 2.72 (t, *J* = 2.5 Hz, 1H, C<sub>10</sub>H); <sup>13</sup>C NMR (101 MHz, MeOD) δ<sub>C</sub> 166.8 (C<sub>1</sub>), 143.6 (C<sub>2</sub> / C<sub>5</sub> / C<sub>7</sub>), 137.0 (C<sub>2</sub> / C<sub>5</sub> / C<sub>7</sub>), 134.5 (C<sub>2</sub> / C<sub>5</sub> / C<sub>7</sub>), 126.3 (C<sub>3</sub>), 116.1 (C<sub>4</sub>/C<sub>6</sub>), 115.8 (C<sub>4</sub>/C<sub>6</sub>), 79.0 (C<sub>9</sub>), 71.8 (C<sub>10</sub>), 31.6 (C<sub>8</sub>); LC-MS *m/z* (ES<sup>+</sup>) 221 ([M+H]<sup>+</sup>, 100 %); HRMS *m/z* found 221.0567 [M+H]<sup>+</sup>, expected 221.0557.

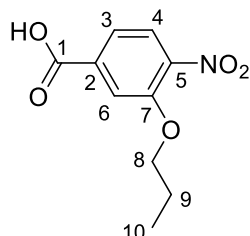
### 7.1.22 {4-amino-6-[(4-methylphenyl)amino]-1,3,5-triazin-2-yl}methyl 4-nitro-3-(prop-2-yn-1-ylamino)benzoate (**T3**)



To a suspension of **13** (21.0 mg, 0.10 mmol, 1.1 eq) in anhydrous toluene (10 mL) under N<sub>2</sub>, thionyl chloride (50 μL, 0.73 mmol, 8.5 eq) was added. The reaction mixture was heated to 75 °C under N<sub>2</sub> for 2 hr, after which time the solvent was removed *in vacuo*. The crude material was then

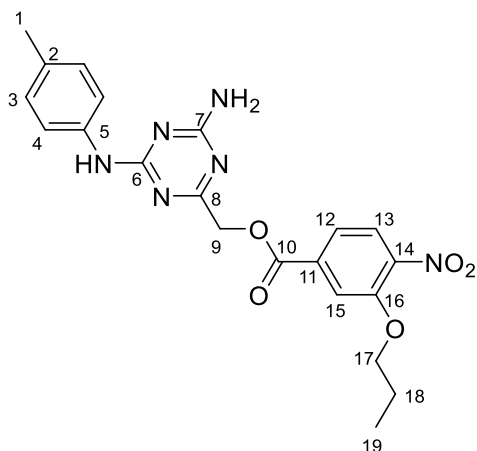
resuspended in anhydrous DCM (5 mL), **10** (20 mg 0.09 mmol, 1.1 eq) and triethylamine (24  $\mu$ L, 0.17, 2 eq) was added, and the reaction refluxed overnight, resulting in the formation of a precipitate. The solvent was removed *in vacuo* and the crude mixture dissolved in EtOAc (10 mL) before being washed with 2 % HCl (2x 10 mL), brine (1  $\times$  10 mL), dried over MgSO<sub>4</sub>, filtered, and concentrated *in vacuo*. Subsequent purification by silica gel chromatography (hexane: EtOAc, 1:1) afforded **T3** (26.2 mg, 63.7 %) as a yellow solid: R<sub>f</sub> 0.3 (hexane / EtOAc, 1:1); <sup>1</sup>H NMR (400 MHz, DMSO-d<sub>6</sub>)  $\delta$ <sub>H</sub> 9.40 (s, 1H, NH), 8.39 (t, *J* = 6.1 Hz, 1H, NH), 8.24 (d, *J* = 8.8 Hz, 1H, C<sub>13</sub>H), 7.72 (d, *J* = 1.8 Hz, 1H, C<sub>15</sub>H), 7.67 – 7.40 (m, 2H, C<sub>3</sub>H), 7.33 (dd, *J* = 8.8, 1.8 Hz, 1H, C<sub>12</sub>H), 7.27-6.54 (m, 4H, C<sub>4</sub>H, NH), 5.11 (s, 2H, C<sub>9</sub>H<sub>2</sub>), 4.25 (d (weak), *J* = 6.1, 2H, C<sub>17</sub>H<sub>2</sub>), 3.22 (s, 1H, C<sub>19</sub>H), 2.16 (s, 3H, C<sub>1</sub>H<sub>3</sub>); <sup>13</sup>C NMR (101 MHz, DMSO-d<sub>6</sub>)  $\delta$ <sub>C</sub> 172.1 (C<sub>10</sub>), 166.5 (C<sub>6</sub>/C<sub>7</sub>/C<sub>8</sub>), 164.5 (C<sub>6</sub>/C<sub>7</sub>/C<sub>8</sub>), 163.9 (C<sub>6</sub>/C<sub>7</sub>/C<sub>8</sub>), 143.5, 136.9, 135.7, 134.1, 130.9, 128.6 (C<sub>3</sub>/C<sub>4</sub>/C<sub>13</sub>), 127.0 (C<sub>3</sub>/C<sub>4</sub>/C<sub>13</sub>), 120.2 (C<sub>3</sub>/C<sub>4</sub>), 116.4 (C<sub>12</sub>/C<sub>15</sub>), 115.8 (C<sub>12</sub>/C<sub>15</sub>), 80.1 (C<sub>18</sub>), 74.4 (C<sub>19</sub>), 65.5 (C<sub>9</sub>), 31.9 (C<sub>17</sub>), 20.3 (C<sub>1</sub>); LC-MS *m/z* (ES<sup>+</sup>) 434 ([M+H]<sup>+</sup>, 100 %); HRMS *m/z* (ES<sup>+</sup>) found 434.1595 [M+Na]<sup>+</sup>, expected 434.1577.

### 7.1.23 4-nitro-3-(propan-1-yloxy)benzoic acid (**23**)



The following procedure was adapted from a previously reported method.<sup>281</sup> A mixture of propanol (50  $\mu$ L, 0.65 mmol, 1.2 eq) and NaHMDS (1.30 mL of a 1 M solution in THF, 1.30 mmol, 2.4 eq) was stirred for 10 min at RT under N<sub>2</sub>. 3-fluoro-4-nitrobenzoic acid (100 mg, 0.54 mmol, 1.0 eq) was suspended in anhydrous THF (0.50 mL), and added dropwise to the reaction flask. The reaction was stirred overnight at RT, after which time the solvent was removed *in vacuo*. The resulting pale yellow solid was washed with 2 % HCl and dried under vacuum to afford **23** (93.8 mg, 76 %) as a pale yellow solid. The product was found to be pure by TLC analysis: R<sub>f</sub> 0.6 (EtOAc, 1 % AcOH); <sup>1</sup>H NMR (400 MHz, MeOD)  $\delta$ <sub>H</sub> 7.95 – 7.81 (m, 2H, C<sub>4</sub>H, C<sub>6</sub>H), 7.81 – 7.68 (m, 1H, C<sub>3</sub>H), 4.18 (t, *J* = 6.3 Hz, 2H, C<sub>8</sub>H), 1.87 (sextet, *J* = 6.9 Hz, 2H, C<sub>9</sub>H), 1.09 (t, *J* = 7.4 Hz, 3H, C<sub>10</sub>H); <sup>13</sup>C NMR (101 MHz, MeOD)  $\delta$ <sub>C</sub> 166.3 (C<sub>1</sub>), 151.4, 142.8, 135.3, 124.5 (C<sub>3</sub>), 121.1 (C<sub>6</sub>), 115.2 (C<sub>4</sub>), 71.0 (C<sub>8</sub>), 22.0 (C<sub>9</sub>), 9.3 (C<sub>10</sub>); LC-MS *m/z* (ES<sup>+</sup>) 224 ([M+H]<sup>+</sup>, 100 %); HRMS *m/z* (ES<sup>+</sup>) found 224.0551[M-H]<sup>-</sup>, expected 224.0559.

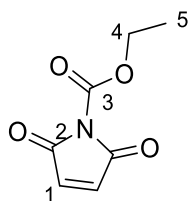
### 7.1.24 {4-amino-6-[(4-methylphenyl)amino]-1,3,5-triazin-2-yl)methyl 4-nitro-3-(propan-1 yloxy)benzoate (T2c)



To a suspension of **23** (22.0 mg, 0.10 mmol, 1.1 eq) in anhydrous toluene (10 mL) under  $N_2$ , thionyl chloride (50  $\mu$ L, 0.73 mmol, 8.5 eq) was added. The reaction mixture was heated to 75  $^{\circ}$ C under  $N_2$  for 4 hr, after which time the solvent was removed *in vacuo*. The crude material was then resuspended in anhydrous DCM (5 mL), **10** (20 mg 0.09 mmol, 1.0eq) and triethylamine (24  $\mu$ L, 0.17, 2.0 eq) was added, and the reaction refluxed overnight. The mixture was diluted with EtOAc (5 mL) before being washed with 2 % HCl (2x 10 mL), brine (1  $\times$  10 mL), dried over  $MgSO_4$ , filtered, and concentrated *in vacuo*. Subsequent purification by silica gel chromatography (hexane: EtOAc, 1:1) afforded **T2c** (24.0 mg, 57.7 %) as a colourless solid:  $R_f$  0.3 (hexane / EtOAc, 1:1);  $^1H$  NMR (400 MHz, DMSO- $d_6$ )  $\delta_H$  9.47 (s, 1H, NH), 8.04 (d,  $J$  = 8.3 Hz, 1H,  $C_{13}H$ ), 7.87 – 7.71 (m, 2H,  $C_{12}H$ ,  $C_{15}H$ ), 7.74-6.54 (m, 6H,  $C_3H$ ,  $C_4H$ ,  $NH_2$ ), 5.15 (s, 2H,  $C_9H_2$ ), 4.16 (m, 2H,  $C_{17}H_2$ ) 2.17 (s, 3H,  $C_1H_3$ ), 1.73 (sextet,  $J$  = 6.9 Hz, 2H,  $C_{18}H_2$ ), 0.96 (t,  $J$  = 7.4 Hz, 3H,  $C_{19}H_3$ );  $^{13}C$  NMR (101 MHz, DMSO- $d_6$ )  $\delta_C$  166.3 ( $C_{10}$ ), 164.5 ( $C_6/C_7/C_8$ ), 164.2 ( $C_6/C_7/C_8$ ), 151.3, 143.1, 137.2, 134.6, 129.0, 125.5 ( $C_{12}/C_{15}$ ), 122.0 ( $C_{12}/C_{15}$ ), 120.8 ( $C_3/C_4$ ), 116.0 ( $C_{13}$ ), 71.5 ( $C_{17}$ ), 65.5 ( $C_9$ ), 22.2 ( $C_{18}$ ), 20.8 ( $C_1$ ), 10.6 ( $C_{19}$ ); LC-MS  $m/z$  ( $ES^+$ ) 439 ( $[M+H]^+$ , 100 %); HRMS  $m/z$  ( $ES^+$ ) found 439.1746  $[M+H]^+$ , expected 439.1730.

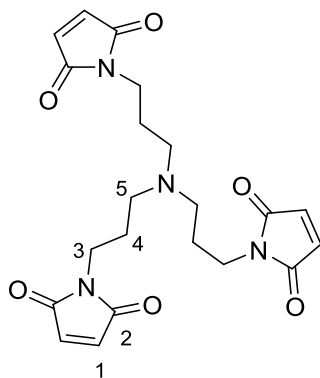


### 7.1.25 Ethyl 2,5-dioxo-2,5-dihydro-1H-pyrrole-1-carboxylate (**29**)



The following procedure was adapted from a previously reported method.<sup>310</sup> A solution of maleimide (1.00 g, 10.40 mmol, 1.0 eq) in EtOAc (20 mL) was cooled to 5 °C. *N*-Methylmorpholine (1.26 mL, 11.44 mmol, 1.1 eq) was added, followed by dropwise addition of ethyl chloroformate (1 mL, 10.40 mmol, 1.0 eq) in EtOAc (10 mL). The reaction was stirred at RT for 1 hr, before being diluted with EtOAc (50 mL), washed with brine (2x 50 mL), dried over MgSO<sub>4</sub>, filtered, and concentrated *in vacuo*. Subsequent purification by silica gel chromatography (hexane: EtOAc, 1:3) afforded **29** (1.19 g, 68 %) as a colourless solid: R<sub>f</sub> 0.7 (hexane / EtOAc, 1:3); <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ<sub>H</sub> 6.83 (s, 2H, C<sub>1</sub>H), 4.42 (q, *J* = 7.1 Hz, 2H, C<sub>4</sub>H<sub>2</sub>), 1.39 (t, *J* = 7.1 Hz, 3H, C<sub>5</sub>H<sub>3</sub>); <sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>) δ<sub>C</sub> 165.8 (C<sub>2</sub>), 147.6 (C<sub>3</sub>), 135.2 (C<sub>1</sub>), 64.0 (C<sub>4</sub>), 14.1 (C<sub>5</sub>); *m/z* (ES<sup>+</sup>) 170 ([M+H]<sup>+</sup>, 50 %).

### 7.1.26 Tris(3-maleimidopropyl)amine (Trap C)



The following procedure was adapted from a previously reported method.<sup>310</sup> Tris(3-aminopropyl)amine (120 μL, 0.59 mmol, 1.0 eq) in a saturated solution of NaHCO<sub>3</sub> in THF/ H<sub>2</sub>O (1:1 v/v, 4 mL) was cooled to 0 °C. On addition of **29** (300 mg, 1.77 mmol, 3.0 eq) the reaction was diluted with further saturated solution of NaHCO<sub>3</sub> in THF/ H<sub>2</sub>O (1:1 v/v, 17 mL) and stirred for 4 hr at 0 °C. The solution was extracted with EtOAc (3 × 80 mL), washed with brine (3 × 80 mL), dried over MgSO<sub>4</sub>, filtered, and concentrated *in vacuo*. Subsequent purification by silica gel chromatography (hexane: EtOAc, 1:3) afforded **Trap C** (112 mg, 44.4 %) as a colourless solid: R<sub>f</sub> 0.5 (hexane / EtOAc, 1:3); <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ<sub>H</sub> 6.68 (s, 6H, C<sub>1</sub>H), 3.60 – 3.51 (m, 6H, C<sub>3</sub>H<sub>2</sub>), 2.39 (t, *J* = 7.3 Hz, 6H, C<sub>5</sub>H<sub>2</sub>), 1.68 (tt, *J* =

7.3, 6.3 Hz, 6H,  $C_4H_2$ );  $^{13}C$  NMR (101 MHz,  $CDCl_3$ )  $\delta_c$  170.8( $C_2$ ), 134.1 ( $C_1$ ), 51.1 ( $C_5$ ), 36.2( $C_3$ ), 26.0 ( $C_4$ ); LC-MS  $m/z$  ( $ES^+$ ) 429 ( $[M+H]^+$ , 100 %); HRMS  $m/z$  ( $ES^+$ ) found 429.1789  $[M+H]^+$ , expected 429.1774.

## 7.2 Peptide synthesis

### 7.2.1 General SPPS procedures

Resins, synthesis grade reagents, Fmoc protected amino acids, and pseudoproline building blocks were purchased from Novobiochem or Sigma-Aldrich and used without further purification. Resins were preloaded with Fmoc-Gly-OH and the loading estimated using standard procedures.<sup>312</sup> Solid phase peptide synthesis was undertaken manually or was automated using an Invatis ResPep SL system.

### 7.2.2 Fmoc Deprotection

The resin was swelled in DMF for 10 min. After removal of DMF by filtration, the resin was treated with 20 % piperidine (v/v) in DMF (3 × 2 mL) for 5 min. After each treatment the resin was washed with DMF (3 × 2 mL), DCM (3 × 2 mL) and further DMF (3 × 2 mL).

### 7.2.3 Peptide synthesis

#### 7.2.3.1 Manual coupling

Fmoc deprotected resin (typically 20 μM) was swelled in DMF for 10 min. The amino acid (5.0 eq. for natural amino acids, 3.0 eq. for unnatural amino acids) was activated by shaking with coupling reagent (5.0 eq. HBTU for natural amino acids, 3.0 eq. HATU for unnatural amino acids) and DIPEA (10.0 eq. for natural amino acids, 6.0 eq. for unnatural amino acids) in DMF (1 mL) for 10 min. The activated reagent mixture was then added to the drained resin and shaken for 1 hour at RT. The mixture was then removed by filtration, and the resin washed with DMF (3 × 2 mL), DCM (3 × 2 mL) and further DMF (3 × 2 mL). All coupling reactions were performed twice. After the final coupling of a given peptide synthesis, the resin was washed with DMF (3 × 2 mL), DCM (3 × 2 mL), MeOH (3 × 2 mL), and Et<sub>2</sub>O (3 × 2 mL) before being dried overnight *in vacuo*.

#### 7.2.3.2 Automated coupling

The resin (20 μmol per well) was swelled in DMF for 10 min prior to Fmoc deprotection (3 × 20 % piperidine (v/v) in DMF, 400 μL, 10 min) followed by washing with DMF. In a separate mixing tube, the incoming amino acid (5.0 eq. for natural amino acids, 3.0 eq. for unnatural amino acids, both as a 0.5 M solution in NMP) was activated with coupling reagent (5.0 eq. HBTU for natural amino acids, 3.0 eq. HATU for unnatural amino acids, both as a 0.5 M solution in DMF) and NMM (10 eq for natural amino acids, 6.0 eq. for unnatural amino acids). The activated reagent mixture was added to the swelled resin and left for 40 min. All couplings were performed in duplicate. The resin was then

washed with DCM (2 × 400 µL) and DMF (8 × 400 µL) before proceeding to the next Fmoc deprotection and coupling cycle. Once the final amino acid had been coupled, the resin was Fmoc deprotected and washed with DMF (3 × 2 mL), DCM (3 × 2 mL), MeOH (3 × 2 mL), and Et<sub>2</sub>O (3 × 2 mL) before being dried overnight *in vacuo*.

## 7.2.4 N-terminal manipulations

### 7.2.4.1 Acetylation

Fmoc deprotected resin (typically 20 µM) was swelled in DMF for 10 min and then treated with a mixture of 10 % Ac<sub>2</sub>O, 20 % DIPEA in DMF (1 mL) for 1h. The resin was then washed with DMF (3 × 2 mL), DCM (3 × 2 mL), MeOH (3 × 2 mL), and Et<sub>2</sub>O (3 × 2 mL) before being dried overnight *in vacuo*.

### 7.2.4.2 TAMRA labelling

Fmoc Deprotected resin (typically 20 µM) was swelled in DMF for 10 min. 5(6)-carboxytetramethylrhodamine (1.5 eq.) was activated by shaking with HATU (1.49 eq.) and DIPEA (3 eq.) in DMF (1 mL) for 10 min. The activated reagent mixture was added to the drained resin and shaken for 2 hour at RT. The mixture was then removed by filtration, and the resin washed with DMF (3 × 2 mL), DCM (3 × 2 mL) and further DMF (3 × 2 mL) before the coupling was repeated with addition of fresh activated reagent. The resin was washed with DMF (3 × 2 mL), DCM (3 × 2 mL), MeOH (3 × 2 mL), and Et<sub>2</sub>O (3 × 2 mL) before being dried overnight *in vacuo*.

## 7.2.5 Resin displacement

### 7.2.5.1 Cleavage with side chain deprotection

A premixed deprotection cocktail (95 % TFA, 2.5 % TIS, 2.5 % H<sub>2</sub>O, 1.5 mL) was added to dried resin (20 µM) and shaken for 3 hr at RT. The solution was filtered and the resin was washed with additional deprotection cocktail (2 mL). The solution was concentrated under a flow of N<sub>2</sub> before peptide precipitation on addition of ice cold tBME (10mL). The mixture was centrifuged (4000 g, 4 °C, 10 min), filtered, and the pellet washed with further tBME (2 × 10 mL) before being dried overnight *in vacuo* prior to preparative LC-MS purification.

### 7.2.5.2 Mild cleavage preserving side chain protecting groups

This method is suitable for acid sensitive resins such as 2-chlorotrityl chloride. Dried resin was shaken with HFIP in DCM (20 % v/v, 1 mL) for 30 min. The solution was filtered and the resin washed with DCM (2 mL). The combined solution was concentrated *in vacuo* and the crude residue was stored at 4 °C.

### 7.2.6 In solution C-terminal electrophilic trap incorporation

To the crude residue resulting from mild resin displacement (Section 7.2.5.2), PyBop (5 eq) and trimethylamine (10 eq) in DCM (4mL) were added. The reaction was stirred for 5 min before either **VS** or **VME** (5 eq) was added and the reaction stirred overnight. The solution was diluted with DCM (20 mL), washed with 1 M KHSO<sub>4</sub> (2 × 20 mL) and dried *in vacuo*. Side chains were removed as described in Section 7.2.5.1 prior to preparative LC-MS purification.

## 7.2.7 Peptide Characterisation

Peptide	Sequence	MW (g/mol)	ES <sup>+</sup> peaks (Da)	Retention time (min)	Accurate mass (Da)	Scale and Yield	Solvent stock
<b>Pg-12-mer-VS</b>	Ac-Pg-STLHLVLRLRG- <b>VS</b>	1517.9	507 (+3), 760 (+2)	10.3	F: +2 = 759.9238 (M = 1517.8476) E: M = 1517.8501	40 μM, 8.2 %	H <sub>2</sub> O
<b>Pg-12-mer-VME</b>	Ac-Pg- STLHLVLRLRG- <b>VME</b>	1497.9	501 (+3), 750 (+2)	10.6	F: +2 = 749.9379 (M = 1497.8758) E: M = 1497.8779	40 μM, 16.6 %	H <sub>2</sub> O
<b>12-mer-VS</b>	Ac-STLHLVLRLRG- <b>VS</b>	1422.8	476 (+3), 713 (+2)	1.1 (6.6)	F: +2 = 712.4081 (M = 1422.8162) E: M = 1422.8130	40 μM, 2.8 %	DMSO
<b>12-mer-VME</b>	Ac-STLHLVLRLRG- <b>VME</b>	1402.8	469 (+3), 703 (+2)	0.7 (7.06)	F: +2 = 702.4234 (M = 1402.8468) E: M = 1402.8408	40 μM, 13 %	DMSO
<b>12-mer</b>	Ac-STLHLVLRLRGG	1362.8	456 (+3), 683 (+2)	11.7	F: +2 = 682.4046 (M = 1362.8092) E: M = 1362.8096	50 μM, 8.8 %	DMSO
<b>T-12-mer</b>	TAMRA- STLHLVLRLRGG	1732.9	435 (+4), 579 (+3) 868 (+2)	12.2/12.7	F: +2 = 867.4806 (M = 1732.9612) E: M = 1732.9413	20 μM, 20 %	DMSO
<b>T-15-mer</b>	TAMRA-QKESTLHLVLRLRGG	2118.1	425 (+5), 531 (+4) 708 (+3)	11.6/11.9	F: +2 = 1060.0673 (M = 2118.1346) E: M = 2118.1375	20 μM, 30 %	DMSO
<b>NLS</b>	PKKKRKV-Pg-STLHLVLRLRG- <b>VME</b>	2320.5	465 (+5), 581 (+4) 774 (+3)	10.1	n.d	20 μM, 1.7 %	H <sub>2</sub> O
<b>CCP1</b>	[R-Ahx-R]3-Pg- STLHLVLRLRG- <b>VME</b>	2731.7	456 (+6), 547 (+5) 684 (+4)	9.7	n.d	20 μM, 4.2 %	H <sub>2</sub> O
<b>CCP2</b>	[R-Ahx-R]3-ILFQMKWHK- STLHLVLRLRG- <b>VME</b>	3943.4	439 (+9), 494 (+8) 564 (+7), 659 (+6)	10.7	n.d	20 μM, 2.3 %	H <sub>2</sub> O

Table 5: Summary of characterisation data obtained for peptide probes and inhibitors. For accurate mass, F = Found mass, E= expected mass.

## 7.3 Biochemical methods

### 7.3.1 General methods

Ultra pure water was obtained using a MilliQ® Millipore water purification system. SDS-PAGE was carried out on 12 % Bis-Tris unless otherwise stated, using NuPAGE® LDS sample loading buffer and Precision plus protein standard all blue marker (Bio-Rad). In-gel fluorescence was measured using either an ETTAN Dige or a Typhoon FLA 9500 Imager (both GE Healthcare). Chemiluminescence was recorded using a LAS-4000 imaging system (GE Healthcare). Absorbance was measured using a SpectraMax M2/M2e microplate reader (Molecular devices). Cell culture media and reagents were obtained from Sigma-Aldrich. AzTB was synthesised as previously reported.<sup>189</sup> NEM (Sigma-Aldrich), IA (Sigma-Aldrich), DTT (AGTC Bioproducts), and **Biotin-Ahx-Ub-VME** (UbiQ) were purchased. Epirubicin was provided as a 3.4 nM stock in DMSO by Eric Lam.

### 7.3.2 Cell culture

HeLa and MCF7 cells were grown in DMEM supplemented with 10 % v/v FBS in a humidified 5 % CO<sub>2</sub> atmosphere at 37 °C. During normal passaging, cells were detached from the plate with 0.5 % trypsin. Cells were grown to 70-90 % confluence prior to treatment.

### 7.3.3 Cell lysis

The protein concentration of all lysates was determined using the BioRad Protein assay following the manufacturer's instructions and using a BSA produced standard absorbance curve.

#### 7.3.3.1 Shear Lysis

The culture media was aspirated and the cells washed twice with PBS. Lysis buffer (50 mM HEPES, pH 7.2, 10 mM NaCl, 5 mM MgCl<sub>2</sub>, 1 mM DTT, 1 × EDTA-free complete protease inhibitor (Roche Diagnostics), was added to the plate and the cells were scraped into an Eppendorf on ice and shear force lysed (25 G needle). The lysates were centrifuged (17 000 *g*, 20 min), their protein concentration determined, and stored at -80 °C until further use.

#### 7.3.3.2 SDS-free fractionation

Buffers are as described in Table 6. The culture media was aspirated and the cells washed twice with PBS. Fresh PBS (1 mL) was then added and the cells scraped into an Eppendorf tube on ice. The cells were pelleted (1000 *g*, 5 min) and the supernatant discarded. The cells were then resuspended in buffer 1 (200 µL), followed by addition of buffer 2 (200 µL) and rotated at 4 °C for 15 min. The

samples were spun gently (720 *g*, 1 min) and the supernatant transferred to a clean eppendorf. The pellet was gently washed with 100  $\mu$ L of buffer 3, spun (720 *g*, 1 min), and the supernatant transferred to the clean eppendorf. This is the cytoplasmic fraction. To the remaining pellet, buffer 4 (500  $\mu$ L) was added and the samples were rotated at 4 °C for 1 hr and then centrifuged (17000 *g*, 10m min). The supernatant was retained as the nuclear extract. The protein concentration of each fraction was determined, and the samples stored at -80 °C until further use.

Buffer 1	25 mM HEPES, 5 mM KCl, 0.5 mM MgCl <sub>2</sub> , 1 mM DTT, 1 × EDTA-free complete protease inhibitor
Buffer 2	Buffer 1 plus 1 % NP-40
Buffer 3	1:1 mixture of buffer 1 and 2
Buffer 4	25 mM HEPES, 350 mM NaCl, 10 % sucrose, 1 mM DTT, 1 × EDTA-free complete protease inhibitor

**Table 6: Buffers used during SDS-free fractionation**

### 7.3.3.3 SDS-free whole cell lysis

The culture media was aspirated and the cells washed twice with PBS. Lysis buffer (1 % Triton X-100, 1 × EDTA-free complete protease inhibitor (Roche Diagnostic) in PBS), was added to the plate and the cells were scraped into an Eppendorf on ice and left for 20 min. The lysates were centrifuged (17 000 *g*, 20 min), their protein concentration determined, and stored at -80 °C until further use.

### 7.3.3.4 Whole cell lysis

The following lysis protocol was used following in-cell labelling assays only. The culture media was aspirated and the cells washed twice with PBS. Lysis buffer (1 % Triton X-100, 0.1 % SDS, 1 × EDTA-free complete protease inhibitor (Roche Diagnostis) in PBS), was added to the plate and the cells were scraped into an Eppendorf on ice and left for 20 min. The lysates were centrifuged (17 000 *g*, 20 min), their protein concentration determined, and stored at -80 °C until further use.

## 7.3.4 In-cell labelling assays

Total DMSO was normalised for each experiment ( $\leq$  2 %). For experiments with **PYR-41-P**, Cysteine-free DMEM was employed for probe incubations.

### 7.3.4.1 Probe labelling

Probes were diluted in DMEM to a relevant concentration and incubated at 37 °C for 5 min. In some cases, cells were pre-incubated with NEM (200  $\mu$ M) in DMEM for 20 min at 37 °C prior to addition of probe. The cell media was aspirated and replaced with the probe containing media ( $\leq$  0.5 % DMSO final concentration) for a defined time period prior to cell lysis as described in Section 7.3.3.4.



#### *7.3.4.2 Competition studies*

Inhibitors of interest were diluted in DMEM to a relevant concentration and incubated at 37 °C for 5 min. The cell media was aspirated and replaced with inhibitor containing media (DMSO final concentration of 0.5 – 1 % for small molecule probes, 2 % for peptide studies due to poor solubility of competitors) and incubated at 37 °C for 30 min. After this time, the media was removed, probe was added to the media at a relevant concentration and media was gently pipetted back onto the plate. Plates were incubated at 37 °C for a further 30 min (1 hr for peptide studies) prior to cell lysis as described in Section 7.3.3.4.

### **7.3.5 In-lysate labelling assays**

Lysate was generated using methods described in Section 1.3.3.1 and Section 1.3.3.2. For DDR studies, cells were treated with either 1 µM or 0.1 µM epirubicin in DMEM (0.2 % DMSO final concentration) for 6 and 24 hr prior to cell lysis.

#### *7.3.5.1 Probe labelling*

Typically, 40 µg of lysate was incubated with a relevant concentration of probe (< 2 % DMSO final concentration) in a total volume of 40 µL. In some cases, lysate was pre-incubated with 10 mM NEM for 20 min at RT prior to addition of probe. The mixtures were incubated at 37 °C for 1 hr. Protein was subsequently precipitated using either the ice cold MeOH or MeOH/CHCl<sub>3</sub> method (Table 7), air-dried, and resuspended in 2 % SDS in PBS (1 mg/mL) prior to CuAAC (see Section 7.3.6).

#### *7.3.5.2 Competition studies*

Typically, 40 µg of lysate was incubated with an inhibitor of interest at a relevant concentration in a total volume of 40 µL. The mixtures were incubated at 37 °C for 30 min, before a fixed concentration of probe was added (< 2 % DMSO final concentration) and the mixtures were incubated at 37 °C for a further 1 hr. Protein was subsequently precipitated using either the ice cold MeOH or MeOH/CHCl<sub>3</sub> method (Table 7), air-dried, and resuspended in 0.2 % SDS (1 mg/mL) in PBS prior to CuAAC (see Section 7.3.6).

### **7.3.6 CuAAC and in-gel fluorescence**

Typically, 40 µg lysates were adjusted to 1 mg/mL in the presence of 0.2 % SDS. A premixed 'click cocktail' (100 µM AzTB, 1 mM CuSO<sub>4</sub>, 1 mM TCEP, 100 µM TBTA final concentrations) was added and the reactions were shaken vigorously at RT for 1 hr, before being quenched by 10 mM EDTA. The protein was precipitated either using the ice cold MeOH or MeOH/CHCl<sub>3</sub> method (Table 7), Protein

was subsequently air-dried and resuspended in 36  $\mu\text{L}$  of 2 % SDS in PBS. To a 10  $\mu\text{g}$  portion (9  $\mu\text{L}$ ) 4  $\times$  working SDS sample loading buffer (for reducing gels, sample loading buffer contained 4 %  $\beta$ -mercaptoethanol) was added (4  $\mu\text{L}$ ) and the mixtures were boiled at 96  $^{\circ}\text{C}$  for 6 min. Samples were resolved by SDS-PAGE and visualised by in-gel fluorescence. Gel loading was verified using Coomassie stain.

Ice cold MeOH precipitation	MeOH/ $\text{CHCl}_3$ precipitation
<ol style="list-style-type: none"> <li>1. Add 10 <math>\times</math> volume ice cold MeOH</li> <li>2. Store overnight at <math>-80^{\circ}\text{C}</math></li> <li>3. Pellet protein (17000 <math>g</math>, 20 min)</li> <li>4. Remove MeOH and Wash (10 <math>\times</math> MeOH)</li> <li>5. Pellet (17000 <math>g</math>, 10 min)</li> <li>6. Remove MeOH and air-dry</li> </ol>	<ol style="list-style-type: none"> <li>1. Add 2 <math>\times</math> volume MeOH, vortex</li> <li>2. Add 0.5 <math>\times</math> volume <math>\text{CHCl}_3</math>, vortex</li> <li>3. Add 1 <math>\times</math> volume <math>\text{H}_2\text{O}</math>, vortex</li> <li>4. Spin sample (17000 <math>g</math>, 5 min)</li> <li>5. Remove aqueous layer</li> <li>6. Wash (10 <math>\times</math> MeOH)</li> <li>7. Pellet (17000 <math>g</math>, 10 min)</li> <li>8. Wash (1 mL MeOH) and pellet (17000 <math>g</math>, 10 min)</li> <li>9. Remove MeOH and air-dry</li> </ol>

**Table 7: Protein precipitation protocols**

### 7.3.7 Western blot analysis

Proteins were transferred onto PVDF membrane by wet transfer using Tris-glycine transfer buffer supplemented with 0,1 % SDS, 10 % MeOH, or an iBlot<sup>®</sup> Gel Transfer Device (ThermoFisher). Membranes were washed with TBS-T (1  $\times$  TBS, 0.1 % Tween-20) and blocked (5 % milk powder in TBS-T, 1h) before incubating with primary antibody diluted in 3 % milk in TBS-T (see Table 8 for antibody specific dilutions) either overnight (4  $^{\circ}\text{C}$ ) or for 1 hr at RT. The membrane was then washed (3  $\times$  5 min, TBS-T) and incubated with secondary antibody at the appropriate dilution in 3 % milk in TBS-T for 1 hr at RT. For NeutrAvidin-HRP blots, membranes were blocked with 3 % BSA and the protein-conjugate diluted in 0.3 % BSA. The membrane was washed (4  $\times$  5 min, TBS-T) and developed with Luminata crescendo Western HRP substrate (Millipore).

Antibody	Supplier	Cat.no	Dilution	Secondary	/clonal
Anti- $\gamma$ -H2A.X	Cell signalling	97185	1:1000	rabbit	mono
Anti-Tubulin	Santa Cruz	sc-53646	1:200	Mouse	mono
Anti-mouse	Advansta	R-05071-500	1:20,000	n/a	-
Anti-Rabbit	Advansta	R-05072-500	1:10,000	n/a	-
NeutrAvidin-HRP (Biotin binding protein conjugate)	Invitrogen	A2664	1:2000	n/a	-

**Table 8: Antibodies and protein conjugates used in western blot studies.**

### 7.3.8 Live cell imaging

MCF7 cells were seeded in 12 well plates at ~50 % confluence for live cell imaging applications. **T-12-mer**, **T-15-mer**, and **12-mer** were separately diluted in DMEM to 10  $\mu$ M and incubated at 37 °C for 5 min. The cell media was then aspirated and replaced with the probe containing media ( $\leq$  0.5 % DMSO final concentration) for 1 hr. Hoechst nuclear stain (ThermoFisher) was diluted in Hank's PBS (8.12  $\mu$ M final concentration). The cell media was aspirated, replaced with this staining solution and further incubated for 30 min. The cells were then gently washed with Hank's PBS before imaging using Cy3 and DAPI filters with a Zeiss Z1 observer microscope. All imaging was performed with the help of Ernest So (Imperial College London).

### 7.3.9 E2-E3 trap assembly

For the following experiments, SDS-PAGE was carried out using pre-cast 4-12 % gradient Bis-Tris gels (Expedeon).

#### 7.3.9.1 Mono-Ubiquitin incorporation

To UbG76C, (1  $\mu$ L, 4 mg/mL) in MBS buffer (18.5  $\mu$ L, 20 mM MES, 150 mM NaCl, 0.5 mM TCEP, pH 6.5) on ice, **Trap C** (0.5  $\mu$ L of a 500 mM solution in DMSO) was added. The solution was mixed gently and incubated for 3 min at 0 °C, before continuing to the ligase addition step described in Section 7.3.9.2. For analysis, DTT was added (1  $\mu$ L, 20 mM, 1mM final concentration) to quench the reaction. 4  $\times$  working SDS sample loading buffer was added (7  $\mu$ L) and the samples were resolved by SDS-PAGE and visualised by Coomassie stain.

#### 7.3.9.2 E2 / E3 ligase additions

Following incubation on ice as described in Section 7.3.9.1, excess maleimide was removed using a Zebra™ 7 kDa MWCO Spin Desalting column (Thermo Scientific) following the manufacturer's protocol. UbcH7 (1.5  $\mu$ L, 5.6 mg/mL) and a selected HHARI fragment (HH94-397 (W379A), 5.2  $\mu$ L, 3.2 mg/mL; or R<sub>2</sub>, 1.5  $\mu$ L, 15.7 mg/mL) was then added and incubated for a further 2 min at 0 °C. DTT was added (1  $\mu$ L, 20 mM, 1mM final concentration) to quench the reaction. 4  $\times$  working SDS sample loading buffer was added (7  $\mu$ L) and the samples were resolved by SDS-PAGE and visualised by Coomassie stain.

## 7.4 Chemical proteomics

### 7.4.1 General methods

R10K8 DMEM media was purchased from Dundee life products and supplemented with 10 % v/v dialysed FBS (Sigma-Aldrich). During normal passaging, R10K8 labelled cells were detached using a trypsin free cell dissociation buffer (Sigma-Aldrich). All proteomic buffers were prepared fresh and filtered (0.2  $\mu$ M) before use. Work areas were cleaned with 70 % IMS and hair tied back to avoid sample contamination. Protein LowBind tubes (Eppendorf) were used for all CuAAC, pull down and stage tipping applications.

### 7.4.2 Lysate preparation

#### 7.4.2.1 Non-labelled lysate

Non-R10K8 labelled cell lysates used for non-quantitative, LFQ, or Spike-in proteomic experiments were prepared as described in Section 7.3.4.

#### 7.4.2.2 Spike-in SILAC standard formation

Cells were grown in DMEM growth media containing  $^{13}\text{C}$   $^{15}\text{N}$  labelled arginine and  $^{13}\text{C}$   $^{15}\text{N}$  labelled lysine (R10K8) for > 5 passages to ensure sufficient incorporation of heavy label. Further, cells were started from a frozen stock which had previously been shown to be >97 % heavy amino acid incorporated.<sup>287</sup> This level of incorporation is sufficient for SILAC applications.<sup>313</sup> For in-cell labelling, the spike-in standard was formed by treating these cells with probe at a defined concentration in R10K8 DMEM for 30 min or 1 hr ( $\leq$  0.5 % DMSO final concentration). The experiment was performed on a large scale (at least 10x 10 cm plates) in order to produce sufficient standard for use in all proteomic experiments for a given probe. Cells on each plate were lysed as described in Section 7.3.3.4 and pooled into a single 'master spike' prior to protein concentration determination.

For in lysate analysis of **Pg-12-mer-VME**, cells were heavy incorporated as described above and lysed as described in sections 7.3.3. Enough plates were lysed in order to produce a sufficient quantity of standard for use in all proteomic experiments for a given probe (typically 10  $\times$  10 cm plates). The spike-in standard was formed by treating this lysate with probe (50  $\mu$ M) for 1 hr. Protein was subsequently precipitated by the MeOH/ $\text{CHCl}_3$  method (Table 7), air-dried, and resuspended in 0.2 % SDS in PBS.

### 7.4.3 Capture and enrichment of MS-based proteomics

Preliminary analysis of **PYR-41-P**, **Pg-12-mer-VS** and **Pg-12-mer-VME** in-lysate labelling was performed without replicates on a 100 µg scale. All other samples were prepared in biological replicates, with non-SILAC samples prepared on a 400 µg scale and Spike-in SILAC samples prepared on a 600 µg scale. Samples were subjected to CuAAC as before (see Section 7.3.6) with the following modifications: CuAAC was undertaken at a protein concentration of 2 mg/mL; lysate was pre-mixed (in a defined ratio of spike-in standard: non labelled sample) prior to CuAAC for spike-in SILAC samples. After precipitating and air-drying, the protein pellet was re-suspended in 2 % SDS / PBS at 10 mg/mL before being diluted 10 × with PBS (1 mg/mL, 0.2 % SDS / PBS final concentration). Alternatively, following **Biotin-Ahx-Ub-VME** incubation, excess probe was removed by buffer exchange (1 mg/mL, 0.2 % SDS / PBS final concentration) using an Amicon® Ultra 10 kDa MWCO Spin filter.

For sample enrichment, 30 µL of NeutrAvidin agarose resin (Thermo Scientific) was pre-washed (0.2 % SDS / PBS, 3 × 300 µL). Beads were gently vortexed (1min) and collected by centrifugation (3000 g, 2 min). Samples were then added to the resin and incubated with gentle shaking for 2 hr at RT. The supernatant was removed, and the beads were washed consecutively with: 1 % SDS in PBS (3 × 300 µL); 4M Urea in PBS (2 × 300 µL); 50 mM AMBIC (5 × 300 µL).

### 7.4.4 On bead reduction, alkylation and digest

The washed beads were reduced with 10 mM DTT / 50 mM AMBIC (50 µL total volume), spun down briefly and incubated with shaking at 55 °C for 30 min. The samples were then spun briefly and washed with further 50 mM AMBIC (2 × 300 µL). Exposed cysteines were alkylated by addition of 10 mM IA / 50 mM AMBIC (50 µL total volume) and incubation in the dark without shaking for 30 min. The samples were then washed with further 50 mM AMBIC (2 × 300 µL), treated with trypsin (5 µL, 0.2 mg/mL, Promega) in 50 mM AMBIC (50 µL) and digested overnight shaking at 37 °C. Following digestion 50 mM AMBIC was added (80 µL), the samples were shaken (10 min), spun down (3 000 g, 2 min) and the supernatant retained. The beads were further washed by shaking with 0.1 % TFA / H<sub>2</sub>O (80 µL, 10 min), spun down (3 000 g, 2 min) and the supernatant retained. The combined supernatants were desalted by stage tipping using SDB-XC (Styrene Divinyl Benzene) extraction disks, eluted with 79 % MeCN / H<sub>2</sub>O and dried *in vacuo*.<sup>314</sup> Dried peptides were stored at -80 °C.

## 7.4.5 LC/MS/MS

### 7.4.5.1 Sample preparation

Prior to LC/MS/MS runs, the dried peptide residues were resuspended in 0.5 % TFA / 2 % MeOH in H<sub>2</sub>O (20 µL), spun down (13 000 g, 10 min) and the supernatant transferred into auto-sampler vials.

### 7.4.5.2 LC/MS/MS data acquisition

All analysis was performed on an Easy nLC-1000 system coupled to a QExactive mass spectrometer via an easy-spray source (all Thermo Fisher Scientific). Typically, 3 µL sample injections were employed for enriched samples. Samples were separated using a reverse phase Acclaim PepMap RSLC column (50 cm × 75 µm inner diameter, Thermo Fisher Scientific) across a 2 hr acetonitrile gradient in 0.1 % formic acid at a flow rate of 250 nL/min. The QExactive mass spectrometer was operated in data-dependent mode with survey scans acquired at a resolution of 75,000 at m/z 200 (transient time 256 ms). Up to the top 10 most abundant isotope patterns with charge +2 from the survey scan were selected with an isolation window of 3.0 m/z and fragmented by HCD with normalized collision energies of 25 W. The maximum ion injection times for the survey scan and the MS/MS scans (acquired with a resolution of 17,500 at m/z 200) were 250 and 80 ms, respectively. The ion target value for MS was set to 10<sup>6</sup> and for MS/MS to 10<sup>5</sup>, and the intensity threshold was set to 8.3 × 10<sup>2</sup>.

### 7.4.5.3 Data analysis

The raw data file obtained from each LC-MS/MS acquisition was processed using MaxQuant software (version 1.3.0.5). Peptides were identified from the MS/MS spectra by searching against the *Homo sapiens* UniProtKB database, using the Andromeda search engine. Cysteine carbamidomethylation was set as a fixed modification, whilst methionine oxidation and N-terminal acetylation were set as variable modifications. The multiplicity of the search was set to: 1 for LFQ quantification; 2 for spike-in SILAC quantitation. Heavy labels present were selected (R10 and K8) for spike-in SILAC. All other parameters were used as pre-set by the software. Label free quantification was performed using a built in algorithm with the 'match between runs' (across a 2 min time window) enabled. Peptides and proteins were identified utilising a 0.01 false discovery rate, with "Unique and razor peptides" mode selected for both identification and quantification of proteins (razor peptides are uniquely assigned to protein groups and not to individual proteins). At least 2 razor + unique peptides were required for valid quantification.

Data outputted from MaxQuant was analysed using a combination of Perseus version 1.5 and Microsoft Office Excel 2010. For all data sets, protein identifications by MaxQuant based on

'contaminants', 'only identified by site' and 'reverse' were filtered out in the first instance. Further filtering only allowed inclusion of a protein target if it was identified in at least two biological replicates.

For LFQ quantification, the filtered intensities were logarithmised ( $\text{Log}_2$ ), and missing values (representing low abundance measurements) were replaced with random numbers imputed from a normal distribution chosen to best simulate low abundance values. A student's T-test was applied to determine statistical significance. For spike-in SILAC experiments, the generated H/L ratios were logarithmised ( $\text{Log}_2$ ), and normalised against their column median prior to averaging across biological replicates. Relative ratio changes on competition were determined by normalising all ratios to the probe only control sample. A relative ratio change  $>1.5$  fold (equivalent to  $\text{Log}_2 > 0.58$ ) was deemed a sufficient threshold for target identification. A student's T-test was applied across replicates with 0.05 FDR to determine statistical significance of competition.

The gene names of targets of interests were inputted into the STRING database (<http://string-db.org/>) to generate interaction networks to determine the enrichment of GO annotations. Raw data and data analysis is available as electronic files for further reference. Please refer to Appendix Table 1.

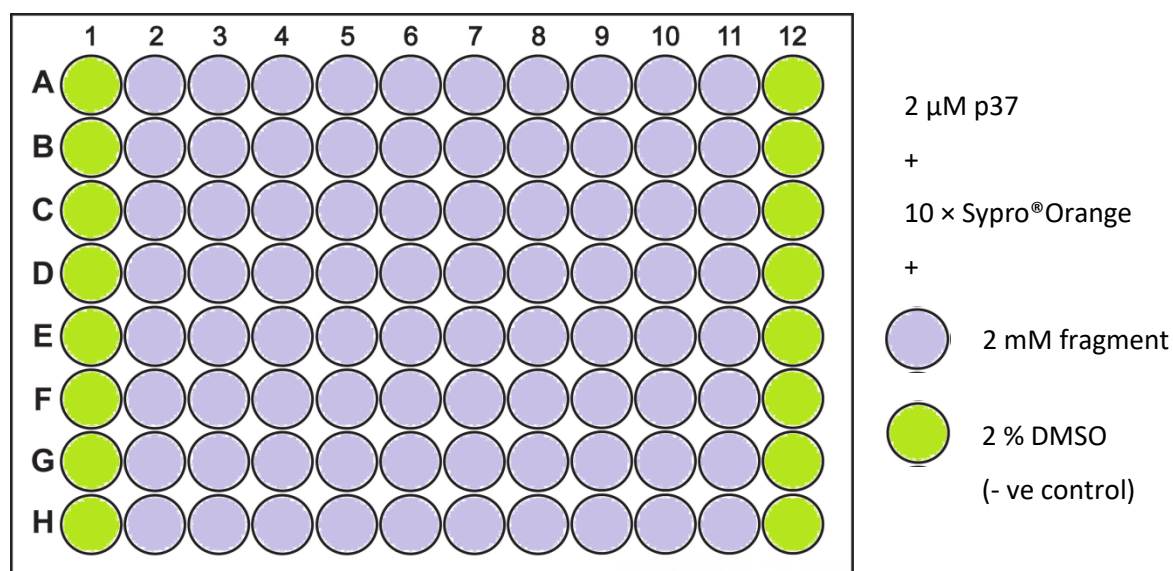
## 7.5 Fragment-Based Drug Discovery (FBDD) workflow

### 7.5.1 Differential Scanning Fluorimetry (DSF)

The DSF assay was performed with indicated concentrations of P37 and fragment, in 96 well semi-skirted real time PCR plates, using a Stratagen MX3005p qPCR machine (Agilent technologies). All experiments were performed using filtered DSF buffer (50 mM HEPES, 150 mM NaCl, 1 mM DTT, pH 7.5).

#### 7.5.1.1 Fragment screen

The Ro3 fragment library (Maybridge) was stored in sealed 96 well 'master plate' arrays as 100 mM DMSO stocks at RT under a N<sub>2</sub> atmosphere using a FluidX system. The first and final lanes contained 100 % DMSO, to allow the inclusion of negative control readings. A DSF buffer solution (2 mL) containing SYPRO® Orange dye (5 µL of 5000x stock in DMSO) and P37 (2.5 µM) was arrayed (16 µL) into a clean PCR plate. To a clean 96 well 'daughter plate', ultra-pure H<sub>2</sub>O (4.5 µL) was dispensed, followed by the fragment array (0.5 µL, 10 mM final concentration). The diluted fragments and controls were then arrayed (4 µL) onto the PCR plate (Figure 1) and mixed (10 × SYPRO® Orange, 2 µM P37, 2 mM fragment, 2 % DMSO final concentrations). Fluorescence was measured in each well over a 25-95 °C gradient.



**Figure 70: Layout of Library plates and subsequent DSF assay plates.**

#### 7.5.1.2 Data analysis and triage

Data was analysed using an Excel Script for DSF analysis, publically available from the Structural genomics consortium (SGC): <ftp:ftp.sgc.ox.ac.uk/pup/biophysics>. T<sub>m</sub> shifts were calculated by fitting



the data to the Boltzmann equation in Graphpad Prism 5 software. Assay data was deemed high quality for plates where the standard deviation of the control reading was  $\leq 0.1^\circ\text{C}$ . Any fragments with a  $\Delta T_m$  value  $> 4$  s.d from the control was carried through the workflow as a possible hit fragment. Hit were grouped by structural similarity using ChemBiofinder Std 13.0 software. Representative fragments from each promising structural group were selected and carried forward for NMR validation.

## 7.5.2 WaterLOGSY experiment

All experiments were performed in DSF buffer (see Section 7.5.1) at 298K (25 °C) using a Bruker Avance III 600 MHz spectrometer.

### 7.5.2.1 Spectra acquisition and analysis

In an NMR tube DSF buffer (450  $\mu\text{L}$ ) and  $\text{D}_2\text{O}$  (45  $\mu\text{L}$ , 10 % final concentration) were mixed and spun down. A 1D  $^1\text{H}$  spectrum of the buffer background was acquired and the system calibrated to maximise water suppression. The fragment (4.5  $\mu\text{L}$  of 100 mM stock, 1 mM final concentration) was then added and the sample spun and mixed. A 1D  $^1\text{H}$  spectrum of the fragment ligand was acquired, followed by a 1D waterLOGSY spectrum. P37 (15  $\mu\text{L}$  of 300  $\mu\text{M}$  stock in DSF buffer, 10  $\mu\text{M}$  final concentrations) was then added, the sample was spun and mixed, and a 1D waterLOGSY spectrum in the presence of protein was acquired. When necessary, a further waterLOGSY spectrum was recorded at higher protein concentration (additional 45  $\mu\text{L}$  P37 and 0.5  $\mu\text{L}$  Fragment, 40  $\mu\text{M}$  and 1mM final concentrations respectively). WaterLOGSY spectra of samples containing P37 alone (at 10 and 40  $\mu\text{M}$ , in DSF buffer, 10 %  $\text{D}_2\text{O}$ ) were also recorded to allow protein background subtraction. Following background subtractions, the  $^1\text{H}$  and waterLOGSY spectra of each fragment with and without protein were overlaid. For a given fragment, a positive inversion of any waterLOGSY in the presence of P37 was described as a validated hit.

## 7.5.3 Micro-scale thermophoresis

All experiments and analysis was undertaken at the Crick Institute (Mill Hill) with the help of Dr Steve Martin.

### 7.5.3.1 P37 labelling

P37 was labelled with NT-647-NHS fluorescent dye using a NT protein labelling kit (Monolith) with some adjustments to the manufacturer's instructions. Briefly, the DSF buffer was exchanged for MST buffer (50 mM HEPES, 100 mM NaCl, 0.5 mM TCEP, pH 7.5). This P37 solution (10  $\mu\text{M}$ , 100  $\mu\text{L}$ ) was

mixed with dye solution (200  $\mu$ M, 100  $\mu$ L) and incubated in the dark for 45 min. Excess dye was removed by gel filtration. The resulting labelled protein, P37-NT-647, was measured by absorbance spectroscopy and found to have a sufficient labelling efficiency of 33 % (6  $\mu$ M protein concentration). A further ubiquitination assay was performed by Katrin Rittinger's group to check labelled protein activity. Labelled and wild type P37 was separately incubated with UBA1, ATP, Ubch7 and Ub for 4 hours at 25 °C. Time course samples were separated by SDS-PAGE and visualised using Coomassie staining.

### *7.5.3.2 Data acquisition and analysis*

0.05 % Tween was added to the MST buffer, to ensure good sample loading into the capillaries. A labelled protein stock (370  $\mu$ L MST buffer, 10  $\mu$ L DMSO, 20  $\mu$ L P37-NT-647) was mixed and kept in foil on ice. 16 vials (100  $\mu$ L eppendorfs) were labelled numerically. Fragment solution (20  $\mu$ L, 10 mM in MST buffer, 2.5 % final DMSO concentration) was added to vial 1. MST buffer (10  $\mu$ L) was added to vials 2-16. Half the solution in vial 1 (10  $\mu$ L) was transferred to vial 2. Vial 2 was mixed by pipetting before solution transfer (10  $\mu$ L) to vial 3. This serial dilution was repeated 15 times across all vials, with 10  $\mu$ L being removed from the final vial to leave 10  $\mu$ L. The labelled protein stock (10  $\mu$ L) was added to each vial, mixing well by pipetting. Samples (100 nM labelled protein, 5 mM- 0.15  $\mu$ M fragment, ~2.5 % DMSO final concentrations) were incubated for 5 min, the capillaries filled (~5  $\mu$ L) and transferred into the Monolith NT.115 (Nanotemper technologies) for fluorescence reading at 670 nm. Capillaries were spun to minimise background.  $K_d$  values were calculated from initial and thermophoretic fluorescence data by Steve Martin using Nanotemper analysis software.

# Bibliography

---

1. Mann, M.; Jensen, O. N., Proteomic analysis of post-translational modifications. *Nat Biotech* **2003**, *21*, 255-261.
2. Ward, J. Activity-Based Protein Profiling of E3 Ubiquitin Ligases in the DNA Damage Response. Manuscript, Imperial College, London, 2012.
3. Bergink, S.; Jentsch, S., Principles of ubiquitin and SUMO modifications in DNA repair. *Nature* **2009**, *458*, 461-467.
4. Al-Hakim, A.; Escribano-Diaz, C.; Landry, M.-C.; O'Donnell, L.; Panier, S.; Szilard, R. K.; Durocher, D., The ubiquitous role of ubiquitin in the DNA damage response. *DNA Repair* **2010**, *9*, 1229-1240.
5. Gideon, G.; Scheid, M.; Hammerling, U.; Boyse, E. A.; Schlesinger, D. H.; Niall, H. D., Isolation of a Polypeptide that has Lymphocyte-Differentiating Properties and is Probably Represented Universally in Living Cells. *Proc Natl Acad Sci U S A* **1975**, *72*, 11-15.
6. Goldknopf, I. L.; Busch, H., Isopeptide linkage between nonhistone and histone 2A polypeptides of chromosomal conjugate-protein A24. *Proc Natl Acad Sci U S A* **1977**, *74*, 864-868.
7. Hunt, L. T.; Dayhoff, m. O., Amino-terminal sequence identity of ubiquitin and the nonhistone component of nuclear protein A24. *Biochem. Biophys. Res. Commun* **1977**, *74*, 650-655.
8. Hershko, A.; Ciechanover, A. H., H.; Haas, A. L.; Rose, I. A., Proposed role of ATP in protein breakdown: Conjugation of proteins with multiple chains of the polypeptide of ATP-dependent proteolysis. *Proc Natl Acad Sci U S A* **1980**, *77*, 1783-1786.
9. Wilkinson, K. D.; Urban, M. K.; Haas, A. L., Ubiquitin is the ATP-dependent proteolysis factor I of rabbit reticulocytes. *J Biol Chem* **1980**, *255*, 7529-7532.
10. Ciechanover, A., Intracellular protein degradation: from a vague idea thru the lysosome and the ubiquitin-proteasome system and onto human diseases and drug targeting. *Hematology Am. Soc. Hematol. Educ. Program* **2006**, *1*, 1-12, 505-6.
11. Ciechanover A; Elias S; Heller H; Ferber S; A., H., Characterization of the heat-stable polypeptide of the ATP-dependent proteolytic system from reticulocytes. *J. Biol. Chem.* **1980**, *255*, 7525-7528.
12. Ciechanover A; Heller H; Elias S; Haas AL; A, H., ATP-dependent conjugation of reticulocyte proteins with the polypeptide required for protein degradation. *Proc Natl Acad Sci U S A* **1980**, *77*, 1365-1368.
13. Hicke, L., Protein regulation by monoubiquitin. *Nat Rev Mol Cell Biol* **2001**, *2*, 195-201.
14. Haglund, K.; Di Fiore, P. P.; Dikic, I., Distinct monoubiquitin signals in receptor endocytosis. *Trends Biochem Sci* **2003**, *28*, 598-603.
15. Huang, T. T.; D'Andrea, A. D., Regulation of DNA repair by ubiquitylation. *Nat Rev Mol Cell Biol* **2006**, *7*, 323-34.
16. Shilatifard, A., Chromatin modifications by methylation and ubiquitination: implications in the regulation of gene expression. *Annu Rev Biochem* **2006**, *75*, 243-69.
17. Pickart, C. M.; Cohen, R. E., Proteasomes and their kin: proteases in the machine age. *Nat Rev Mol Cell Biol* **2004**, *5*, 177-87.
18. Chen, Z. J.; Sun, L. J., Nonproteolytic functions of ubiquitin in cell signaling. *Mol Cell* **2009**, *33*, 275-86.
19. Ikeda, F.; Dikic, I., Atypical ubiquitin chains: new molecular signals. 'Protein Modifications: Beyond the Usual Suspects' review series. *EMBO Rep* **2008**, *9*, 536-42.
20. Kim, H. T.; Kim, K. P.; Lledias, F.; Kisselev, A. F.; Scaglione, K. M.; Skowyra, D.; Gygi, S. P.; Goldberg, A. L., Certain pairs of ubiquitin-conjugating enzymes (E2s) and ubiquitin-protein ligases (E3s) synthesize nondegradable forked ubiquitin chains containing all possible isopeptide linkages. *J Biol Chem* **2007**, *282*, 17375-86.

21. Meyer, H. J.; Rape, M., Enhanced protein degradation by branched ubiquitin chains. *Cell* **2014**, *157*, 910-21.
22. Kulathu, Y.; Komander, D., Atypical ubiquitylation - the unexplored world of polyubiquitin beyond Lys48 and Lys63 linkages. *Nat Rev Mol Cell Biol* **2012**, *13*, 508-23.
23. Komander, D.; Reyes-Turcu, F.; Licchesi, J. D.; Odenwaelder, P.; Wilkinson, K. D.; Barford, D., Molecular discrimination of structurally equivalent Lys 63-linked and linear polyubiquitin chains. *EMBO Rep* **2009**, *10*, 466-73.
24. Kirisako, T.; Kamei, K.; Murata, S.; Kato, M.; Fukumoto, H.; Kanie, M.; Sano, S.; Tokunaga, F.; Tanaka, K.; Iwai, K., A ubiquitin ligase complex assembles linear polyubiquitin chains. *EMBO J* **2006**, *25*, 4877-87.
25. Tokunaga, F.; Sakata, S.; Saeki, Y.; Satomi, Y.; Kirisako, T.; Kamei, K.; Nakagawa, T.; Kato, M.; Murata, S.; Yamaoka, S.; Yamamoto, M.; Akira, S.; Takao, T.; Tanaka, K.; Iwai, K., Involvement of linear polyubiquitylation of NEMO in NF-kappaB activation. *Nat Cell Biol* **2009**, *11*, 123-32.
26. Ciechanover, A.; Heller, H.; Katz-Etzion, R.; Hershko, A., Activation of the heat-stable polypeptide of the ATP-dependent proteolytic system. *Proc Natl Acad Sci U S A* **1981**, *78*, 761-5.
27. Hershko, A.; Ciechanover, A.; Rose, I. A., Identification of the active amino acid residue of the polypeptide of ATP-dependent protein breakdown. *J Biol Chem* **1981**, *256*, 1525-8.
28. Hershko, A.; Heller, H.; Elias, S.; Ciechanover, A., Components of ubiquitin-protein ligase system. Resolution, affinity purification, and role in protein breakdown. *J Biol Chem* **1983**, *258*, 8206-14.
29. Lub, S.; Maes, K.; Menu, E.; De Bruyne, E.; Vanderkerken, K.; Van Valckenborgh, E., Novel strategies to target the ubiquitin proteasome system in multiple myeloma. *Oncotarget* **2016**, *7*, 6521-37.
30. Morreale, F. E.; Walden, H., Types of Ubiquitin Ligases. *Cell* **2016**, *165*, 248-248
31. Stewart, M. D.; Ritterhoff, T.; Klevit, R. E.; Brzovic, P. S., E2 enzymes: more than just middle men. *Cell Res* **2016**, *26*, 423-40.
32. Wenzel, D. M.; Lissounov, A.; Brzovic, P. S.; Klevit, R. E., UBC7 reactivity profile reveals parkin and HHARI to be RING/HECT hybrids. *Nature* **2011**, *474*, 105-108.
33. Pruneda, J. N.; Stoll, K. E.; Bolton, L. J.; Brzovic, P. S.; Klevit, R. E., Ubiquitin in motion: structural studies of the ubiquitin-conjugating enzyme approximately ubiquitin conjugate. *Biochemistry* **2011**, *50*, 1624-33.
34. Pruneda, J. N.; Littlefield, P. J.; Soss, S. E.; Nordquist, K. A.; Chazin, W. J.; Brzovic, P. S.; Klevit, R. E., Structure of an E3:E2~Ub complex reveals an allosteric mechanism shared among RING/U-box ligases. *Mol Cell* **2012**, *47*, 933-42.
35. Dou, H.; Buetow, L.; Sibbet, G. J.; Cameron, K.; Huang, D. T., BIRC7-E2 ubiquitin conjugate structure reveals the mechanism of ubiquitin transfer by a RING dimer. *Nat Struct Mol Biol* **2012**, *19*, 876-83.
36. Branigan, E.; Plechanovova, A.; Jaffray, E. G.; Naismith, J. H.; Hay, R. T., Structural basis for the RING-catalyzed synthesis of K63-linked ubiquitin chains. *Nat Struct Mol Biol* **2015**, *22*, 597-602.
37. Plechanovova, A.; Jaffray, E. G.; Tatham, M. H.; Naismith, J. H.; Hay, R. T., Structure of a RING E3 ligase and ubiquitin-loaded E2 primed for catalysis. *Nature* **2012**, *489*, 115-20.
38. Heride, C.; Urbe, S.; Clague, M. J., Ubiquitin code assembly and disassembly. *Curr Biol* **2014**, *24*, R215-20.
39. Hofmann, R. M.; Pickart, C. M., Noncanonical MMS2-encoded ubiquitin-conjugating enzyme functions in assembly of novel polyubiquitin chains for DNA repair. *Cell* **1999**, *96*, 645-53.
40. Wu, H.; Pomeroy, S. L.; Ferreira, M.; Teider, N.; Mariani, J.; Nakayama, K. I.; Hatakeyama, S.; Tron, V. A.; Saltibus, L. F.; Spyropoulos, L.; Leng, R. P., UBE4B promotes Hdm2-mediated degradation of the tumor suppressor p53. *Nat Med* **2011**, *17*, 347-55.
41. Metzger, M. B.; Hristova, V. A.; Weissman, A. M., HECT and RING finger families of E3 ubiquitin ligases at a glance. *J Cell Sci* **2012**, *125*, 531-7.

42. Paul, I.; Ghosh, M. K., The E3 ligase CHIP: insights into its structure and regulation. *Biomed Res Int* **2014**, *2014*, 918183.
43. Sang, Y.; Yan, F.; Ren, X., The role and mechanism of CRL4 E3 ubiquitin ligase in cancer and its potential therapy implications. *Oncotarget* **2015**, *6*, 42590-602.
44. Ogunjimi, A. A.; Briant, D. J.; Pece-Barbara, N.; Le Roy, C.; Di Guglielmo, G. M.; Kavsak, P.; Rasmussen, R. K.; Seet, B. T.; Sicheri, F.; Wrana, J. L., Regulation of Smurf2 ubiquitin ligase activity by anchoring the E2 to the HECT domain. *Mol Cell* **2005**, *19*, 297-308.
45. Bonni, S.; Wang, H. R.; Causing, C. G.; Kavsak, P.; Stroschein, S. L.; Luo, K.; Wrana, J. L., TGF-beta induces assembly of a Smad2-Smurf2 ubiquitin ligase complex that targets SnoN for degradation. *Nat Cell Biol* **2001**, *3*, 587-95.
46. Kamadurai, H. B.; Qiu, Y.; Deng, A.; Harrison, J. S.; Macdonald, C.; Actis, M.; Rodrigues, P.; Miller, D. J.; Souphron, J.; Lewis, S. M.; Kurinov, I.; Fujii, N.; Hammel, M.; Piper, R.; Kuhlman, B.; Schulman, B. A., Mechanism of ubiquitin ligation and lysine prioritization by a HECT E3. *Elife* **2013**, *2*, e00828.
47. Lechtenberg, B. C.; Rajput, A.; Sanishvili, R.; Dobaczewska, M. K.; Ware, C. F.; Mace, P. D.; Riedl, S. J., Structure of a HOIP/E2~ubiquitin complex reveals RBR E3 ligase mechanism and regulation. *Nature* **2016**, *529*, 546-50.
48. Hoppe, T., Multiubiquitylation by E4 enzymes: 'one size' doesn't fit all. *Trends Biochem Sci* **2005**, *30*, 183-7.
49. Jacq, X.; Kemp, M.; Martin, N. M.; Jackson, S. P., Deubiquitylating enzymes and DNA damage response pathways. *Cell Biochem Biophys* **2013**, *67*, 25-43.
50. Reyes-Turcu, F. E.; Ventii, K. H.; Wilkinson, K. D., Regulation and cellular roles of ubiquitin-specific deubiquitinating enzymes. *Annu Rev Biochem* **2009**, *78*, 363-97.
51. Komander, D.; Clague, M. J.; Urbe, S., Breaking the chains: structure and function of the deubiquitinases. *Nat Rev Mol Cell Biol* **2009**, *10*, 550-63.
52. Spratt, Donald E.; Walden, H.; Shaw, Gary S., RBR E3 ubiquitin ligases: new structures, new insights, new questions. *Biochemical Journal* **2014**, *458*, 421-437.
53. Ronau, J. A.; Beckmann, J. F.; Hochstrasser, M., Substrate specificity of the ubiquitin and Ubl proteases. *Cell Res* **2016**, *26*, 441-56.
54. Metzger, M. B.; Pruneda, J. N.; Klevit, R. E.; Weissman, A. M., RING-type E3 ligases: master manipulators of E2 ubiquitin-conjugating enzymes and ubiquitination. *Biochim Biophys Acta* **2014**, *1843*, 47-60.
55. Scott, D.; Oldham, N. J.; Strachan, J.; Searle, M. S.; Layfield, R., Ubiquitin-binding domains: mechanisms of ubiquitin recognition and use as tools to investigate ubiquitin-modified proteomes. *Proteomics* **2015**, *15*, 844-61.
56. Deveraux, Q.; Ustrell, V.; Pickart, C.; Rechsteiner, M., A 26-S Protease Subunit That Binds Ubiquitin Conjugates. *Journal of Biological Chemistry* **1994**, *269*, 7059-7061.
57. Husnjak, K.; Dikic, I., Ubiquitin-binding proteins: decoders of ubiquitin-mediated cellular functions. *Annu Rev Biochem* **2012**, *81*, 291-322.
58. Dikic, I.; Wakatsuki, S.; Walters, K. J., Ubiquitin-binding domains - from structures to functions. *Nat Rev Mol Cell Biol* **2009**, *10*, 659-71.
59. Burroughs, A. M.; Balaji, S.; Iyer, L. M.; Aravind, L., Small but versatile: the extraordinary functional and structural diversity of the beta-grasp fold. *Biol Direct* **2007**, *2*, 18.
60. van der Veen, A. G.; Ploegh, H. L., Ubiquitin-like proteins. *Annu Rev Biochem* **2012**, *81*, 323-57.
61. Mahajan, R.; Delphin, C.; Guan, T.; Gerace, L.; Melchior, F., A small ubiquitin-related polypeptide involved in targeting RanGAP1 to nuclear pore complex protein RanBP2. *Cell* **1997**, *88*, 97-107.
62. Gareau, J. R.; Lima, C. D., The SUMO pathway: emerging mechanisms that shape specificity, conjugation and recognition. *Nat Rev Mol Cell Biol* **2010**, *11*, 861-71.

63. Xu, Y.; Plechanovova, A.; Simpson, P.; Marchant, J.; Leidecker, O.; Kraatz, S.; Hay, R. T.; Matthews, S. J., Structural insight into SUMO chain recognition and manipulation by the ubiquitin ligase RNF4. *Nat Commun* **2014**, *5*, 4217.
64. Guzzo, C. M.; Matunis, M. J., Expanding SUMO and ubiquitin-mediated signaling through hybrid SUMO-ubiquitin chains and their receptors. *Cell Cycle* **2013**, *12*, 1015-7.
65. Hoegel, C.; Pfander, B.; Moldovan, G. L.; Pyrowolakis, G.; Jentsch, S., RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. *Nature* **2002**, *419*, 135-41.
66. Jeram, S. M.; Srikumar, T.; Zhang, X. D.; Anne Eisenhauer, H.; Rogers, R.; Pedrioli, P. G.; Matunis, M.; Raught, B., An improved SUMmOn-based methodology for the identification of ubiquitin and ubiquitin-like protein conjugation sites identifies novel ubiquitin-like protein chain linkages. *Proteomics* **2010**, *10*, 254-65.
67. Deshaies, R. J.; Emberley, E. D.; Saha, A., Control of cullin-ring ubiquitin ligase activity by nedd8. *Subcell Biochem* **2010**, *54*, 41-56.
68. Souphron, J.; Waddell, M. B.; Paydar, A.; Tokgoz-Gromley, Z.; Roussel, M. F.; Schulman, B. A., Structural dissection of a gating mechanism preventing misactivation of ubiquitin by NEDD8's E1. *Biochemistry* **2008**, *47*, 8961-9.
69. Liu, J.; Furukawa, M.; Matsumoto, T.; Xiong, Y., NEDD8 modification of CUL1 dissociates p120(CAND1), an inhibitor of CUL1-SKP1 binding and SCF ligases. *Mol Cell* **2002**, *10*, 1511-8.
70. Zheng, J.; Yang, X.; Harrell, J. M.; Ryzhikov, S.; Shim, E. H.; Lykke-Andersen, K.; Wei, N.; Sun, H.; Kobayashi, R.; Zhang, H., CAND1 binds to unneddylated CUL1 and regulates the formation of SCF ubiquitin E3 ligase complex. *Mol Cell* **2002**, *10*, 1519-26.
71. Jeon, Y. J.; Yoo, H. M.; Chung, C. H., ISG15 and immune diseases. *Biochim Biophys Acta* **2010**, *1802*, 485-96.
72. Okumura, A.; Pitha, P. M.; Harty, R. N., ISG15 inhibits Ebola VP40 VLP budding in an L-domain-dependent manner by blocking Nedd4 ligase activity. *Proc Natl Acad Sci U S A* **2008**, *105*, 3974-9.
73. Malakhova, O. A.; Zhang, D. E., ISG15 inhibits Nedd4 ubiquitin E3 activity and enhances the innate antiviral response. *J Biol Chem* **2008**, *283*, 8783-7.
74. Theng, S. S.; Wang, W.; Mah, W. C.; Chan, C.; Zhuo, J.; Gao, Y.; Qin, H.; Lim, L.; Chong, S. S.; Song, J.; Lee, C. G., Disruption of FAT10-MAD2 binding inhibits tumor progression. *Proc Natl Acad Sci U S A* **2014**, *111*, E5282-91.
75. Canaan, A.; DeFuria, J.; Perelman, E.; Schultz, V.; Seay, M.; Tuck, D.; Flavell, R. A.; Snyder, M. P.; Obin, M. S.; Weissman, S. M., Extended lifespan and reduced adiposity in mice lacking the FAT10 gene. *Proc Natl Acad Sci U S A* **2014**, *111*, 5313-8.
76. Watanabe, J.; Nakagawa, M.; Watanabe, N.; Nakamura, M., Ubiquitin-like protein MNSFbeta covalently binds to Bcl-G and enhances lipopolysaccharide/interferon gamma-induced apoptosis in macrophages. *FEBS J* **2013**, *280*, 1281-93.
77. Nakamura, M.; Nakagawa, M.; Watanabe, J., Ubiquitin-like protein MNSFbeta negatively regulates T cell function and survival. *Immunol Invest* **2015**, *44*, 1-12.
78. Nakamura, M.; Watanabe, N.; Notsu, K., Ubiquitin-like protein MNSFbeta covalently binds to cytosolic 10-formyltetrahydrofolate dehydrogenase and regulates thymocyte function. *Biochem Biophys Res Commun* **2015**, *464*, 1096-100.
79. Tatsumi, K.; Yamamoto-Mukai, H.; Shimizu, R.; Waguri, S.; Sou, Y. S.; Sakamoto, A.; Taya, C.; Shitara, H.; Hara, T.; Chung, C. H.; Tanaka, K.; Yamamoto, M.; Komatsu, M., The Ufm1-activating enzyme Uba5 is indispensable for erythroid differentiation in mice. *Nat Commun* **2011**, *2*, 181.
80. Xu, J.; Zhang, J.; Wang, L.; Zhou, J.; Huang, H.; Wu, J.; Zhong, Y.; Shi, Y., Solution structure of Urm1 and its implications for the origin of protein modifiers. *Proc Natl Acad Sci U S A* **2006**, *103*, 11625-30.
81. Groll, M.; Bajorek, M.; Kohler, A.; Moroder, L.; Rubin, D. M.; Huber, R.; Glickman, M. H.; Finley, D., A gated channel into the proteasome core particle. *Nat Struct Biol* **2000**, *7*, 1062-7.

82. McBride, A.; Ryan, P. Y., Proteasome inhibitors in the treatment of multiple myeloma. *Expert Rev Anticancer Ther* **2013**, *13*, 339-58.
83. Kane, R. C.; Bross, P. F.; Farrell, A. T.; Pazdur, R., Velcade: U.S. FDA approval for the treatment of multiple myeloma progressing on prior therapy. *Oncologist* **2003**, *8*, 508-13.
84. Crawford, L. J.; Walker, B.; Irvine, A. E., Proteasome inhibitors in cancer therapy. *J Cell Commun Signal* **2011**, *5*, 101-10.
85. Jentsch, S.; McGrath, J. P.; Varshavsky, A., The yeast DNA repair gene RAD6 encodes a ubiquitin-conjugating enzyme. *Nature* **1987**, *329*, 131-4.
86. Svejstrup, J. Q., Contending with transcriptional arrest during RNAPII transcript elongation. *Trends Biochem Sci* **2007**, *32*, 165-71.
87. Daulny, A.; Tansey, W. P., Damage control: DNA repair, transcription, and the ubiquitin-proteasome system. *DNA Repair* **2009**, *8*, 444-448.
88. van Cuijk, L.; Vermeulen, W.; Marteijn, J. A., Ubiquitin at work: the ubiquitous regulation of the damage recognition step of NER. *Exp Cell Res* **2014**, *329*, 101-9.
89. Fousteri, M.; Mullenders, L. H., Transcription-coupled nucleotide excision repair in mammalian cells: molecular mechanisms and biological effects. *Cell Res* **2008**, *18*, 73-84.
90. Soucy, T. A.; Smith, P. G.; Milhollen, M. A.; Berger, A. J.; Gavin, J. M.; Adhikari, S.; Brownell, J. E.; Burke, K. E.; Cardin, D. P.; Critchley, S.; Cullis, C. A.; Doucette, A.; Garnsey, J. J.; Gaulin, J. L.; Gershman, R. E.; Lublinsky, A. R.; McDonald, A.; Mizutani, H.; Narayanan, U.; Olhava, E. J.; Peluso, S.; Rezaei, M.; Sintchak, M. D.; Talreja, T.; Thomas, M. P.; Traore, T.; Vyskocil, S.; Weatherhead, G. S.; Yu, J.; Zhang, J.; Dick, L. R.; Claiborne, C. F.; Rolfe, M.; Bolen, J. B.; Langston, S. P., An inhibitor of NEDD8-activating enzyme as a new approach to treat cancer. *Nature* **2009**, *458*, 732-736.
91. Zhu, W.; Dutta, A., An ATR- and BRCA1-mediated Fanconi anemia pathway is required for activating the G2/M checkpoint and DNA damage repair upon rereplication. *Mol Cell Biol* **2006**, *26*, 4601-11.
92. Lin, J. J.; Milhollen, M. A.; Smith, P. G.; Narayanan, U.; Dutta, A., NEDD8-targeting drug MLN4924 elicits DNA rereplication by stabilizing Cdt1 in S phase, triggering checkpoint activation, apoptosis, and senescence in cancer cells. *Cancer Res* **2010**, *70*, 10310-20.
93. Guihard, S.; Ramolu, L.; Macabre, C.; Wasyluk, B.; Noel, G.; Abecassis, J.; Jung, A. C., The NEDD8 conjugation pathway regulates p53 transcriptional activity and head and neck cancer cell sensitivity to ionizing radiation. *Int J Oncol* **2012**, *41*, 1531-40.
94. Jazaeri, A. A.; Shibata, E.; Park, J.; Bryant, J. L.; Conaway, M. R.; Modesitt, S. C.; Smith, P. G.; Milhollen, M. A.; Berger, A. J.; Dutta, A., Overcoming platinum resistance in preclinical models of ovarian cancer using the neddylation inhibitor MLN4924. *Mol Cancer Ther* **2013**, *12*, 1958-67.
95. Kee, Y.; Huang, M.; Chang, S.; Moreau, L. A.; Park, E.; Smith, P. G.; D'Andrea, A. D., Inhibition of the Nedd8 system sensitizes cells to DNA interstrand cross-linking agents. *Mol Cancer Res* **2012**, *10*, 369-77.
96. Nawrocki, S. T.; Kelly, K. R.; Smith, P. G.; Espitia, C. M.; Possemato, A.; Beausoleil, S. A.; Milhollen, M.; Blakemore, S.; Thomas, M.; Berger, A.; Carew, J. S., Disrupting protein NEDDylation with MLN4924 is a novel strategy to target cisplatin resistance in ovarian cancer. *Clin Cancer Res* **2013**, *19*, 3577-90.
97. Ma, T.; Chen, Y.; Zhang, F.; Yang, C. Y.; Wang, S.; Yu, X., RNF111-dependent neddylation activates DNA damage-induced ubiquitination. *Mol Cell* **2013**, *49*, 897-907.
98. Brown, J. S.; Jackson, S. P., Ubiquitylation, neddylation and the DNA damage response. *Open Biol* **2015**, *5*, 150018.
99. Kolas, N. K.; Chapman, J. R.; Nakada, S.; Ylanko, J.; Chahwan, R.; Sweeney, F. D.; Panier, S.; Mendez, M.; Wildenhain, J.; Thomson, T. M.; Pelletier, L.; Jackson, S. P.; Durocher, D., Orchestration of the DNA-damage response by the RNF8 ubiquitin ligase. *Science* **2007**, *318*, 1637-40.
100. Huen, M. S.; Grant, R.; Manke, I.; Minn, K.; Yu, X.; Yaffe, M. B.; Chen, J., RNF8 transduces the DNA-damage signal via histone ubiquitylation and checkpoint protein assembly. *Cell* **2007**, *131*, 901-14.

101. Mailand, N.; Bekker-Jensen, S.; Faustrup, H.; Melander, F.; Bartek, J.; Lukas, C.; Lukas, J., RNF8 ubiquitylates histones at DNA double-strand breaks and promotes assembly of repair proteins. *Cell* **2007**, *131*, 887-900.
102. Bekker-Jensen, S.; Rendtlew Danielsen, J.; Fugger, K.; Gromova, I.; Nerstedt, A.; Lukas, C.; Bartek, J.; Lukas, J.; Mailand, N., HERC2 coordinates ubiquitin-dependent assembly of DNA repair factors on damaged chromosomes. *Nat Cell Biol* **2010**, *12*, 80-6; sup pp 1-12.
103. Mattioli, F.; Vissers, J. H.; van Dijk, W. J.; Ikpa, P.; Citterio, E.; Vermeulen, W.; Marteiijn, J. A.; Sixma, T. K., RNF168 ubiquitinates K13-15 on H2A/H2AX to drive DNA damage signaling. *Cell* **2012**, *150*, 1182-95.
104. Nakada, S.; Tai, I.; Panier, S.; Al-Hakim, A.; Iemura, S.; Juang, Y. C.; O'Donnell, L.; Kumakubo, A.; Munro, M.; Sicheri, F.; Gingras, A. C.; Natsume, T.; Suda, T.; Durocher, D., Non-canonical inhibition of DNA damage-dependent ubiquitination by OTUB1. *Nature* **2010**, *466*, 941-6.
105. Patel, K. J.; Joenje, H., Fanconi anemia and DNA replication repair. *DNA Repair (Amst)* **2007**, *6*, 885-90.
106. Helleday, T., The underlying mechanism for the PARP and BRCA synthetic lethality: clearing up the misunderstandings. *Mol Oncol* **2011**, *5*, 387-93.
107. Mateo, J.; Carreira, S.; Sandhu, S.; Miranda, S.; Mossop, H.; Perez-Lopez, R.; Nava Rodrigues, D.; Robinson, D.; Omlin, A.; Tunariu, N.; Boysen, G.; Porta, N.; Flohr, P.; Gillman, A.; Figueiredo, I.; Paulding, C.; Seed, G.; Jain, S.; Ralph, C.; Protheroe, A.; Hussain, S.; Jones, R.; Elliott, T.; McGovern, U.; Bianchini, D.; Goodall, J.; Zafeiriou, Z.; Williamson, C. T.; Ferraldeschi, R.; Riisnaes, R.; Ebbs, B.; Fowler, G.; Roda, D.; Yuan, W.; Wu, Y.-M.; Cao, X.; Brough, R.; Pemberton, H.; A'Hern, R.; Swain, A.; Kunju, L. P.; Eeles, R.; Attard, G.; Lord, C. J.; Ashworth, A.; Rubin, M. A.; Knudsen, K. E.; Feng, F. Y.; Chinnaiyan, A. M.; Hall, E.; de Bono, J. S., DNA-Repair Defects and Olaparib in Metastatic Prostate Cancer. *New England Journal of Medicine* **2015**, *373*, 1697-1708.
108. Mattern, M. R.; Wu, J.; Nicholson, B., Ubiquitin-based anticancer therapy: Carpet bombing with proteasome inhibitors vs surgical strikes with E1, E2, E3, or DUB inhibitors. *Biochim Biophys Acta* **2012**.
109. Tokunaga, F., Linear ubiquitination-mediated NF-kappaB regulation and its related disorders. *J Biochem* **2013**, *154*, 313-23.
110. Basak, S.; Shih, V. F.; Hoffmann, A., Generation and activation of multiple dimeric transcription factors within the NF-kappaB signaling system. *Mol Cell Biol* **2008**, *28*, 3139-50.
111. Pal, A.; Young, M. A.; Donato, N. J., Emerging potential of therapeutic targeting of ubiquitin-specific proteases in the treatment of cancer. *Cancer Res* **2014**, *74*, 4955-66.
112. Miyamoto, S., Nuclear initiated NF-kappaB signaling: NEMO and ATM take center stage. *Cell Res* **2011**, *21*, 116-30.
113. Mackay, C.; Carroll, E.; Ibrahim, A. F.; Garg, A.; Inman, G. J.; Hay, R. T.; Alpi, A. F., E3 ubiquitin ligase HOIP attenuates apoptotic cell death induced by cisplatin. *Cancer Res* **2014**, *74*, 2246-57.
114. Traenckner, E. B.; Wilk, S.; Baeuerle, P. A., A proteasome inhibitor prevents activation of NF-kappa B and stabilizes a newly phosphorylated form of I kappa B-alpha that is still bound to NF-kappa B. *EMBO J* **1994**, *13*, 5433-41.
115. Wilkins, M. R.; Pasquali, C.; Appel, R. D.; Ou, K.; Golaz, O.; Sanchez, J. C.; Yan, J. X.; Gooley, A. A.; Hughes, G.; Humphery-Smith, I.; Williams, K. L.; Hochstrasser, D. F., From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Biotechnology (N Y)* **1996**, *14*, 61-5.
116. Tyers, M.; Mann, M., From genomics to proteomics. *Nature* **2003**, *422*, 193-197.
117. Karas, M.; Hillenkamp, F., Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* **1988**, *60*, 2299-301.
118. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M., Electrospray ionization for mass spectrometry of large biomolecules. *Science* **1989**, *246*, 64-71.
119. Cody, R. B., Jr.; Amster, I. J.; McLafferty, F. W., Peptide mixture sequencing by tandem Fourier-transform mass spectrometry. *Proc Natl Acad Sci U S A* **1985**, *82*, 6367-70.



120. Olsen, J. V.; Macek, B.; Lange, O.; Makarov, A.; Horning, S.; Mann, M., Higher-energy C-trap dissociation for peptide modification analysis. *Nature methods* **2007**, *4*, 709-12.
121. Cox, J.; Mann, M., MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **2008**, *26*, 1367-72.
122. Szklarczyk, D.; Franceschini, A.; Wyder, S.; Forslund, K.; Heller, D.; Huerta-Cepas, J.; Simonovic, M.; Roth, A.; Santos, A.; Tsafou, K. P.; Kuhn, M.; Bork, P.; Jensen, L. J.; von Mering, C., STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Research* **2015**, *43*, D447-D452.
123. Washburn, M. P.; Wolters, D.; Yates, J. R., Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* **2001**, *19*, 242-247.
124. Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M. C.; Yates, J. R., 3rd, Protein analysis by shotgun/bottom-up proteomics. *Chem Rev* **2013**, *113*, 2343-94.
125. Yates, J. R., 3rd, The revolution and evolution of shotgun proteomics for large-scale proteome analysis. *J Am Chem Soc* **2013**, *135*, 1629-40.
126. Huber, C.; Huber, L., Special focus on top-down proteomics. *Proteomics* **2010**, *10*, 3564-5.
127. Siuti, N.; Kelleher, N. L., Decoding protein modifications using top-down mass spectrometry. *Nature methods* **2007**, *4*, 817-21.
128. Beaudette, P.; Popp, O.; Dittmar, G., Proteomic techniques to probe the ubiquitin landscape. *Proteomics* **2016**, *16*, 273-87.
129. Durbin, K. R.; Fornelli, L.; Fellers, R. T.; Doubleday, P. F.; Narita, M.; Kelleher, N. L., Quantitation and Identification of Thousands of Human Proteoforms below 30 kDa. *J Proteome Res* **2016**, *15*, 976-82.
130. Ong, S.-E.; Mann, M., A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). *Nat. Protocols* **2007**, *1*, 2650-2660.
131. Schwanhäusser, B.; Gossen, M.; Dittmar, G.; Selbach, M., Global analysis of cellular protein translation by pulsed SILAC. *PROTEOMICS* **2009**, *9*, 205-209.
132. Geiger, T.; Wisniewski, J. R.; Cox, J.; Zanivan, S.; Kruger, M.; Ishihama, Y.; Mann, M., Use of stable isotope labeling by amino acids in cell culture as a spike-in standard in quantitative proteomics. *Nat Protoc* **2011**, *6*, 147-57.
133. Schmidt, A.; Kellermann, J.; Lottspeich, F., A novel strategy for quantitative proteomics using isotope-coded protein labels. *Proteomics* **2005**, *5*, 4-15.
134. Boersema, P. J.; Raijmakers, R.; Lemeer, S.; Mohammed, S.; Heck, A. J., Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nat Protoc* **2009**, *4*, 484-94.
135. Thompson, A.; Schafer, J.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, G.; Neumann, T.; Johnstone, R.; Mohammed, A. K.; Hamon, C., Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem* **2003**, *75*, 1895-904.
136. Dayon, L.; Hainard, A.; Licker, V.; Turck, N.; Kuhn, K.; Hochstrasser, D. F.; Burkhard, P. R.; Sanchez, J. C., Relative quantification of proteins in human cerebrospinal fluids by MS/MS using 6-plex isobaric tags. *Anal Chem* **2008**, *80*, 2921-31.
137. Hebert, A. S.; Merrill, A. E.; Bailey, D. J.; Still, A. J.; Westphall, M. S.; Strieter, E. R.; Pagliarini, D. J.; Coon, J. J., Neutron-encoded mass signatures for multiplexed proteome quantification. *Nature methods* **2013**, *10*, 332-4.
138. Bantscheff, M.; Schirle, M.; Sweetman, G.; Rick, J.; Kuster, B., Quantitative mass spectrometry in proteomics: a critical review. *Analytical and Bioanalytical Chemistry* **2007**, *389*, 1017-1031.
139. Cox, J.; Hein, M. Y.; Luber, C. A.; Paron, I.; Nagaraj, N.; Mann, M., Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics* **2014**, *13*, 2513-26.
140. Gerber, S. A.; Rush, J.; Stemman, O.; Kirschner, M. W.; Gygi, S. P., Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc Natl Acad Sci U S A* **2003**, *100*, 6940-5.

141. Peng, J.; Schwartz, D.; Elias, J. E.; Thoreen, C. C.; Cheng, D.; Marsischky, G.; Roelofs, J.; Finley, D.; Gygi, S. P., A proteomics approach to understanding protein ubiquitination. *Nat Biotech* **2003**, *21*, 921-926.
142. Kirkpatrick, D. S.; Hathaway, N. A.; Hanna, J.; Elsasser, S.; Rush, J.; Finley, D.; King, R. W.; Gygi, S. P., Quantitative analysis of in vitro ubiquitinated cyclin B1 reveals complex chain topology. *Nat Cell Biol* **2006**, *8*, 700-710.
143. Phu, L.; Izrael-Tomasevic, A.; Matsumoto, M. L.; Bustos, D.; Dynek, J. N.; Fedorova, A. V.; Bakalarski, C. E.; Arnott, D.; Deshayes, K.; Dixit, V. M.; Kelley, R. F.; Vucic, D.; Kirkpatrick, D. S., Improved quantitative mass spectrometry methods for characterizing complex ubiquitin signals. *Mol Cell Proteomics* **2011**, *10*, M110.003756.
144. Matsumoto, M. L.; Wickliffe, K. E.; Dong, K. C.; Yu, C.; Bosanac, I.; Bustos, D.; Phu, L.; Kirkpatrick, D. S.; Hymowitz, S. G.; Rape, M.; Kelley, R. F.; Dixit, V. M., K11-Linked Polyubiquitination in Cell Cycle Control Revealed by a K11 Linkage-Specific Antibody. *Mol Cell* **2010**, *39*, 477-484.
145. Impens, F.; Radoshevich, L.; Cossart, P.; Ribet, D., Mapping of SUMO sites and analysis of SUMOylation changes induced by external stimuli. *Proc Natl Acad Sci U S A* **2014**, *111*, 12432-7.
146. Tammsalu, T.; Matic, I.; Jaffray, E. G.; Ibrahim, A. F.; Tatham, M. H.; Hay, R. T., Proteome-wide identification of SUMO2 modification sites. *Sci Signal* **2014**, *7*, rs2.
147. Beckett, D.; Kovaleva, E.; Schatz, P. J., A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Sci* **1999**, *8*, 921-9.
148. Franco, M.; Seyfried, N. T.; Brand, A. H.; Peng, J.; Mayor, U., A novel strategy to isolate ubiquitin conjugates reveals wide role for ubiquitination during neural development. *Mol Cell Proteomics* **2011**, *10*, M110 002188.
149. Hjerpe, R.; Aillet, F.; Lopitz-Otsoa, F.; Lang, V.; England, P.; Rodriguez, M. S., Efficient protection and isolation of ubiquitylated proteins using tandem ubiquitin-binding entities. *EMBO Rep* **2009**, *10*, 1250-8.
150. Raasi, S.; Pickart, C. M., Rad23 ubiquitin-associated domains (UBA) inhibit 26 S proteasome-catalyzed proteolysis by sequestering lysine 48-linked polyubiquitin chains. *J Biol Chem* **2003**, *278*, 8951-9.
151. Raasi, S.; Varadan, R.; Fushman, D.; Pickart, C. M., Diverse polyubiquitin interaction properties of ubiquitin-associated domains. *Nat Struct Mol Biol* **2005**, *12*, 708-14.
152. Yoshida, Y.; Saeki, Y.; Murakami, A.; Kawawaki, J.; Tsuchiya, H.; Yoshihara, H.; Shindo, M.; Tanaka, K., A comprehensive method for detecting ubiquitinated substrates using TR-TUBE. *Proc Natl Acad Sci U S A* **2015**, *112*, 4630-5.
153. Mark, K. G.; Loveless, T. B.; Toczyski, D. P., Isolation of ubiquitinated substrates by tandem affinity purification of E3 ligase-polyubiquitin-binding domain fusions (ligase traps). *Nat Protoc* **2016**, *11*, 291-301.
154. Lowe, E. D.; Hasan, N.; Trempe, J. F.; Fonso, L.; Noble, M. E.; Endicott, J. A.; Johnson, L. N.; Brown, N. R., Structures of the Dsk2 UBL and UBA domains and their complex. *Acta Crystallogr D Biol Crystallogr* **2006**, *62*, 177-88.
155. Seibenhener, M. L.; Babu, J. R.; Geetha, T.; Wong, H. C.; Krishna, N. R.; Wooten, M. W., Sequestosome 1/p62 is a polyubiquitin chain binding protein involved in ubiquitin proteasome degradation. *Mol Cell Biol* **2004**, *24*, 8055-68.
156. Layfield, R.; Tooth, D.; Landon, M.; Dawson, S.; Mayer, J.; Alban, A., Purification of polyubiquitinated proteins by S5a-affinity chromatography. *Proteomics* **2001**, *1*, 773-7.
157. Hershko, A.; Eytan, E.; Ciechanover, A.; Haas, A. L., Immunochemical analysis of the turnover of ubiquitin-protein conjugates in intact cells. Relationship to the breakdown of abnormal proteins. *J Biol Chem* **1982**, *257*, 13964-70.
158. Fujimuro, M.; Sawada, H.; Yokosawa, H., Production and characterization of monoclonal antibodies specific to multi-ubiquitin chains of polyubiquitinated proteins. *FEBS Lett* **1994**, *349*, 173-80.

159. Newton, K.; Matsumoto, M. L.; Wertz, I. E.; Kirkpatrick, D. S.; Lill, J. R.; Tan, J.; Dugger, D.; Gordon, N.; Sidhu, S. S.; Fellouse, F. A.; Komuves, L.; French, D. M.; Ferrando, R. E.; Lam, C.; Compaan, D.; Yu, C.; Bosanac, I.; Hymowitz, S. G.; Kelley, R. F.; Dixit, V. M., Ubiquitin chain editing revealed by polyubiquitin linkage-specific antibodies. *Cell* **2008**, *134*, 668-78.
160. Xu, G.; Paige, J. S.; Jaffrey, S. R., Global analysis of lysine ubiquitination by ubiquitin remnant immunoaffinity profiling. *Nat Biotechnol* **2010**, *28*, 868-73.
161. Kim, W.; Bennett, E. J.; Huttlin, E. L.; Guo, A.; Li, J.; Possemato, A.; Sowa, M. E.; Rad, R.; Rush, J.; Comb, M. J.; Harper, J. W.; Gygi, S. P., Systematic and quantitative assessment of the ubiquitin-modified proteome. *Mol Cell* **2011**, *44*, 325-40.
162. Wagner, S. A.; Beli, P.; Weinert, B. T.; Nielsen, M. L.; Cox, J.; Mann, M.; Choudhary, C., A proteome-wide, quantitative survey of in vivo ubiquitylation sites reveals widespread regulatory roles. *Mol Cell Proteomics* **2011**, *10*, M111.013284.
163. Emanuele, Michael J.; Elia, Andrew E. H.; Xu, Q.; Thoma, Claudio R.; Izhar, L.; Leng, Y.; Guo, A.; Chen, Y.-N.; Rush, J.; Hsu, Paul W.-C.; Yen, H.-Chi S.; Elledge, Stephen J., Global Identification of Modular Cullin-RING Ligase Substrates. *Cell* **2011**, *147*, 459-474.
164. Lindsten, K.; Menendez-Benito, V.; Masucci, M. G.; Dantuma, N. P., A transgenic mouse model of the ubiquitin/proteasome system. *Nat Biotechnol* **2003**, *21*, 897-902.
165. Dantuma, N. P.; Lindsten, K.; Glas, R.; Jellne, M.; Masucci, M. G., Short-lived green fluorescent proteins for quantifying ubiquitin/proteasome-dependent proteolysis in living cells. *Nat Biotechnol* **2000**, *18*, 538-43.
166. van Wijk, S. J.; Fiskin, E.; Putyrski, M.; Pampaloni, F.; Hou, J.; Wild, P.; Kensche, T.; Grecco, H. E.; Bastiaens, P.; Dikic, I., Fluorescence-based sensors to monitor localization and functions of linear and K63-linked ubiquitin chains in cells. *Mol Cell* **2012**, *47*, 797-809.
167. Sims, J. J.; Scavone, F.; Cooper, E. M.; Kane, L. A.; Youle, R. J.; Boeke, J. D.; Cohen, R. E., Polyubiquitin-sensor proteins reveal localization and linkage-type dependence of cellular ubiquitin signaling. *Nature methods* **2012**, *9*, 303-9.
168. Smith, A. M.; Ammar, R.; Nislow, C.; Giaever, G., A survey of yeast genomic assays for drug and target discovery. *Pharmacol Ther* **2010**, *127*, 156-64.
169. Isasa, M.; Rose, C. M.; Elsasser, S.; Navarrete-Perea, J.; Paulo, J. A.; Finley, D. J.; Gygi, S. P., Multiplexed, Proteome-Wide Protein Expression Profiling: Yeast Deubiquitylating Enzyme Knockout Strains. *Journal of Proteome Research* **2015**, *14*, 5306-5317.
170. Cong, L.; Ran, F. A.; Cox, D.; Lin, S.; Barretto, R.; Habib, N.; Hsu, P. D.; Wu, X.; Jiang, W.; Marraffini, L. A.; Zhang, F., Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **2013**, *339*, 819-823.
171. Lee, S. Y.; Pullen, L.; Virgil, D. J.; Castañeda, C. A.; Abeykoon, D.; Bolon, D. N. A.; Fushman, D., Alanine Scan of Core Positions in Ubiquitin Reveals Links between Dynamics, Stability, and Function. *Journal of Molecular Biology* **2014**, *426*, 1377-1389.
172. Pitluk, Z. W.; McDonough, M.; Sangan, P.; Gonda, D. K., Novel CDC34 (UBC3) ubiquitin-conjugating enzyme mutants obtained by charge-to-alanine scanning mutagenesis. *Mol Cell Biol* **1995**, *15*, 1210-9.
173. Heal, W. P.; Wickramasinghe, S. R.; Tate, E. W., Activity based chemical proteomics: profiling proteases as drug targets. *Curr Drug Discov Technol* **2008**, *5*, 200-12.
174. Cravatt, B. F.; Sorensen, E. J., Chemical strategies for the global analysis of protein function. *Curr Opin Chem Biol* **2000**, *4*, 663-8.
175. Jeffery, D. A.; Bogoy, M., Chemical proteomics and its application to drug discovery. *Curr Opin Biotechnol* **2003**, *14*, 87-95.
176. Weerapana, E.; Simon, G. M.; Cravatt, B. F., Disparate proteome reactivity profiles of carbon electrophiles. *Nat Chem Biol* **2008**, *4*, 405-7.
177. Borodovsky, A.; Kessler, B. M.; Casagrande, R.; Overkleeft, H. S.; Wilkinson, K. D.; Ploegh, H. L., A novel active site-directed probe specific for deubiquitylating enzymes reveals proteasome association of USP14. *EMBO J* **2001**, *20*, 5187-96.

178. Borodovsky, A.; Ovaa, H.; Kolli, N.; Gan-Erdene, T.; Wilkinson, K. D.; Ploegh, H. L.; Kessler, B. M., Chemistry-based functional proteomics reveals novel members of the deubiquitinating enzyme family. *Chem Biol* **2002**, *9*, 1149-59.
179. Altun, M.; Kramer, H. B.; Willems, L. I.; McDermott, J. L.; Leach, C. A.; Goldenberg, S. J.; Kumar, K. G.; Konietzny, R.; Fischer, R.; Kogan, E.; Mackeen, M. M.; McGouran, J.; Khoronenkova, S. V.; Parsons, J. L.; Dianov, G. L.; Nicholson, B.; Kessler, B. M., Activity-based chemical proteomics accelerates inhibitor development for deubiquitylating enzymes. *Chem Biol* **2011**, *18*, 1401-12.
180. Kramer, H. B.; Nicholson, B.; Kessler, B. M.; Altun, M., Detection of ubiquitin–proteasome enzymatic activities in cells: Application of activity-based probes to inhibitor development. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **2012**, *1823*, 2029-2037.
181. Mulder, M. P. C.; El Oualid, F.; ter Beek, J.; Ovaa, H., A Native Chemical Ligation Handle that Enables the Synthesis of Advanced Activity-Based Probes: Diubiquitin as a Case Study. *Chembiochem* **2014**, *15*, 946-949.
182. Iphofer, A.; Kummer, A.; Nimtz, M.; Ritter, A.; Arnold, T.; Frank, R.; van den Heuvel, J.; Kessler, B. M.; Jansch, L.; Franke, R., Profiling ubiquitin linkage specificities of deubiquitinating enzymes with branched ubiquitin isopeptide probes. *Chembiochem* **2012**, *13*, 1416-20.
183. Love, K. R.; Pandya, R. K.; Spooner, E.; Ploegh, H. L., Ubiquitin C-Terminal Electrophiles Are Activity-Based Probes for Identification and Mechanistic Study of Ubiquitin Conjugating Machinery. *ACS Chem Biol* **2009**, *4*, 275-287.
184. Pao, K. C.; Stanley, M.; Han, C.; Lai, Y. C.; Murphy, P.; Balk, K.; Wood, N. T.; Corti, O.; Corvol, J. C.; Muqit, M. M.; Virdee, S., Probes of ubiquitin E3 ligases enable systematic dissection of parkin activation. *Nat Chem Biol* **2016**, *12*, 324-31.
185. de Jong, A.; Merckx, R.; Berlin, I.; Rodenko, B.; Wijdeven, R. H.; El Atmioui, D.; Yalcin, Z.; Robson, C. N.; Neefjes, J. J.; Ovaa, H., Ubiquitin-based probes prepared by total synthesis to profile the activity of deubiquitinating enzymes. *Chembiochem* **2012**, *13*, 2251-8.
186. van Swieten, P. F.; Leeuwenburgh, M. A.; Kessler, B. M.; Overkleeft, H. S., Bioorthogonal organic chemistry in living cells: novel strategies for labeling biomolecules. *Organic & Biomolecular Chemistry* **2005**, *3*, 20-27.
187. Martell, J.; Weerapana, E., Applications of copper-catalyzed click chemistry in activity-based protein profiling. *Molecules* **2014**, *19*, 1378-93.
188. 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 Chem Int Ed Engl* **2002**, *41*, 2596-2599.
189. Heal, W. P.; Jovanovic, B.; Bessin, S.; Wright, M. H.; Magee, A. I.; Tate, E. W., Bioorthogonal chemical tagging of protein cholesterylation in living cells. *Chem Commun (Camb)* **2011**, *47*, 4081-3.
190. Broncel, M.; Serwa, R. A.; Ciepla, P.; Krause, E.; Dallman, M. J.; Magee, A. I.; Tate, E. W., Multifunctional reagents for quantitative proteome-wide analysis of protein modification in human cells and dynamic profiling of protein lipidation during vertebrate development. *Angew Chem Int Ed Engl* **2015**, *54*, 5948-51.
191. McGouran, J. F.; Kramer, H. B.; Mackeen, M. M.; di Gleria, K.; Altun, M.; Kessler, B. M., Fluorescence-based active site probes for profiling deubiquitinating enzymes. *Org Biomol Chem* **2012**, *10*, 3379-83.
192. An, H.; Statsyuk, A. V., Development of activity-based probes for ubiquitin and ubiquitin-like protein signaling pathways. *J Am Chem Soc* **2013**, *135*, 16948-62.
193. Weerapana, E.; Wang, C.; Simon, G. M.; Richter, F.; Khare, S.; Dillon, M. B. D.; Bachovchin, D. A.; Mowen, K.; Baker, D.; Cravatt, B. F., Quantitative reactivity profiling predicts functional cysteines in proteomes. *Nature* **2010**, *468*, 790-795.
194. Wang, C.; Weerapana, E.; Blewett, M. M.; Cravatt, B. F., A chemoproteomic platform to quantitatively map targets of lipid-derived electrophiles. *Nat Meth* **2014**, *11*, 79-85.
195. Abo, M.; Weerapana, E., A Caged Electrophilic Probe for Global Analysis of Cysteine Reactivity in Living Cells. *J Am Chem Soc* **2015**, *137*, 7087-90.

196. Chan, E. W.; Chattopadhyaya, S.; Panicker, R. C.; Huang, X.; Yao, S. Q., Developing photoactive affinity probes for proteomic profiling: hydroxamate-based probes for metalloproteases. *J Am Chem Soc* **2004**, *126*, 14435-46.
197. Lomenick, B.; Olsen, R. W.; Huang, J., Identification of direct protein targets of small molecules. *ACS Chem Biol* **2011**, *6*, 34-46.
198. Martinez Molina, D.; Jafari, R.; Ignatushchenko, M.; Seki, T.; Larsson, E. A.; Dan, C.; Sreekumar, L.; Cao, Y.; Nordlund, P., Monitoring drug target engagement in cells and tissues using the cellular thermal shift assay. *Science* **2013**, *341*, 84-7.
199. Roux, K. J.; Kim, D. I.; Raida, M.; Burke, B., A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J Cell Biol* **2012**, *196*, 801-10.
200. Coyaud, E.; Mis, M.; Laurent, E. M.; Dunham, W. H.; Couzens, A. L.; Robitaille, M.; Gingras, A. C.; Angers, S.; Raught, B., BioID-based Identification of Skp Cullin F-box (SCF)beta-TrCP1/2 E3 Ligase Substrates. *Mol Cell Proteomics* **2015**, *14*, 1781-95.
201. Sakamoto, K. M.; Kim, K. B.; Kumagai, A.; Mercurio, F.; Crews, C. M.; Deshaies, R. J., Protacs: chimeric molecules that target proteins to the Skp1-Cullin-F box complex for ubiquitination and degradation. *Proc Natl Acad Sci U S A* **2001**, *98*, 8554-9.
202. Ito, T.; Ando, H.; Suzuki, T.; Ogura, T.; Hotta, K.; Imamura, Y.; Yamaguchi, Y.; Handa, H., Identification of a primary target of thalidomide teratogenicity. *Science* **2010**, *327*, 1345-50.
203. Lu, J.; Qian, Y.; Altieri, M.; Dong, H.; Wang, J.; Raina, K.; Hines, J.; Winkler, J. D.; Crew, A. P.; Coleman, K.; Crews, C. M., Hijacking the E3 Ubiquitin Ligase Cereblon to Efficiently Target BRD4. *Chem Biol* **2015**, *22*, 755-63.
204. Winter, G. E.; Buckley, D. L.; Paulk, J.; Roberts, J. M.; Souza, A.; Dhe-Paganon, S.; Bradner, J. E., DRUG DEVELOPMENT. Phthalimide conjugation as a strategy for in vivo target protein degradation. *Science* **2015**, *348*, 1376-81.
205. Lai, A. C.; Toure, M.; Hellerschmied, D.; Salami, J.; Jaime-Figueroa, S.; Ko, E.; Hines, J.; Crews, C. M., Modular PROTAC Design for the Degradation of Oncogenic BCR-ABL. *Angew Chem Int Ed Engl* **2016**, *55*, 807-10.
206. Yang, Y.; Kitagaki, J.; Dai, R. M.; Tsai, Y. C.; Lorick, K. L.; Ludwig, R. L.; Pierre, S. A.; Jensen, J. P.; Davydov, I. V.; Oberoi, P.; Li, C. C.; Kenten, J. H.; Beutler, J. A.; Vousden, K. H.; Weissman, A. M., Inhibitors of ubiquitin-activating enzyme (E1), a new class of potential cancer therapeutics. *Cancer Res* **2007**, *67*, 9472-81.
207. Ungermannova, D.; Parker, S. J.; Nasveschuk, C. G.; Chapnick, D. A.; Phillips, A. J.; Kuchta, R. D.; Liu, X., Identification and mechanistic studies of a novel ubiquitin E1 inhibitor. *J Biomol Screen* **2012**, *17*, 421-34.
208. Brownell, J. E.; Sintchak, M. D.; Gavin, J. M.; Liao, H.; Bruzzese, F. J.; Bump, N. J.; Soucy, T. A.; Milhollen, M. A.; Yang, X.; Burkhardt, A. L.; Ma, J.; Loke, H. K.; Lingaraj, T.; Wu, D.; Hamman, K. B.; Spelman, J. J.; Cullis, C. A.; Langston, S. P.; Vyskocil, S.; Sells, T. B.; Mallender, W. D.; Visiers, I.; Li, P.; Claiborne, C. F.; Rolfe, M.; Bolen, J. B.; Dick, L. R., Substrate-assisted inhibition of ubiquitin-like protein-activating enzymes: the NEDD8 E1 inhibitor MLN4924 forms a NEDD8-AMP mimetic in situ. *Mol Cell* **2010**, *37*, 102-11.
209. Ceccarelli, D. F.; Tang, X.; Pelletier, B.; Orlicky, S.; Xie, W.; Plantevin, V.; Neculai, D.; Chou, Y. C.; Ogunjimi, A.; Al-Hakim, A.; Varelas, X.; Koszela, J.; Wasney, G. A.; Vedadi, M.; Dhe-Paganon, S.; Cox, S.; Xu, S.; Lopez-Girona, A.; Mercurio, F.; Wrana, J.; Durocher, D.; Meloche, S.; Webb, D. R.; Tyers, M.; Sicheri, F., An allosteric inhibitor of the human Cdc34 ubiquitin-conjugating enzyme. *Cell* **2011**, *145*, 1075-87.
210. Huang, H.; Ceccarelli, D. F.; Orlicky, S.; St-Cyr, D. J.; Ziemba, A.; Garg, P.; Plamondon, S.; Auer, M.; Sidhu, S.; Marinier, A.; Kleiger, G.; Tyers, M.; Sicheri, F., E2 enzyme inhibition by stabilization of a low-affinity interface with ubiquitin. *Nat Chem Biol* **2014**, *10*, 156-63.
211. Sanders, M. A.; Brahemi, G.; Nangia-Makker, P.; Balan, V.; Morelli, M.; Kothayer, H.; Westwell, A. D.; Shekhar, M. P., Novel inhibitors of Rad6 ubiquitin conjugating enzyme: design, synthesis, identification, and functional characterization. *Mol Cancer Ther* **2013**, *12*, 373-83.

212. Cheng, J.; Fan, Y. H.; Xu, X.; Zhang, H.; Dou, J.; Tang, Y.; Zhong, X.; Rojas, Y.; Yu, Y.; Zhao, Y.; Vasudevan, S. A.; Zhang, H.; Nuchtern, J. G.; Kim, E. S.; Chen, X.; Lu, F.; Yang, J., A small-molecule inhibitor of UBE2N induces neuroblastoma cell death via activation of p53 and JNK pathways. *Cell Death Dis* **2014**, *5*, 1079.
213. Pulvino, M.; Liang, Y.; Oleksyn, D.; DeRan, M.; Van Pelt, E.; Shapiro, J.; Sanz, I.; Chen, L.; Zhao, J., Inhibition of proliferation and survival of diffuse large B-cell lymphoma cells by a small-molecule inhibitor of the ubiquitin-conjugating enzyme Ubc13-Uev1A. *Blood* **2012**, *120*, 1668-77.
214. Strickson, S.; Campbell, D. G.; Emmerich, C. H.; Knebel, A.; Plater, L.; Ritorto, M. S.; Shpiro, N.; Cohen, P., The anti-inflammatory drug BAY 11-7082 suppresses the MyD88-dependent signalling network by targeting the ubiquitin system. *Biochem J* **2013**, *451*, 427-37.
215. Tovar, C.; Rosinski, J.; Filipovic, Z.; Higgins, B.; Kolinsky, K.; Hilton, H.; Zhao, X.; Vu, B. T.; Qing, W.; Packman, K.; Myklebost, O.; Heimbrook, D. C.; Vassilev, L. T., Small-molecule MDM2 antagonists reveal aberrant p53 signaling in cancer: implications for therapy. *Proc Natl Acad Sci U S A* **2006**, *103*, 1888-93.
216. Shangary, S.; Qin, D.; McEachern, D.; Liu, M.; Miller, R. S.; Qiu, S.; Nikolovska-Coleska, Z.; Ding, K.; Wang, G.; Chen, J.; Bernard, D.; Zhang, J.; Lu, Y.; Gu, Q.; Shah, R. B.; Pienta, K. J.; Ling, X.; Kang, S.; Guo, M.; Sun, Y.; Yang, D.; Wang, S., Temporal activation of p53 by a specific MDM2 inhibitor is selectively toxic to tumors and leads to complete tumor growth inhibition. *Proc Natl Acad Sci U S A* **2008**, *105*, 3933-8.
217. Kojima, K.; Burks, J. K.; Arts, J.; Andreeff, M., The novel tryptamine derivative JNJ-26854165 induces wild-type p53- and E2F1-mediated apoptosis in acute myeloid and lymphoid leukemias. *Mol Cancer Ther* **2010**, *9*, 2545-57.
218. Wang, W.; Qin, J. J.; Voruganti, S.; Srivenugopal, K. S.; Nag, S.; Patil, S.; Sharma, H.; Wang, M. H.; Wang, H.; Buolamwini, J. K.; Zhang, R., The pyrido[b]indole MDM2 inhibitor SP-141 exerts potent therapeutic effects in breast cancer models. *Nat Commun* **2014**, *5*, 5086.
219. Lu, Y.; Nikolovska-Coleska, Z.; Fang, X.; Gao, W.; Shangary, S.; Qiu, S.; Qin, D.; Wang, S., Discovery of a nanomolar inhibitor of the human murine double minute 2 (MDM2)-p53 interaction through an integrated, virtual database screening strategy. *J Med Chem* **2006**, *49*, 3759-62.
220. Issaeva, N.; Bozko, P.; Enge, M.; Protopopova, M.; Verhoef, L. G.; Masucci, M.; Pramanik, A.; Selivanova, G., Small molecule RITA binds to p53, blocks p53-HDM-2 interaction and activates p53 function in tumors. *Nat Med* **2004**, *10*, 1321-8.
221. Nakajima, H.; Fujiwara, H.; Furuichi, Y.; Tanaka, K.; Shimbara, N., A novel small-molecule inhibitor of NF-kappaB signaling. *Biochem Biophys Res Commun* **2008**, *368*, 1007-13.
222. Chan, C. H.; Morrow, J. K.; Li, C. F.; Gao, Y.; Jin, G.; Moten, A.; Stagg, L. J.; Ladbury, J. E.; Cai, Z.; Xu, D.; Logothetis, C. J.; Hung, M. C.; Zhang, S.; Lin, H. K., Pharmacological inactivation of Skp2 SCF ubiquitin ligase restricts cancer stem cell traits and cancer progression. *Cell* **2013**, *154*, 556-68.
223. Fischer, E. S.; Bohm, K.; Lydeard, J. R.; Yang, H.; Stadler, M. B.; Cavadini, S.; Nagel, J.; Serluca, F.; Acker, V.; Lingaraju, G. M.; Tichkule, R. B.; Schebesta, M.; Forrester, W. C.; Schirle, M.; Hassiepen, U.; Ottl, J.; Hild, M.; Beckwith, R. E.; Harper, J. W.; Jenkins, J. L.; Thoma, N. H., Structure of the DDB1-CRBN E3 ubiquitin ligase in complex with thalidomide. *Nature* **2014**, *512*, 49-53.
224. Cai, Q.; Sun, H.; Peng, Y.; Lu, J.; Nikolovska-Coleska, Z.; McEachern, D.; Liu, L.; Qiu, S.; Yang, C. Y.; Miller, R.; Yi, H.; Zhang, T.; Sun, D.; Kang, S.; Guo, M.; Leopold, L.; Yang, D.; Wang, S., A potent and orally active antagonist (SM-406/AT-406) of multiple inhibitor of apoptosis proteins (IAPs) in clinical development for cancer treatment. *J Med Chem* **2011**, *54*, 2714-26.
225. Flygare, J. A.; Beresini, M.; Budha, N.; Chan, H.; Chan, I. T.; Cheeti, S.; Cohen, F.; Deshayes, K.; Doerner, K.; Eckhardt, S. G.; Elliott, L. O.; Feng, B.; Franklin, M. C.; Reisner, S. F.; Gazzard, L.; Halladay, J.; Hymowitz, S. G.; La, H.; LoRusso, P.; Maurer, B.; Murray, L.; Plise, E.; Quan, C.; Stephan, J. P.; Young, S. G.; Tom, J.; Tsui, V.; Um, J.; Varfolomeev, E.; Vucic, D.; Wagner, A. J.; Wallweber, H. J.; Wang, L.; Ware, J.; Wen, Z.; Wong, H.; Wong, J. M.; Wong, M.; Wong, S.; Yu, R.; Zobel, K.; Fairbrother, W. J., Discovery of a potent small-molecule antagonist of inhibitor of apoptosis (IAP)

- proteins and clinical candidate for the treatment of cancer (GDC-0152). *J Med Chem* **2012**, *55*, 4101-13.
226. Zeng, X.; Sigoillot, F.; Gaur, S.; Choi, S.; Pfaff, K. L.; Oh, D. C.; Hathaway, N.; Dimova, N.; Cuny, G. D.; King, R. W., Pharmacologic inhibition of the anaphase-promoting complex induces a spindle checkpoint-dependent mitotic arrest in the absence of spindle damage. *Cancer Cell* **2010**, *18*, 382-95.
227. Mund, T.; Lewis, M. J.; Maslen, S.; Pelham, H. R., Peptide and small molecule inhibitors of HECT-type ubiquitin ligases. *Proc Natl Acad Sci U S A* **2014**, *111*, 16736-41.
228. Cao, Y.; Wang, C.; Zhang, X.; Xing, G.; Lu, K.; Gu, Y.; He, F.; Zhang, L., Selective small molecule compounds increase BMP-2 responsiveness by inhibiting Smurf1-mediated Smad1/5 degradation. *Sci Rep* **2014**, *4*, 4965.
229. Edelmann, M. J.; Nicholson, B.; Kessler, B. M., Pharmacological targets in the ubiquitin system offer new ways of treating cancer, neurodegenerative disorders and infectious diseases. *Expert Rev Mol Med* **2011**, *13*, 35.
230. Kapuria, V.; Peterson, L. F.; Fang, D.; Bornmann, W. G.; Talpaz, M.; Donato, N. J., Deubiquitinase inhibition by small-molecule WP1130 triggers aggresome formation and tumor cell apoptosis. *Cancer Res* **2010**, *70*, 9265-76.
231. Lee, B. H.; Lee, M. J.; Park, S.; Oh, D. C.; Elsasser, S.; Chen, P. C.; Gartner, C.; Dimova, N.; Hanna, J.; Gygi, S. P.; Wilson, S. M.; King, R. W.; Finley, D., Enhancement of proteasome activity by a small-molecule inhibitor of USP14. *Nature* **2010**, *467*, 179-84.
232. Mullally, J. E.; Moos, P. J.; Edes, K.; Fitzpatrick, F. A., Cyclopentenone Prostaglandins of the J Series Inhibit the Ubiquitin Isopeptidase Activity of the Proteasome Pathway. *Journal of Biological Chemistry* **2001**, *276*, 30366-30373.
233. D'Arcy, P.; Brnjic, S.; Olofsson, M. H.; Fryknas, M.; Lindsten, K.; De Cesare, M.; Perego, P.; Sadeghi, B.; Hassan, M.; Larsson, R.; Linder, S., Inhibition of proteasome deubiquitinating activity as a new cancer therapy. *Nat Med* **2011**, *17*, 1636-40.
234. Landre, V.; Rotblat, B.; Melino, S.; Bernassola, F.; Melino, G., Screening for E3-ubiquitin ligase inhibitors: challenges and opportunities. *Oncotarget* **2014**, *5*, 7988-8013.
235. D'Arcy, P.; Wang, X.; Linder, S., Deubiquitinase inhibition as a cancer therapeutic strategy. *Pharmacol Ther* **2015**, *147*, 32-54.
236. Claessen, J. H.; Witte, M. D.; Yoder, N. C.; Zhu, A. Y.; Spooner, E.; Ploegh, H. L., Catch-and-release probes applied to semi-intact cells reveal ubiquitin-specific protease expression in Chlamydia trachomatis infection. *ChemBiochem* **2013**, *14*, 343-52.
237. Borodovsky, A.; Ovaa, H.; Meester, W. J.; Venanzi, E. S.; Bogyo, M. S.; Hekking, B. G.; Ploegh, H. L.; Kessler, B. M.; Overkleeft, H. S., Small-molecule inhibitors and probes for ubiquitin- and ubiquitin-like-specific proteases. *ChemBiochem* **2005**, *6*, 287-91.
238. Liu, S.; Hanzlik, R. P., Structure-activity relationships for inhibition of papain by peptide Michael acceptors. *J Med Chem* **1992**, *35*, 1067-75.
239. Burlina, F.; Morris, C.; Behrendt, R.; White, P.; Offer, J., Simplifying native chemical ligation with an N-acylsulfonamide linker. *Chem Commun (Camb)* **2012**, *48*, 2579-81.
240. Tanida, I.; Tanida-Miyake, E.; Komatsu, M.; Ueno, T.; Kominami, E., Human Apg3p/Aut1p homologue is an authentic E2 enzyme for multiple substrates, GATE-16, GABARAP, and MAP-LC3, and facilitates the conjugation of hApg12p to hApg5p. *J Biol Chem* **2002**, *277*, 13739-44.
241. Radoshevich, L.; Murrow, L.; Chen, N.; Fernandez, E.; Roy, S.; Fung, C.; Debnath, J., ATG12 conjugation to ATG3 regulates mitochondrial homeostasis and cell death. *Cell* **2010**, *142*, 590-600.
242. Koulich, E.; Li, X.; DeMartino, G. N., Relative structural and functional roles of multiple deubiquitylating proteins associated with mammalian 26S proteasome. *Mol Biol Cell* **2008**, *19*, 1072-82.
243. He, W.; Dorn, D. C.; Erdjument-Bromage, H.; Tempst, P.; Moore, M. A.; Massague, J., Hematopoiesis controlled by distinct TIF1gamma and Smad4 branches of the TGFbeta pathway. *Cell* **2006**, *125*, 929-41.

244. Ballinger, C. A.; Connell, P.; Wu, Y.; Hu, Z.; Thompson, L. J.; Yin, L. Y.; Patterson, C., Identification of CHIP, a novel tetratricopeptide repeat-containing protein that interacts with heat shock proteins and negatively regulates chaperone functions. *Mol Cell Biol* **1999**, *19*, 4535-45.
245. Chen, Z.; Barbi, J.; Bu, S.; Yang, H. Y.; Li, Z.; Gao, Y.; Jinasena, D.; Fu, J.; Lin, F.; Chen, C.; Zhang, J.; Yu, N.; Li, X.; Shan, Z.; Nie, J.; Gao, Z.; Tian, H.; Li, Y.; Yao, Z.; Zheng, Y.; Park, B. V.; Pan, Z.; Zhang, J.; Dang, E.; Li, Z.; Wang, H.; Luo, W.; Li, L.; Semenza, G. L.; Zheng, S. G.; Loser, K.; Tsun, A.; Greene, M. I.; Pardoll, D. M.; Pan, F.; Li, B., The ubiquitin ligase Stub1 negatively modulates regulatory T cell suppressive activity by promoting degradation of the transcription factor Foxp3. *Immunity* **2013**, *39*, 272-85.
246. Gadad, S. S.; Shandilya, J.; Kishore, A. H.; Kundu, T. K., NPM3, a Member of the Nucleophosmin/Nucleoplasmin Family, Enhances Activator-Dependent Transcription. *Biochemistry* **2010**, *49*, 1355-1357.
247. Li, X. B.; Chen, J.; Deng, M. J.; Wang, F.; Du, Z. W.; Zhang, J. W., Zinc finger protein HZF1 promotes K562 cell proliferation by interacting with and inhibiting INCA1. *Mol Med Rep* **2011**, *4*, 1131-7.
248. Tochio, N.; Umehara, T.; Munemasa, Y.; Suzuki, T.; Sato, S.; Tsuda, K.; Koshihara, S.; Kigawa, T.; Nagai, R.; Yokoyama, S., Solution structure of histone chaperone ANP32B: interaction with core histones H3-H4 through its acidic concave domain. *J Mol Biol* **2010**, *401*, 97-114.
249. Santa-Coloma, T. A., Anp32e (Cpd1) and related protein phosphatase 2 inhibitors. *Cerebellum* **2003**, *2*, 310-20.
250. Winkler, D. D.; Luger, K., The histone chaperone FACT: structural insights and mechanisms for nucleosome reorganization. *J Biol Chem* **2011**, *286*, 18369-74.
251. Oliveira, D. V.; Kato, A.; Nakamura, K.; Ikura, T.; Okada, M.; Kobayashi, J.; Yanagihara, H.; Saito, Y.; Tauchi, H.; Komatsu, K., Histone chaperone FACT regulates homologous recombination by chromatin remodeling through interaction with RNF20. *J Cell Sci* **2014**, *127*, 763-72.
252. British National Formulary. In *BNF*; Association, B. M.; Society, R. P., Eds. March 2011.
253. Khongkow, P.; Karunaratna, U.; Khongkow, M.; Gong, C.; Gomes, A. R.; Yague, E.; Monteiro, L. J.; Kongsema, M.; Zona, S.; Man, E. P.; Tsang, J. W.; Coombes, R. C.; Wu, K. J.; Khoo, U. S.; Medema, R. H.; Freire, R.; Lam, E. W., FOXM1 targets NBS1 to regulate DNA damage-induced senescence and epirubicin resistance. *Oncogene* **2014**, *33*, 4144-55.
254. Kuo, L. J.; Yang, L. X., Gamma-H2AX - a novel biomarker for DNA double-strand breaks. *In Vivo* **2008**, *22*, 305-9.
255. von Stechow, L.; Typas, D.; Carreras Puigvert, J.; Oort, L.; Siddappa, R.; Pines, A.; Vrieling, H.; van de Water, B.; Mullenders, L. H.; Danen, E. H., The E3 ubiquitin ligase ARIH1 protects against genotoxic stress by initiating a 4EHP-mediated mRNA translation arrest. *Mol Cell Biol* **2015**, *35*, 1254-68.
256. Wu, B.; Piloto, S.; Zeng, W.; Hoverter, N. P.; Schilling, T. F.; Waterman, M. L., Ring Finger Protein 14 is a new regulator of TCF/beta-catenin-mediated transcription and colon cancer cell survival. *EMBO Rep* **2013**, *14*, 347-55.
257. Pestov, D. G.; Strezoska, Z.; Lau, L. F., Evidence of p53-dependent cross-talk between ribosome biogenesis and the cell cycle: effects of nucleolar protein Bop1 on G(1)/S transition. *Mol Cell Biol* **2001**, *21*, 4246-55.
258. Strezoska, Z.; Pestov, D. G.; Lau, L. F., Functional inactivation of the mouse nucleolar protein Bop1 inhibits multiple steps in pre-rRNA processing and blocks cell cycle progression. *J Biol Chem* **2002**, *277*, 29617-25.
259. Pryor, A.; Tung, L.; Yang, Z.; Kapadia, F.; Chang, T. H.; Johnson, L. F., Growth-regulated expression and G0-specific turnover of the mRNA that encodes URH49, a mammalian DExH/D box protein that is highly related to the mRNA export protein UAP56. *Nucleic Acids Res* **2004**, *32*, 1857-65.



260. Cunningham, C. W.; Mukhopadhyay, A.; Lushington, G. H.; Blagg, B. S.; Prisinzano, T. E.; Krise, J. P., Uptake, distribution and diffusivity of reactive fluorophores in cells: implications toward target identification. *Mol Pharm* **2010**, *7*, 1301-10.
261. Fischer, R.; Waizenegger, T.; Kohler, K.; Brock, R., A quantitative validation of fluorophore-labelled cell-permeable peptide conjugates: fluorophore and cargo dependence of import. *Biochim Biophys Acta* **2002**, *1564*, 365-74.
262. Khan, S.; Sur, S.; Dankers, P. Y.; da Silva, R. M.; Boekhoven, J.; Poor, T. A.; Stupp, S. I., Post-assembly functionalization of supramolecular nanostructures with bioactive peptides and fluorescent proteins by native chemical ligation. *Bioconjug Chem* **2014**, *25*, 707-17.
263. Gait, M. J. Patent (WO 2013/030569 A2). 07/03/2013, 2013.
264. Lange, A.; Mills, R. E.; Lange, C. J.; Stewart, M.; Devine, S. E.; Corbett, A. H., Classical nuclear localization signals: definition, function, and interaction with importin alpha. *J Biol Chem* **2007**, *282*, 5101-5.
265. Chen, J. J.; Tsu, C. A.; Gavin, J. M.; Milhollen, M. A.; Bruzzese, F. J.; Mallender, W. D.; Sintchak, M. D.; Bump, N. J.; Yang, X.; Ma, J.; Loke, H.-K.; Xu, Q.; Li, P.; Bence, N. F.; Brownell, J. E.; Dick, L. R., Mechanistic Studies of Substrate-assisted Inhibition of Ubiquitin-activating Enzyme by Adenosine Sulfamate Analogues. *Journal of Biological Chemistry* **286**, 40867-40877.
266. Kapuria, V.; Peterson, L. F.; Showalter, H. D.; Kirchhoff, P. D.; Talpaz, M.; Donato, N. J., Protein cross-linking as a novel mechanism of action of a ubiquitin-activating enzyme inhibitor with anti-tumor activity. *Biochem Pharmacol* **2011**, *82*, 341-9.
267. Houille, O.; Fretz, H.; Hilpert, K.; Riederer, M.; Giller, T.; Valdenaire, O., Pyrazolidinedione derivatives and their use as platelet aggregation inhibitors. WO 2005002574 A12005.
268. Cocquet, G.; Ferroud, C.; Guy, A., A mild and efficient procedure for ring-opening reactions of piperidine and pyrrolidine derivatives by single electron transfer photooxidation. *Tetrahedron* **2000**, *56*, 2975-2984.
269. Yao, S.; Gallenkamp, D.; Wolfel, K.; Luke, B.; Schindler, M.; Scherkenbeck, J., Synthesis and SERCA activities of structurally simplified cyclopiazonic acid analogues. *Bioorg Med Chem* **2011**, *19*, 4669-78.
270. Kreuzer, J.; Bach, N. C.; Forler, D.; Sieber, S. A., Target discovery of acivicin in cancer cells elucidates its mechanism of growth inhibition. *Chemical Science* **2015**, *6*, 237-245.
271. Lushchak, V. I., Glutathione Homeostasis and Functions: Potential Targets for Medical Interventions. *Journal of Amino Acids* **2012**, *2012*, 26.
272. Banjac, A.; Perisic, T.; Sato, H.; Seiler, A.; Bannai, S.; Weiss, N.; Kolle, P.; Tschoep, K.; Issels, R. D.; Daniel, P. T.; Conrad, M.; Bornkamm, G. W., The cystine/cysteine cycle: a redox cycle regulating susceptibility versus resistance to cell death. *Oncogene* **2008**, *27*, 1618-28.
273. Ryter, S. W.; Tyrrell, R. M., The heme synthesis and degradation pathways: role in oxidant sensitivity. Heme oxygenase has both pro- and antioxidant properties. *Free Radic Biol Med* **2000**, *28*, 289-309.
274. Panchenko, M. V.; Farber, H. W.; Korn, J. H., Induction of heme oxygenase-1 by hypoxia and free radicals in human dermal fibroblasts. *American Journal of Physiology - Cell Physiology* **2000**, *278*, C92-C101.
275. Clements, C. M.; McNally, R. S.; Conti, B. J.; Mak, T. W.; Ting, J. P., DJ-1, a cancer- and Parkinson's disease-associated protein, stabilizes the antioxidant transcriptional master regulator Nrf2. *Proc Natl Acad Sci U S A* **2006**, *103*, 15091-6.
276. Richarme, G.; Mihoub, M.; Dairou, J.; Bui, L. C.; Leger, T.; Lamouri, A., Parkinsonism-associated protein DJ-1/Park7 is a major protein deglycase that repairs methylglyoxal- and glyoxal-glycated cysteine, arginine, and lysine residues. *J Biol Chem* **2015**, *290*, 1885-97.
277. Giordano, S.; Darley-Usmar, V.; Zhang, J., Autophagy as an essential cellular antioxidant pathway in neurodegenerative disease. *Redox Biol* **2014**, *2*, 82-90.

278. Itoh, K.; Wakabayashi, N.; Katoh, Y.; Ishii, T.; Igarashi, K.; Engel, J. D.; Yamamoto, M., Keap1 represses nuclear activation of antioxidant responsive elements by Nrf2 through binding to the amino-terminal Neh2 domain. *Genes Dev* **1999**, *13*, 76-86.
279. Dinkova-Kostova, A. T.; Holtzclaw, W. D.; Cole, R. N.; Itoh, K.; Wakabayashi, N.; Katoh, Y.; Yamamoto, M.; Talalay, P., Direct evidence that sulfhydryl groups of Keap1 are the sensors regulating induction of phase 2 enzymes that protect against carcinogens and oxidants. *Proc Natl Acad Sci U S A* **2002**, *99*, 11908-13.
280. Kommi, D. N.; Jadhavar, P. S.; Kumar, D.; Chakraborti, A. K., "All-water" one-pot diverse synthesis of 1,2-disubstituted benzimidazoles: hydrogen bond driven 'synergistic electrophile-nucleophile dual activation' by water. *Green Chem* **2013**, *15*, 798-810.
281. Thomas, J. C. Throwing a chemical spanner in the malaria invasion motor: Interaction and dynamics of the Plasmodium MTIP/MyoA complex. PhD Thesis, Imperial College London 2011.
282. Nairoukh, Z.; Avnir, D.; Blum, J., Acid-catalyzed hydration of alkynes in aqueous microemulsions. *ChemSusChem* **2013**, *6*, 430-2.
283. Riedl, T.; Hanaoka, F.; Egly, J. M., The comings and goings of nucleotide excision repair factors on damaged DNA. *EMBO J* **2003**, *22*, 5293-303.
284. Uhlén, M.; Fagerberg, L.; Hallström, B. M.; Lindskog, C.; Oksvold, P.; Mardinoglu, A.; Sivertsson, Å.; Kampf, C.; Sjöstedt, E.; Asplund, A.; Olsson, I.; Edlund, K.; Lundberg, E.; Navani, S.; Szigartyo, C. A.-K.; Odeberg, J.; Djureinovic, D.; Takanen, J. O.; Hober, S.; Alm, T.; Edqvist, P.-H.; Berling, H.; Tegel, H.; Mulder, J.; Rockberg, J.; Nilsson, P.; Schwenk, J. M.; Hamsten, M.; von Feilitzen, K.; Forsberg, M.; Persson, L.; Johansson, F.; Zwahlen, M.; von Heijne, G.; Nielsen, J.; Pontén, F., Tissue-based map of the human proteome. *Science* **2015**, *347*.
285. Swope, D.; Li, J.; Radice, G. L., Beyond cell adhesion: the role of armadillo proteins in the heart. *Cell Signal* **2013**, *25*, 93-100.
286. Jäkel, S.; Mingot, J. M.; Schwarzmaier, P.; Hartmann, E.; Görlich, D., Importins fulfil a dual function as nuclear import receptors and cytoplasmic chaperones for exposed basic domains. *The EMBO Journal* **2002**, *21*, 377-386.
287. Clulow, J. A. Unravelling the targets of electrophilic natural products in cancer with chemical proteomics. PhD Thesis Imperial College London 2015.
288. Kalesh, K. A.; Clulow, J. A.; Tate, E. W., Target profiling of zerumbone using a novel cell-permeable clickable probe and quantitative chemical proteomics. *Chemical Communications* **2015**, *51*, 5497-5500.
289. Mashalidis, E. H.; Sledz, P.; Lang, S.; Abell, C., A three-stage biophysical screening cascade for fragment-based drug discovery. *Nat Protoc* **2013**, *8*, 2309-24.
290. Hajduk, P. J.; Greer, J., A decade of fragment-based drug design: strategic advances and lessons learned. *Nat Rev Drug Discov* **2007**, *6*, 211-9.
291. Pirard, B.; Ertl, P., Evaluation of a Semi-Automated Workflow for Fragment Growing. *Journal of Chemical Information and Modeling* **2015**, *55*, 180-193.
292. Hung, A. W.; Silvestre, H. L.; Wen, S.; Ciulli, A.; Blundell, T. L.; Abell, C., Application of fragment growing and fragment linking to the discovery of inhibitors of Mycobacterium tuberculosis pantothenate synthetase. *Angew Chem Int Ed Engl* **2009**, *48*, 8452-6.
293. Niesen, F. H.; Berglund, H.; Vedadi, M., The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat Protoc* **2007**, *2*, 2212-21.
294. Mayer, M.; Meyer, B., Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy. *Angewandte Chemie International Edition* **1999**, *38*, 1784-1788.
295. Dalvit, C.; Fogliatto, G.; Stewart, A.; Veronesi, M.; Stockman, B., WaterLOGSY as a method for primary NMR screening: practical aspects and range of applicability. *J Biomol NMR* **2001**, *21*, 349-59.
296. Wienken, C. J.; Baaske, P.; Rothbauer, U.; Braun, D.; Duhr, S., Protein-binding assays in biological liquids using microscale thermophoresis. *Nat Commun* **2010**, *1*, 100.

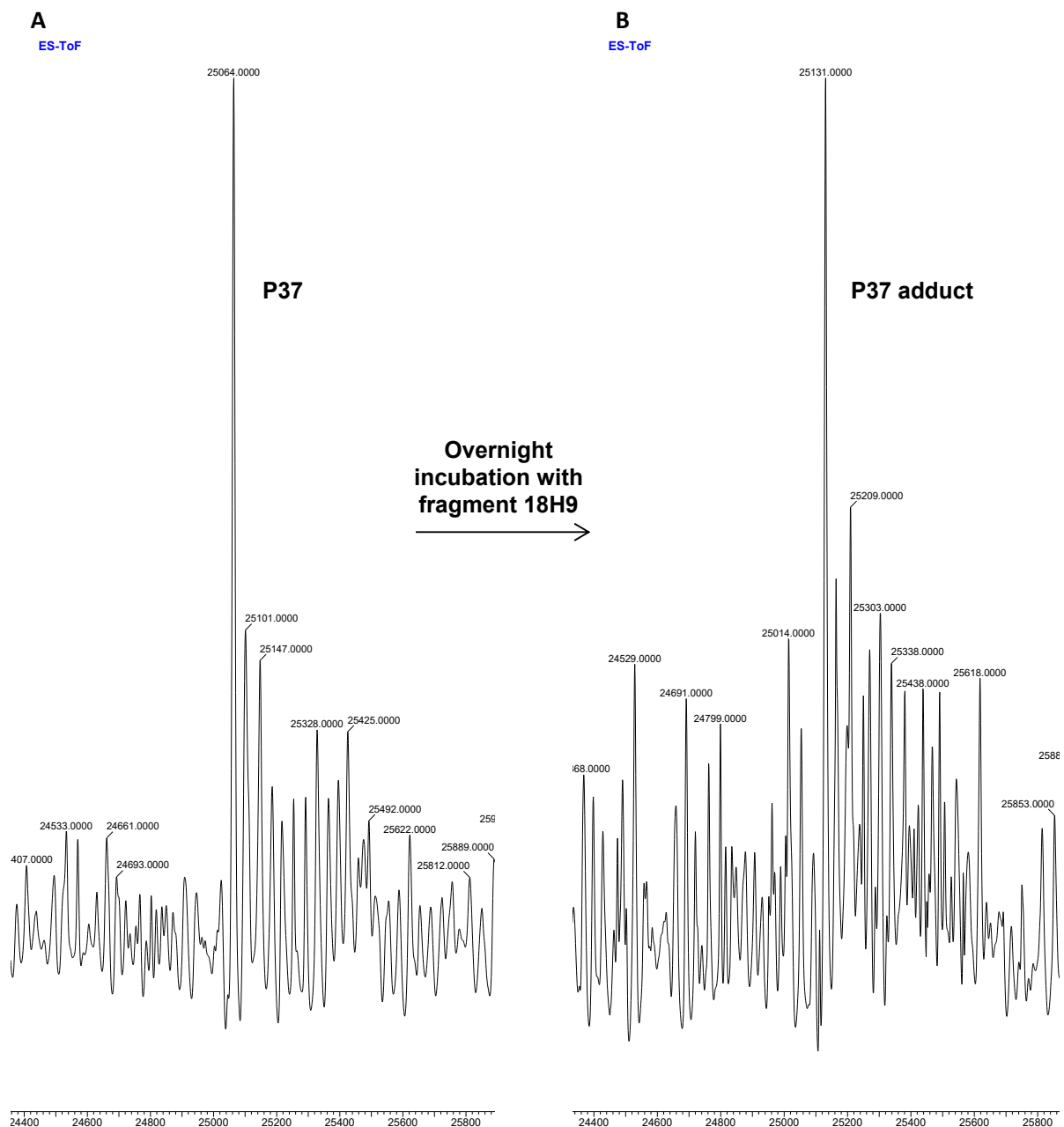
297. Blundell, T. L.; Patel, S., High-throughput X-ray crystallography for drug discovery. *Current Opinion in Pharmacology* **2004**, *4*, 490-496.
298. Stieglitz, B.; Rana, R. R.; Koliopoulos, M. G.; Morris-Davies, A. C.; Schaeffer, V.; Christodoulou, E.; Howell, S.; Brown, N. R.; Dikic, I.; Rittinger, K., Structural basis for ligase-specific conjugation of linear ubiquitin chains by HOIP. *Nature* **2013**, *503*, 422-6.
299. Reinhard, L.; Mayerhofer, H.; Geerlof, A.; Mueller-Dieckmann, J.; Weiss, M. S., Optimization of protein buffer cocktails using ThermoFluor. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications* **2013**, *69*, 209-214.
300. Douse, C. H.; Vrieling, N.; Wenlin, Z.; Cota, E.; Tate, E. W., Targeting a dynamic protein-protein interaction: fragment screening against the malaria myosin A motor complex. *ChemMedChem* **2015**, *10*, 134-43.
301. Cimperman, P.; Baranauskiene, L.; Jachimoviciute, S.; Jachno, J.; Torresan, J.; Michailoviene, V.; Matuliene, J.; Sereikaite, J.; Bumelis, V.; Matulis, D., A quantitative model of thermal stabilization and destabilization of proteins by ligands. *Biophys J* **2008**, *95*, 3222-31.
302. Matulis, D.; Kranz, J. K.; Salemme, F. R.; Todd, M. J., Thermodynamic stability of carbonic anhydrase: measurements of binding affinity and stoichiometry using ThermoFluor. *Biochemistry* **2005**, *44*, 5258-66.
303. Scott, D. E.; Coyne, A. G.; Hudson, S. A.; Abell, C., Fragment-based approaches in drug discovery and chemical biology. *Biochemistry* **2012**, *51*, 4990-5003.
304. Ilari, A.; Savino, C., Protein Structure Determination by X-Ray Crystallography. In *Bioinformatics*, Keith, J., Ed. Humana Press 2008; Vol. 452, pp 63-87.
305. Hartshorn, M. J.; Murray, C. W.; Cleasby, A.; Frederickson, M.; Tickle, I. J.; Jhoti, H., Fragment-based lead discovery using X-ray crystallography. *J Med Chem* **2005**, *48*, 403-13.
306. Morrison, K. E., Parkin mutations and early onset parkinsonism. *Brain* **2003**, *126*, 1250-1.
307. Duda, D. M.; Olszewski, J. L.; Schuermann, J. P.; Kurinov, I.; Miller, D. J.; Nourse, A.; Alpi, A. F.; Schulman, B. A., Structure of HHARI, a RING-IBR-RING ubiquitin ligase: autoinhibition of an Ariadne-family E3 and insights into ligation mechanism. *Structure* **2013**, *21*, 1030-41.
308. Tan, N. G. S.; Ardley, H. C.; Scott, G. B.; Rose, S. A.; Markham, A. F.; Robinson, P. A., Human homologue of ariadne promotes the ubiquitylation of translation initiation factor 4E homologous protein, 4EHP. *FEBS Letters* **2003**, *554*, 501-504.
309. Guimaraes, C. P.; Witte, M. D.; Theile, C. S.; Bozkurt, G.; Kundrat, L.; Blom, A. E. M.; Ploegh, H. L., Site-specific C-terminal and internal loop labeling of proteins using sortase-mediated reactions. *Nat. Protocols* **2013**, *8*, 1787-1799.
310. El-Mahdy, A. F. M.; Shibata, T.; Kabashima, T.; Kai, M., Dendrimer-like polymeric DNAs as chemiluminescence probes for amplified detection of telomere DNA on a solid-phase membrane. *Chemical Communications* **2014**, *50*, 859-861.
311. Armarego, W. L. F.; Perrin, D. D., *Purification of Laboratory Chemicals*. 4 ed.; Butterworth Heinemann, 4th edn 1997, p342.
312. Chan, W. C.; White, P. D., *Fmoc solid Phase Peptide Synthesis*. Oxford University Press.
313. Mann, M., Functional and quantitative proteomics using SILAC. *Nat Rev Mol Cell Biol* **2006**, *7*, 952-8.
314. 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* **2003**, *75*, 663-70.

## Appendices and spectra

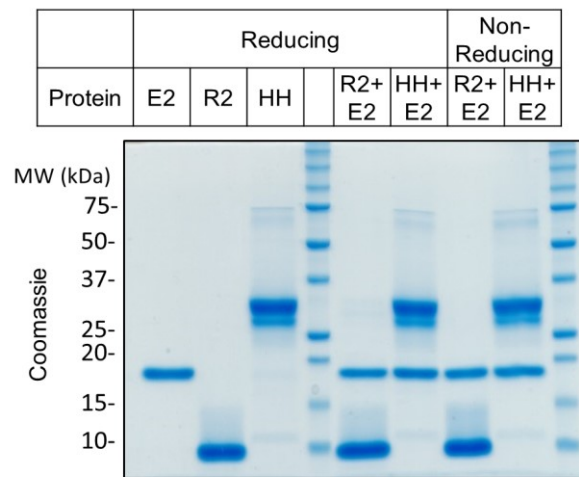
---

**Appendix Table 1: List of electronic files available for further reference. For LC-MS/MS analysis, both raw (Output from MaxQuant search) and analysis data sheets are included.**

<b>File A</b>	List of all known Ub and Ubl machinery and other proteins associated with Ub/Ubl conjugation. The list was assembled from Uniprot: ( <a href="http://www.uniprot.org/">http://www.uniprot.org/</a> )
<b>File 1</b>	LFQ analysis of Pg-12-mer-VS and Pg-12-mer-VME probe targets in HeLa cell lysate. Results from shear force and chemical lysis approaches are included.
<b>File 2</b>	LFQ analysis of Pg-12-mer-VME probe targets in fractionated MCF7 cell lysate. Methods with and without precipitation following CuAAC is compared.
<b>File 3</b>	LFQ analysis of biotin-Ahx-Ub-vme (200 nM) in MCF7 cell lysate.
<b>File 4</b>	Spike-in SILAC experiment of Pg-12-mer-VME (10 $\mu$ M) in competition with 0, 20, and 50 $\mu$ M of 12-mer-vme in MCF7 cell lysate. Data analysis with averaged L/H ratios for each concentration normalised to no competition (12-mer-vme = 0 $\mu$ M). GO-term analysis of competed targets was undertaken using STRING software.
<b>File 5</b>	Spike-in SILAC experiment of Pg-12-mer-VME (10 $\mu$ M) labelling in epirubicin (0, 0.1, and 1.0 $\mu$ M) treated MCF7 cell lysate. Data analysis with averaged L/H ratios for each concentration normalised to no DNA damage (epirubicin = 0 $\mu$ M).
<b>File 6</b>	LFQ analysis of PYR-41-P (50 $\mu$ M) and HeLa cell lysate.
<b>File 7</b>	LFQ analysis of PYR-41-P (50 $\mu$ M) and T2 (100 $\mu$ M) probe labelling in MCF7 cells. Average LFQ intensities are shown. GO-term analysis was undertaken on the top 50 targets ranked by LFQ intensity using STRING software.
<b>File 8</b>	Spike-in SILAC experiment of PYR-41-P (25 $\mu$ M) in competition with 0, 125, and 250 $\mu$ M of PYR-41 in MCF7 cells. Data analysis with averaged L/H ratios for each concentration normalised to no competition (PYR-41 = 0 $\mu$ M).
<b>File 9</b>	Spike-in SILAC experiment of T2 (20 $\mu$ M) in competition with 0, 40, and 200 $\mu$ M of T2c in MCF7 cells. Data analysis with averaged L/H ratios for each concentration normalised to no competition (T2c = 0 $\mu$ M).



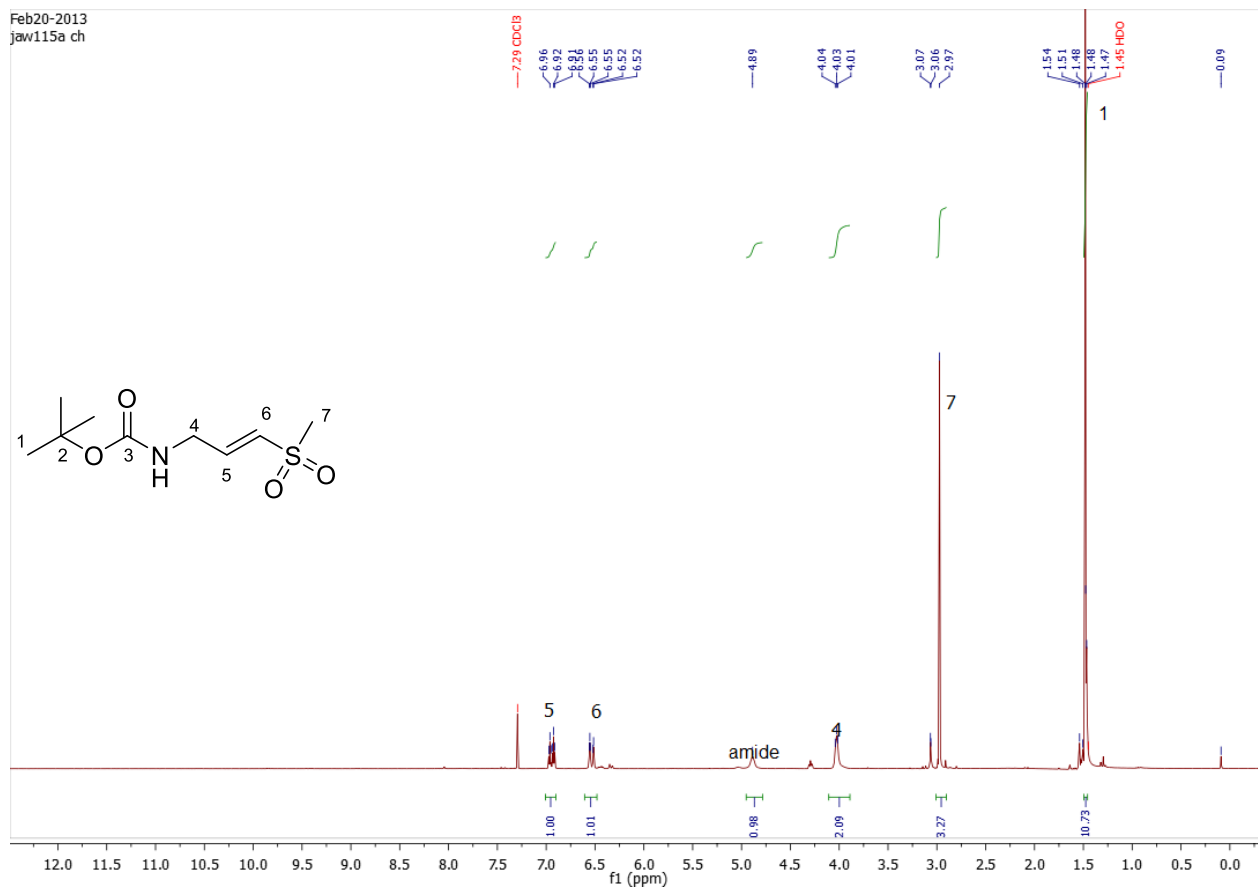
**Appendix Figure 71: MS spectral analysis of fragment binding. P37 was incubated overnight in the absence (A) and presence (B) of fragment 18H9 and analysed by ES-TOF MS. Spectra show a mass shift 68 Da, indicating that 18H9 is covalently reacting with P37.**



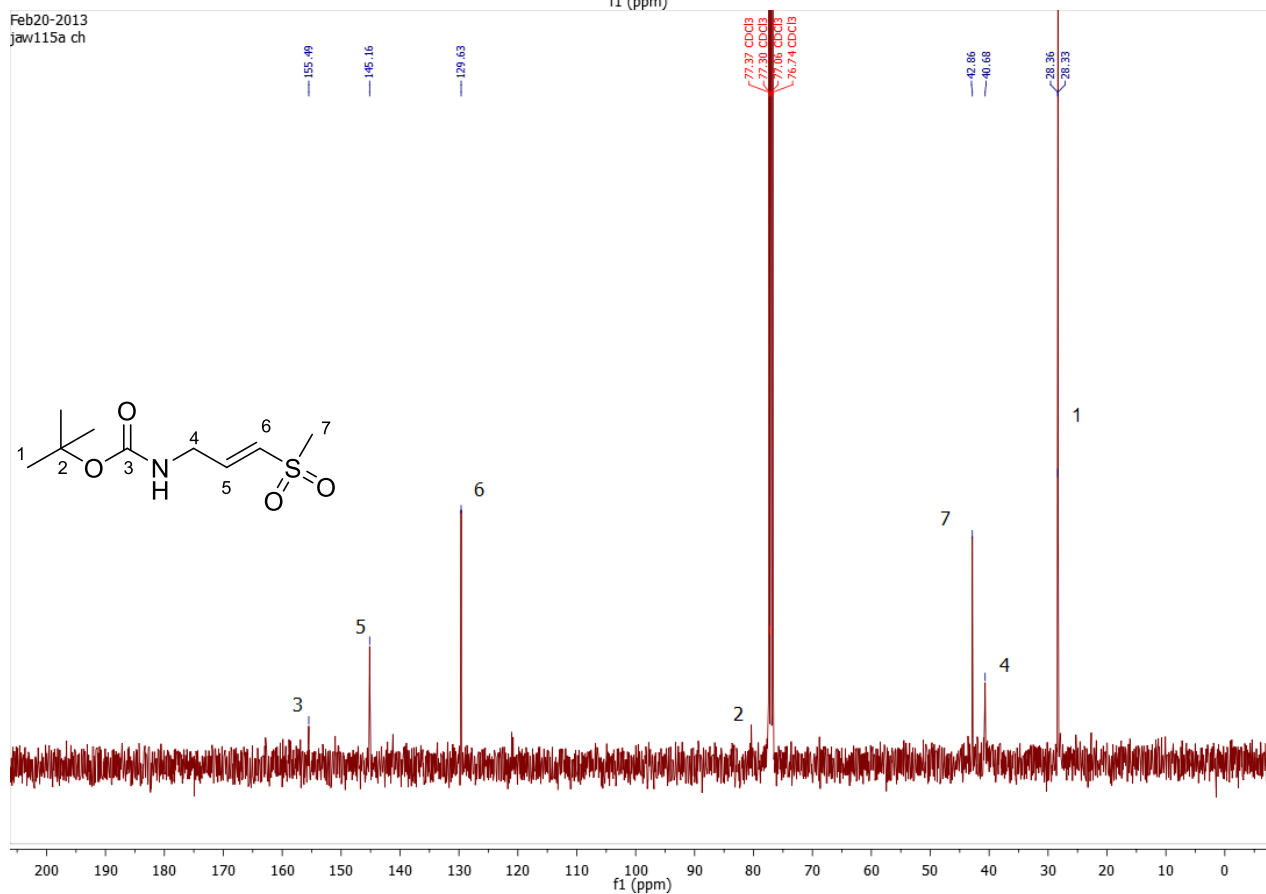
**Appendix Figure2: Determining the ability of Ubch7 (E2) to form heterodimers with R2 and HH which are stable on SDS-PAGE analysis. Proteins were incubated on ice for 3 min, any interactions quenched on addition of DTT, samples separated and analysed by Coomassie staining.**

# NMR Spectra: Compound 2

Feb20-2013  
jaw115a ch

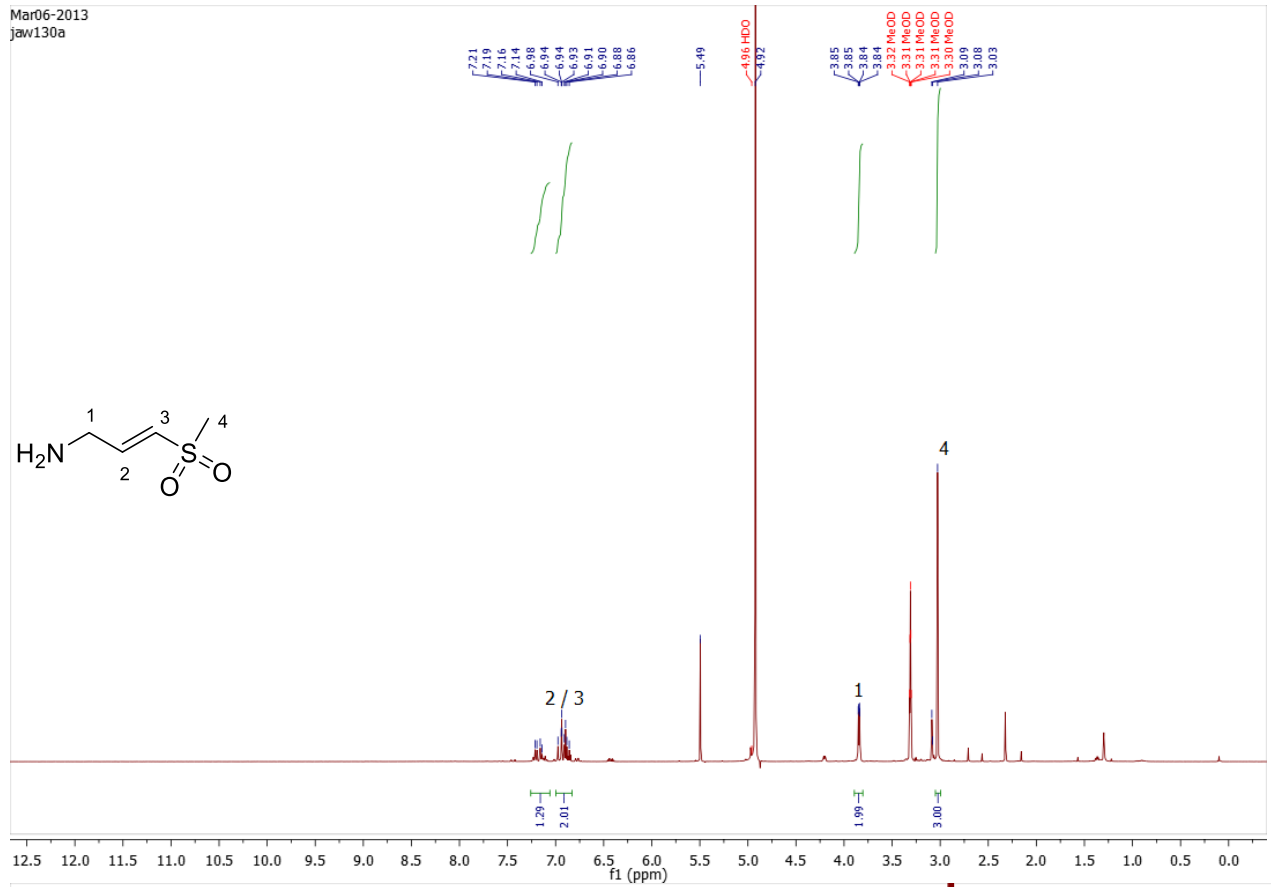


Feb20-2013  
jaw115a ch

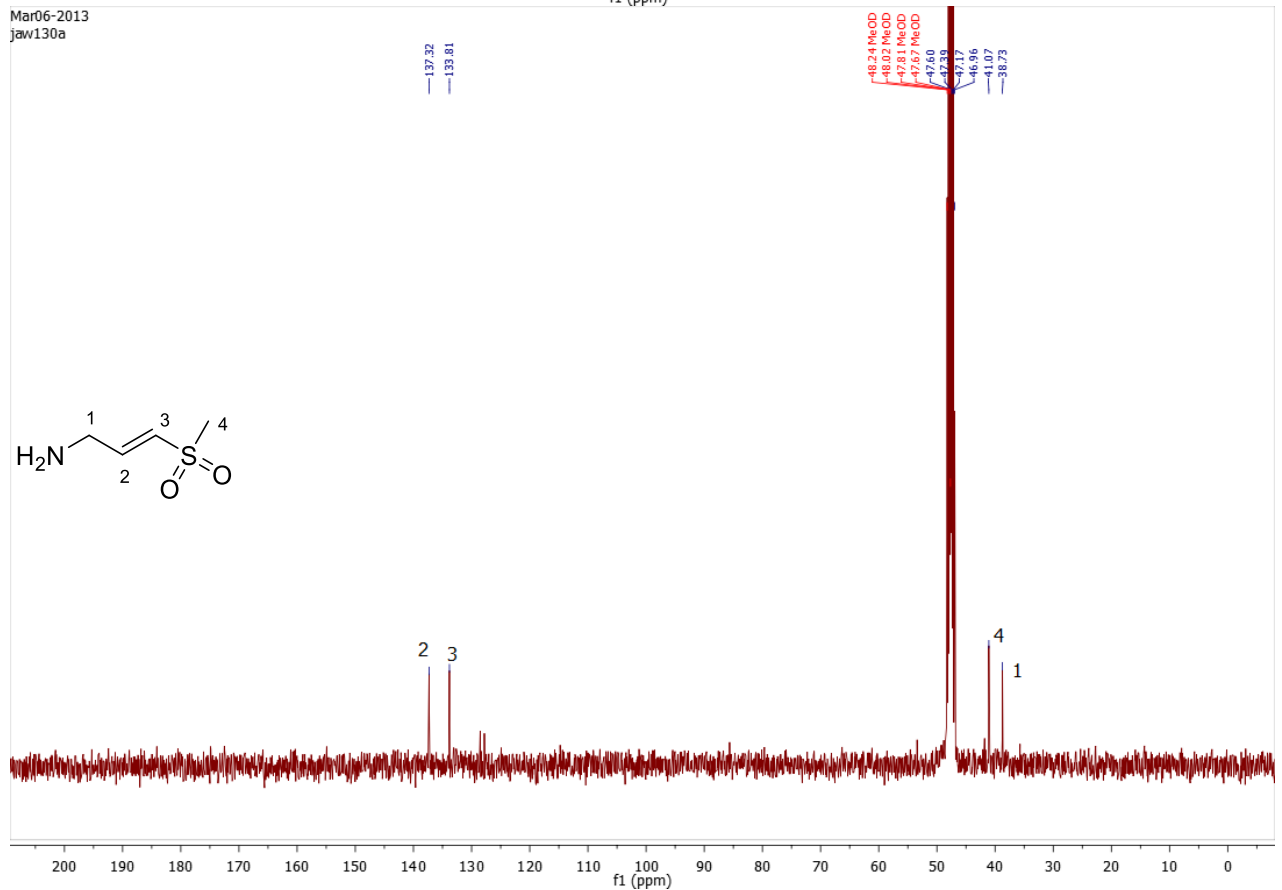


# NMR Spectra: VS (TFA salt)

Mar06-2013  
jaw130a



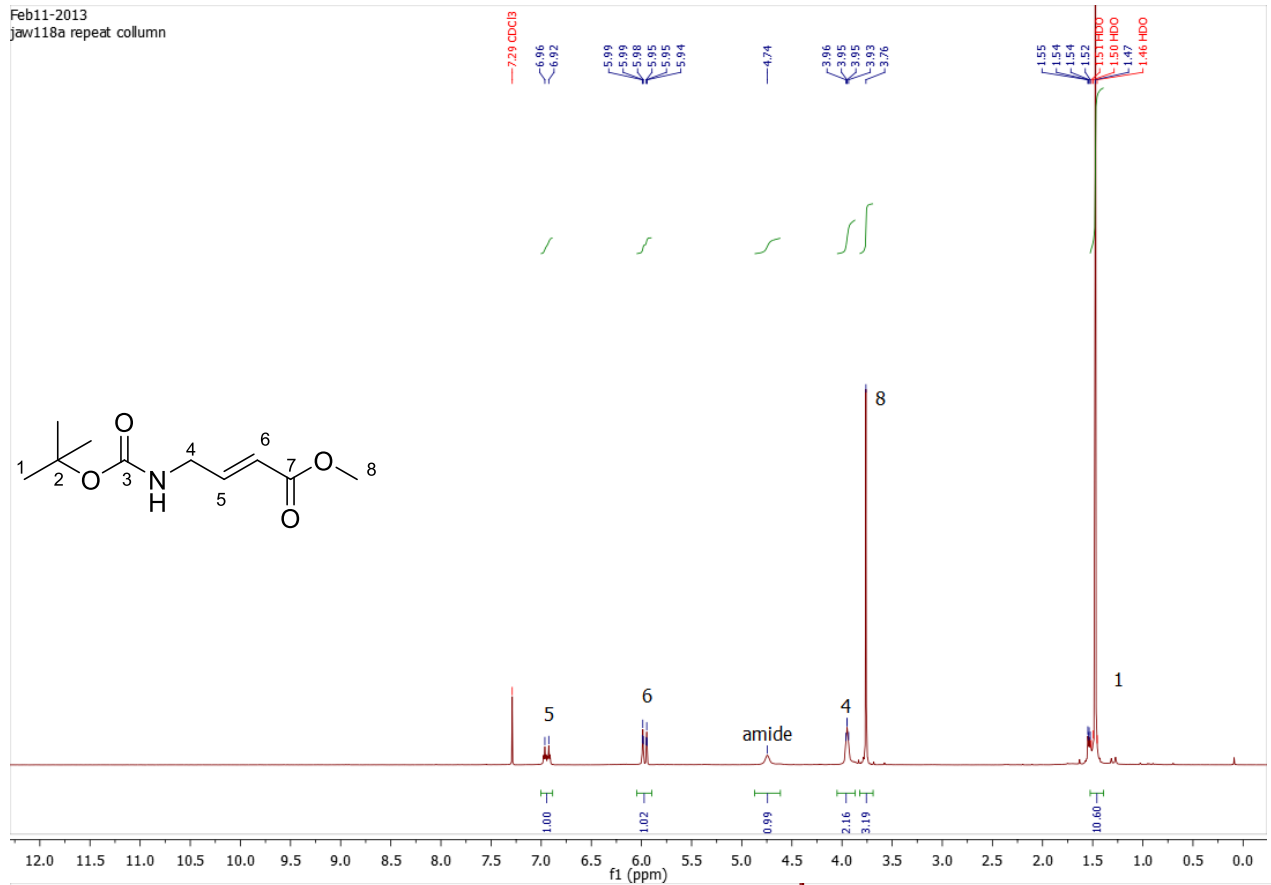
Mar06-2013  
jaw130a



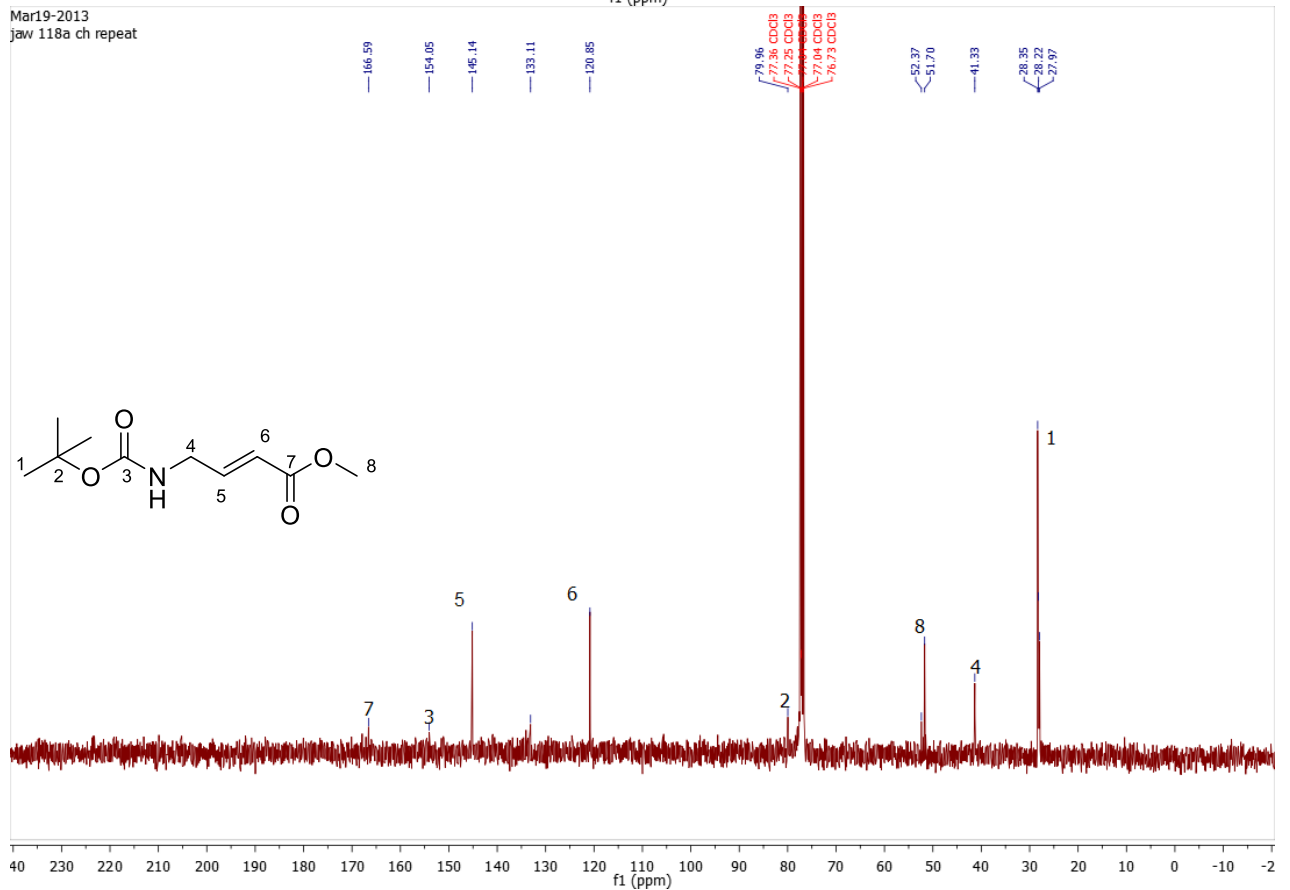


# NMR Spectra: Compound 3

Feb11-2013  
jaw118a repeat column

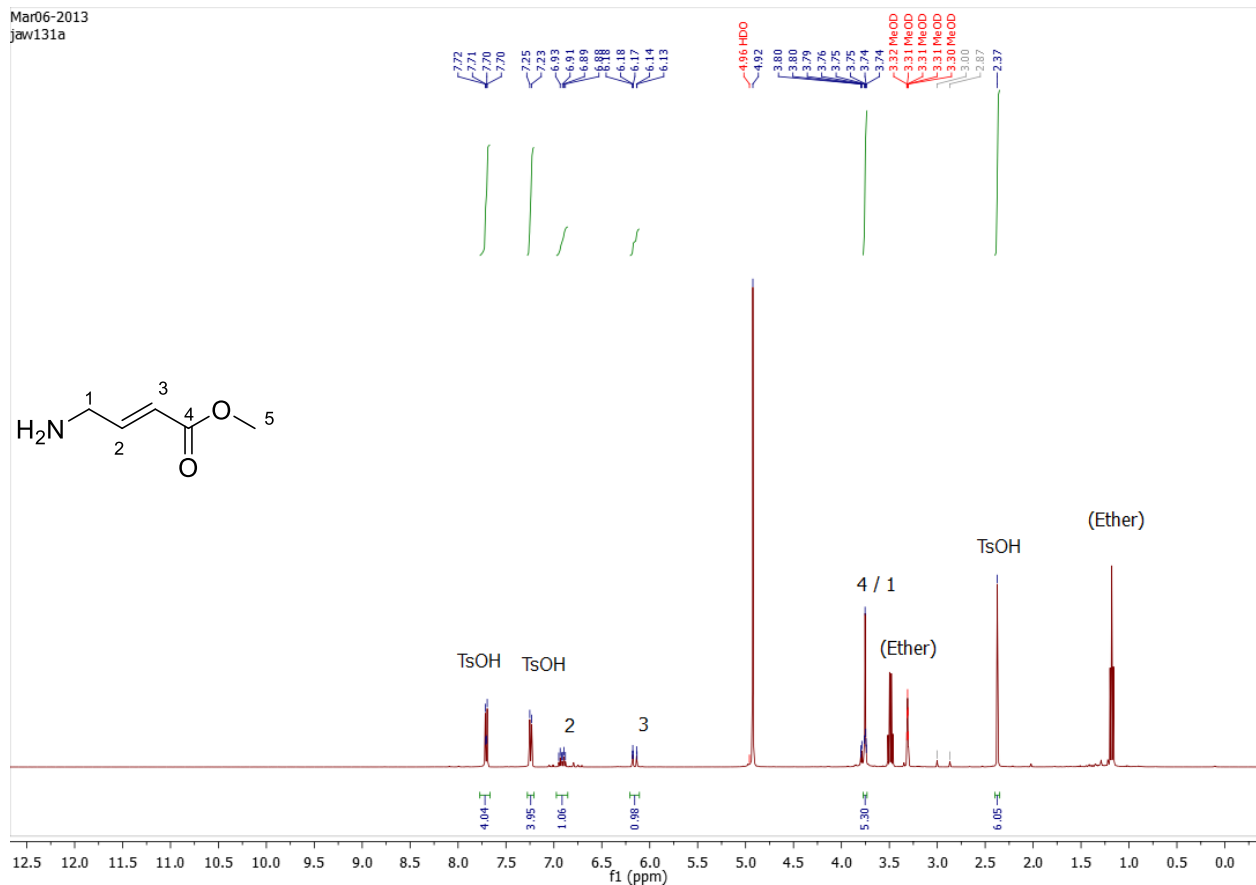


Mar19-2013  
jaw 118a ch repeat

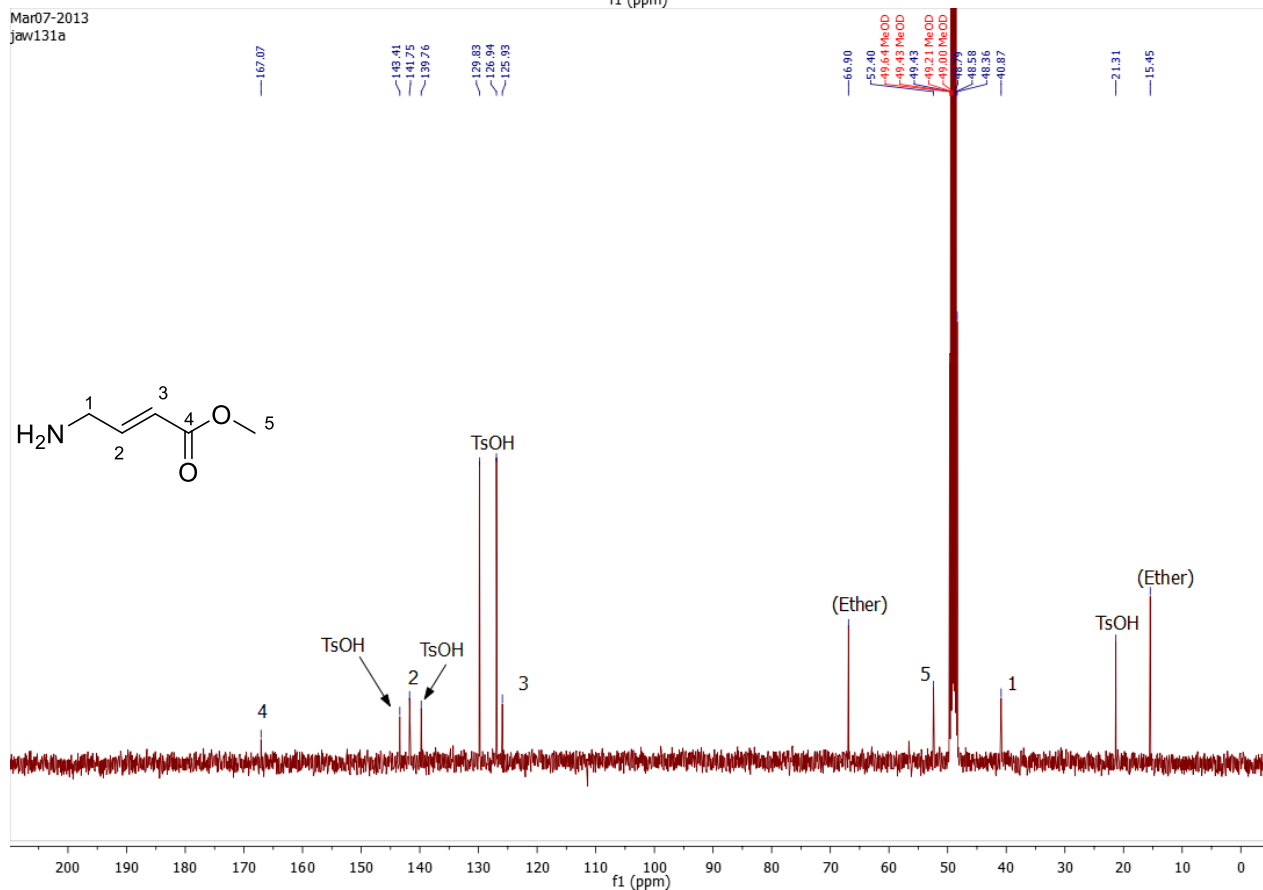


# NMR Spectra: VME (tosylate salt)

Mar06-2013  
jaw131a

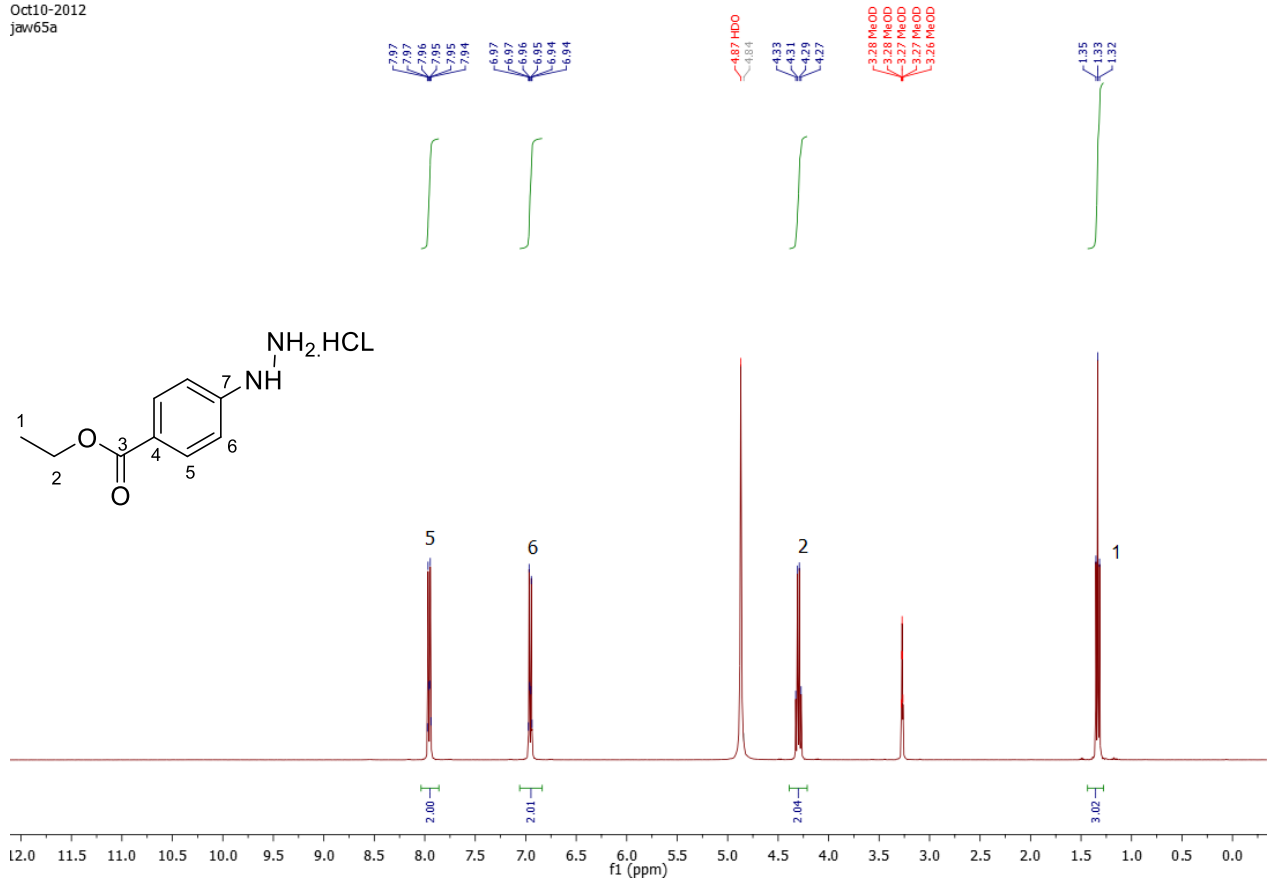


Mar07-2013  
jaw131a

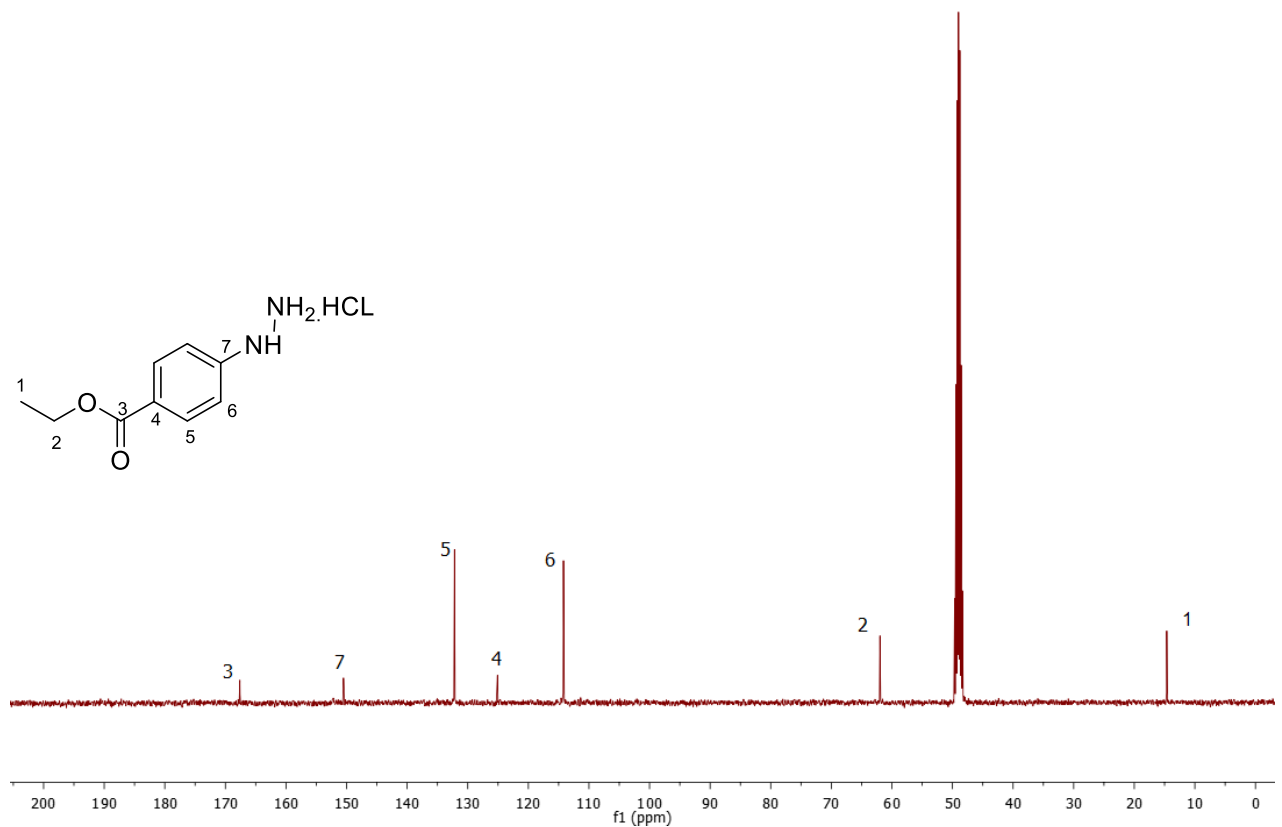


# NMR Spectra: Compound 7

Oct10-2012  
jaw65a

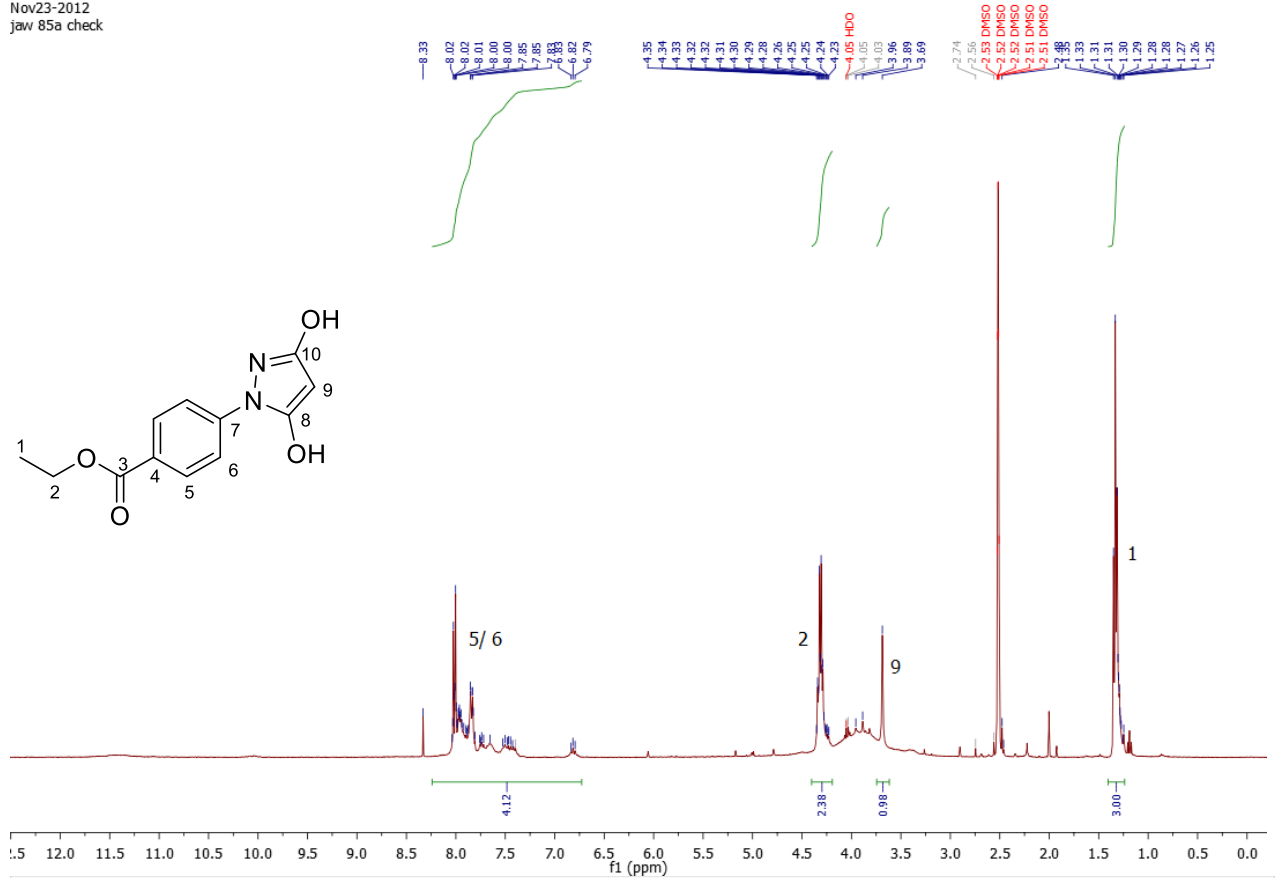


Oct10-2012  
jaw65a carbon

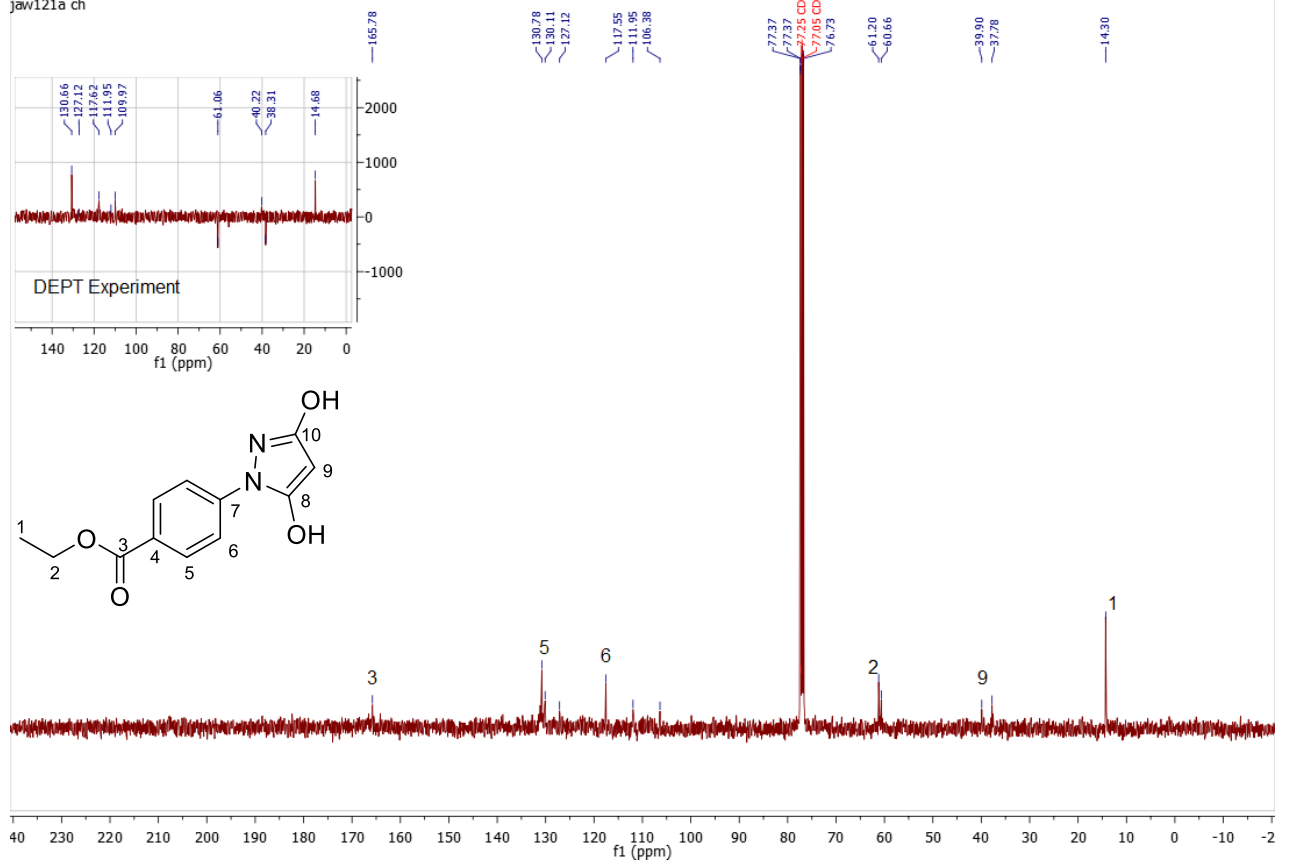


# NMR Spectra: Compound 5

Nov23-2012  
jaw 85a check

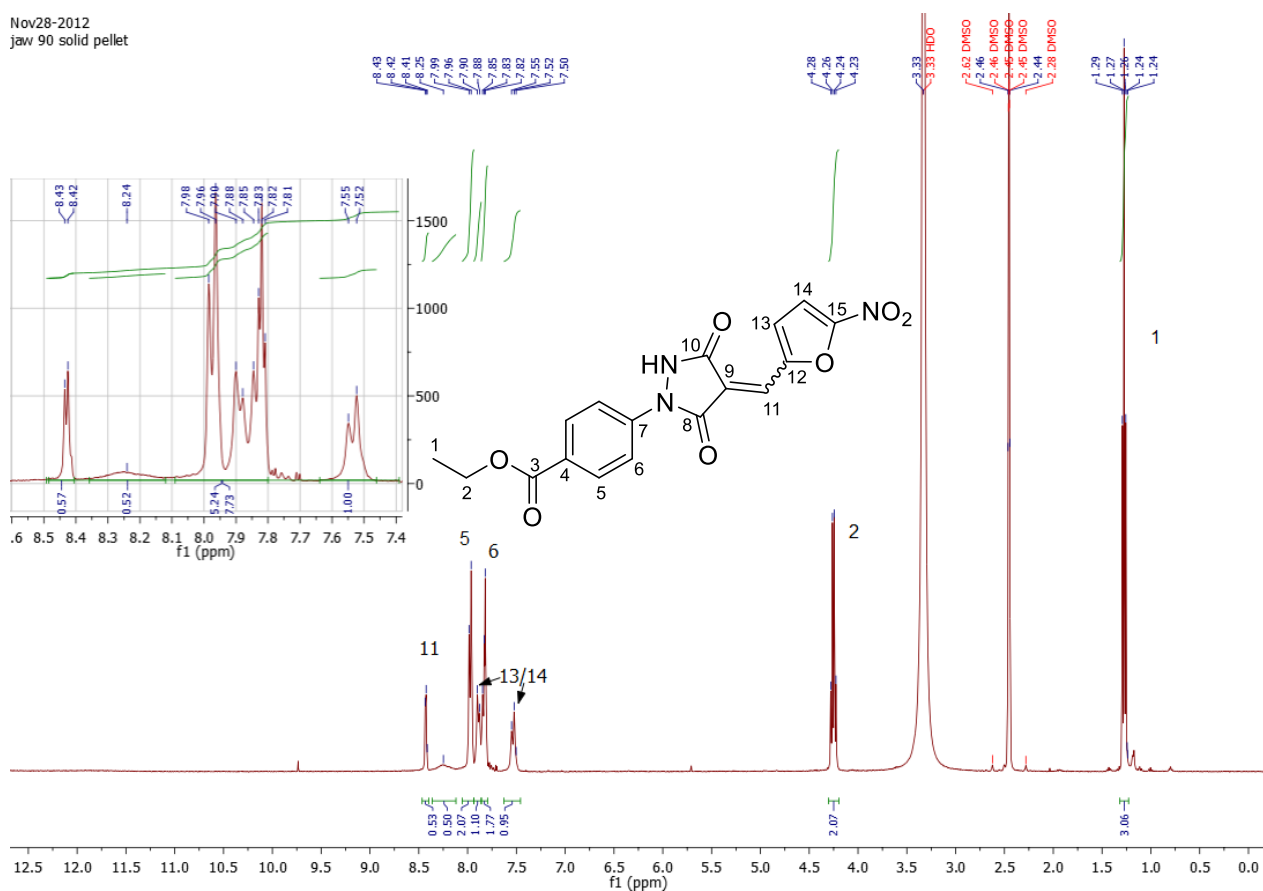


Feb20-2013  
jaw121a ch

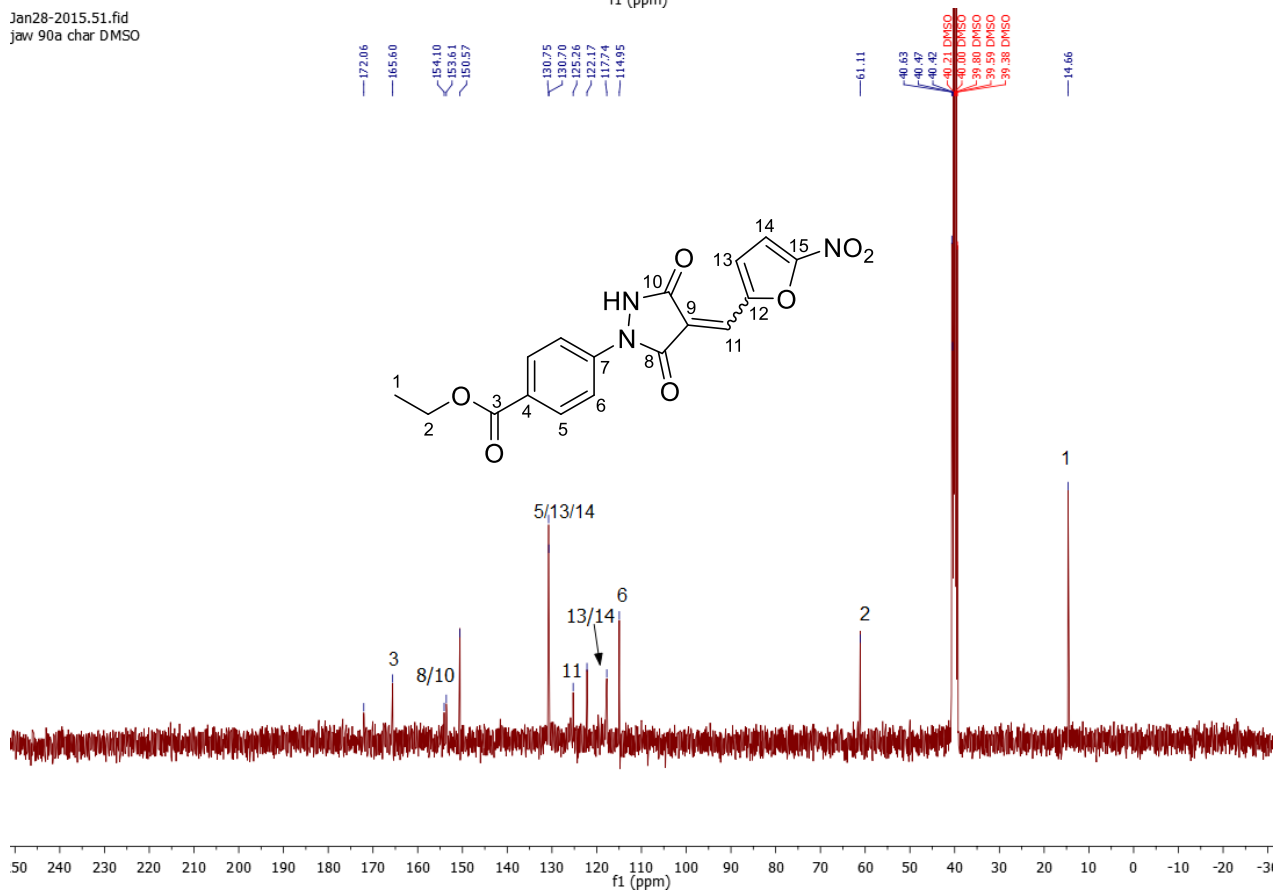


# NMR Spectra: PYR-41

Nov28-2012  
jaw 90 solid pellet

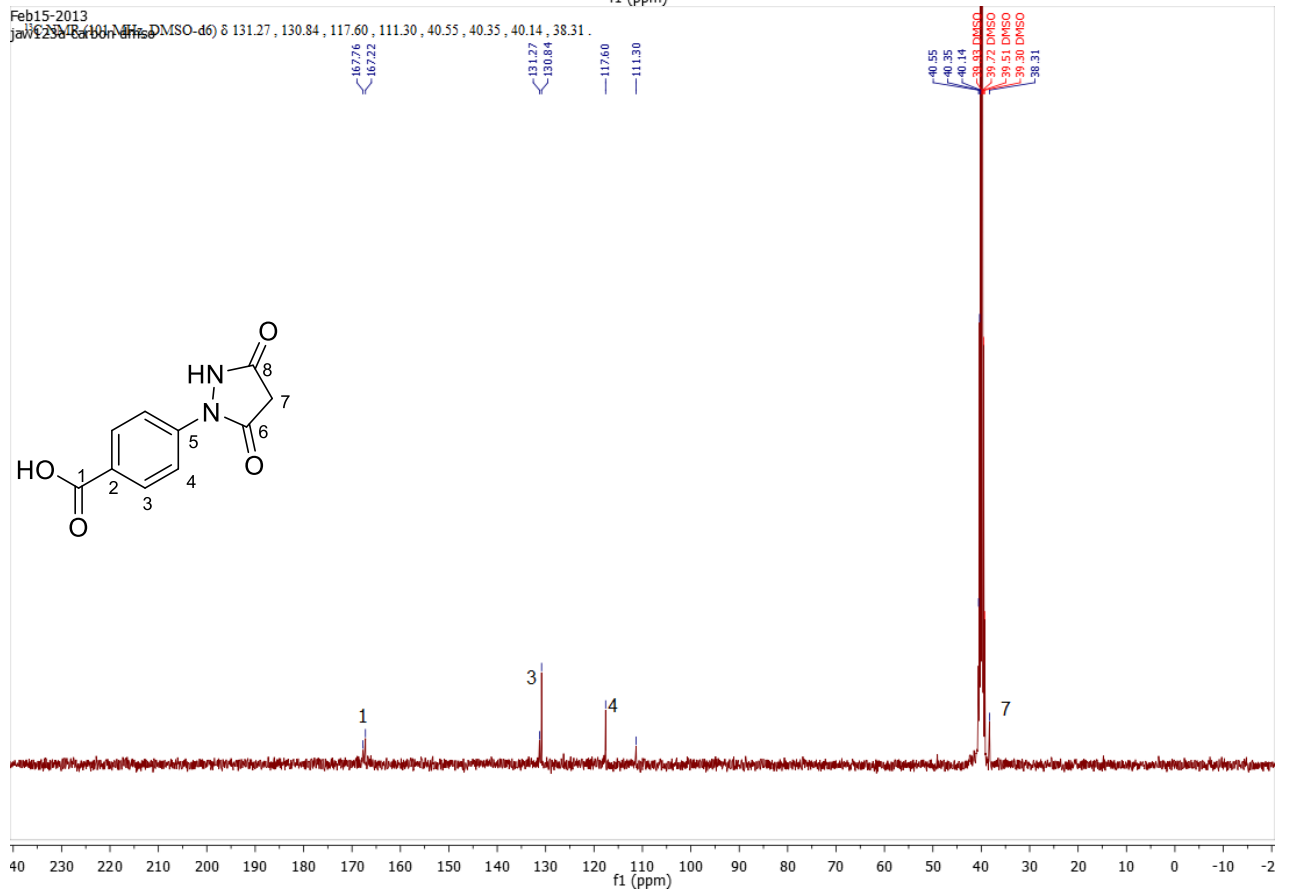
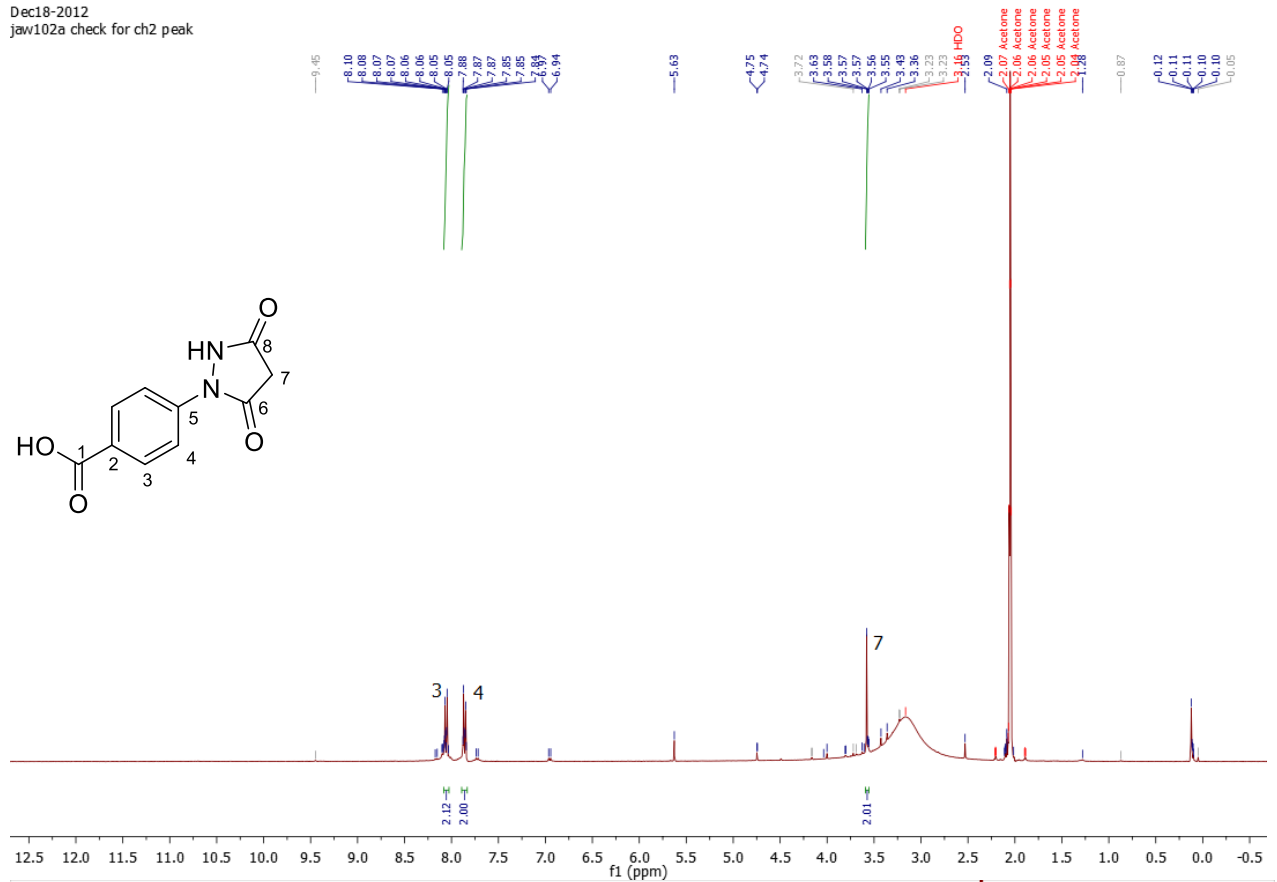


Jan28-2015.51.fid  
jaw 90a char DMSO

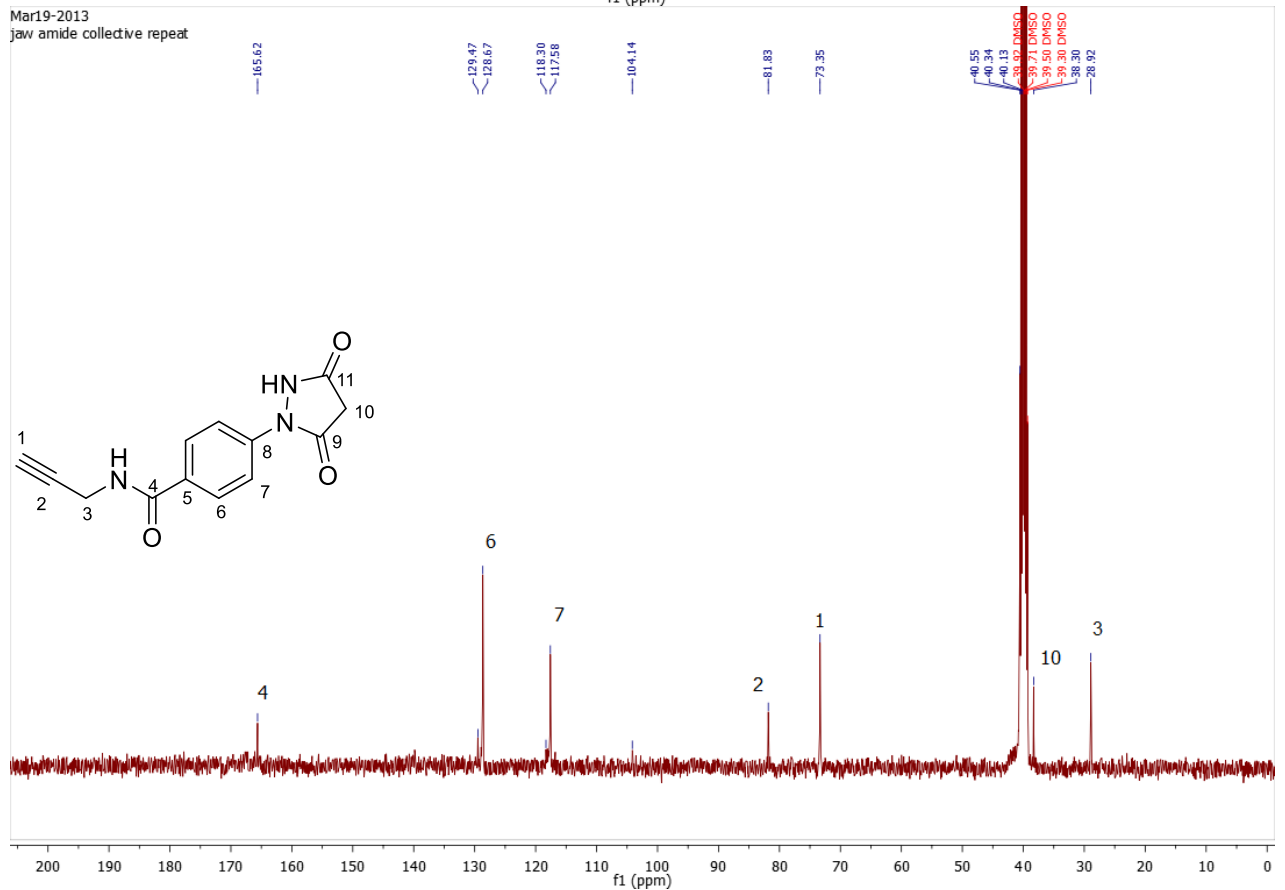
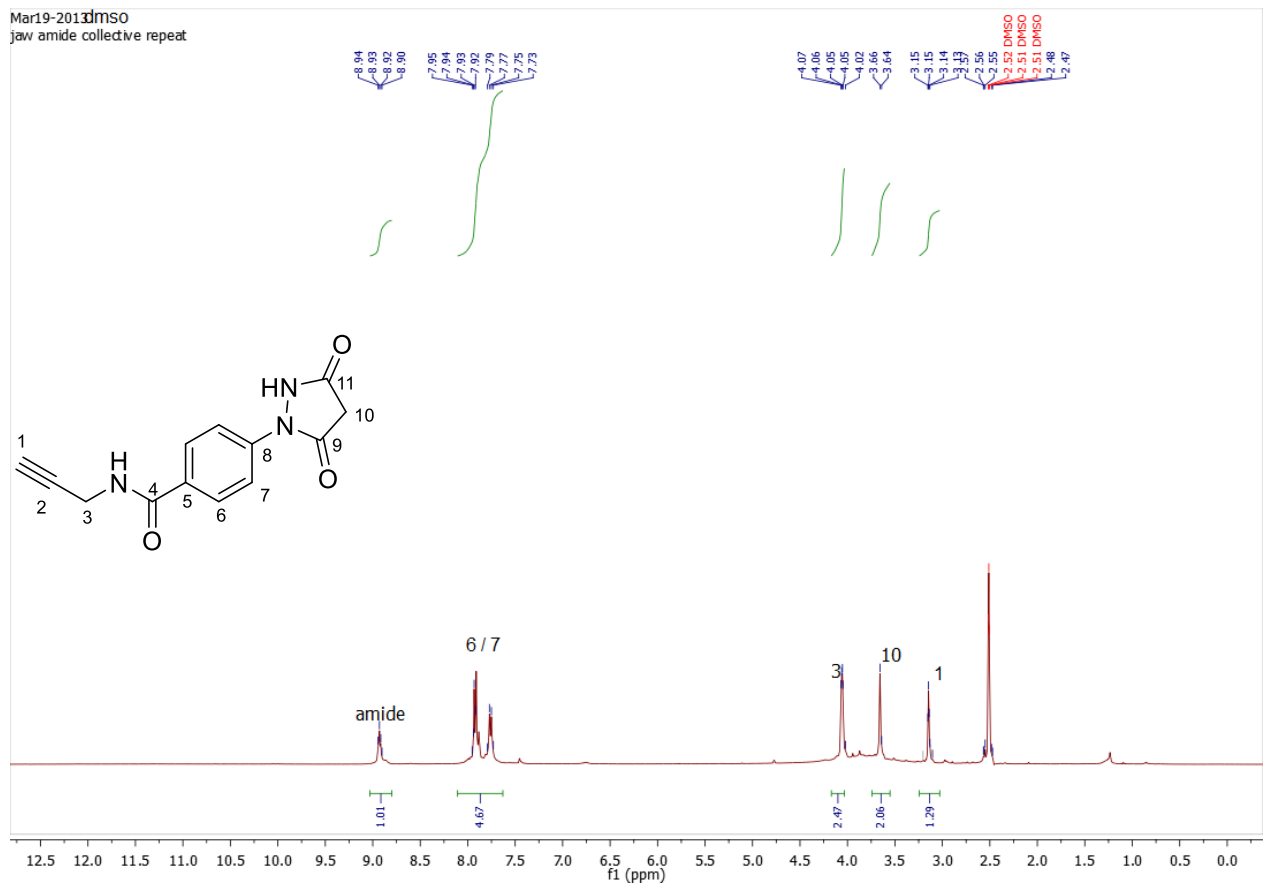


# NMR Spectra: Compound 9

Dec18-2012  
jaw102a check for ch2 peak

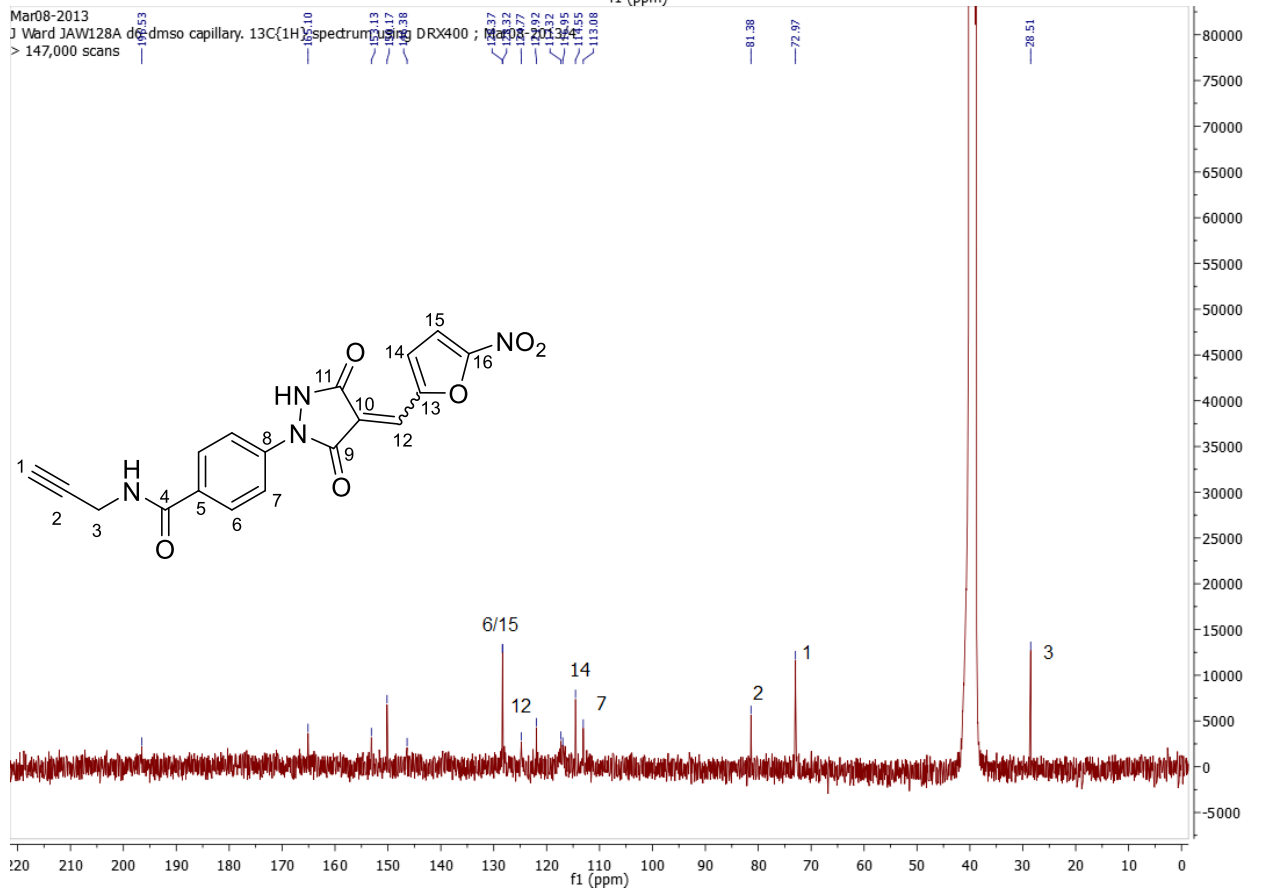
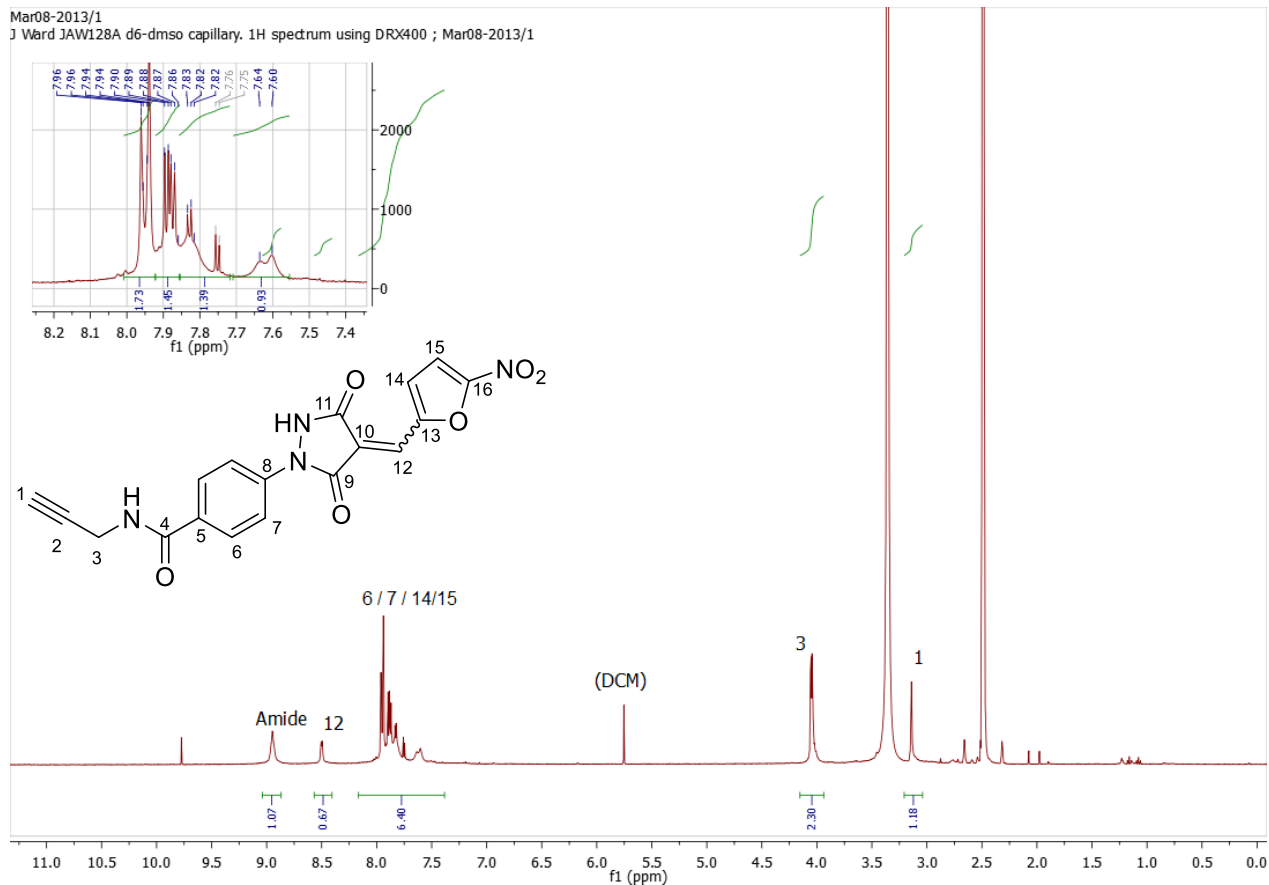


# NMR Spectra: Compound 4



# NMR Spectra: PYR-41-P

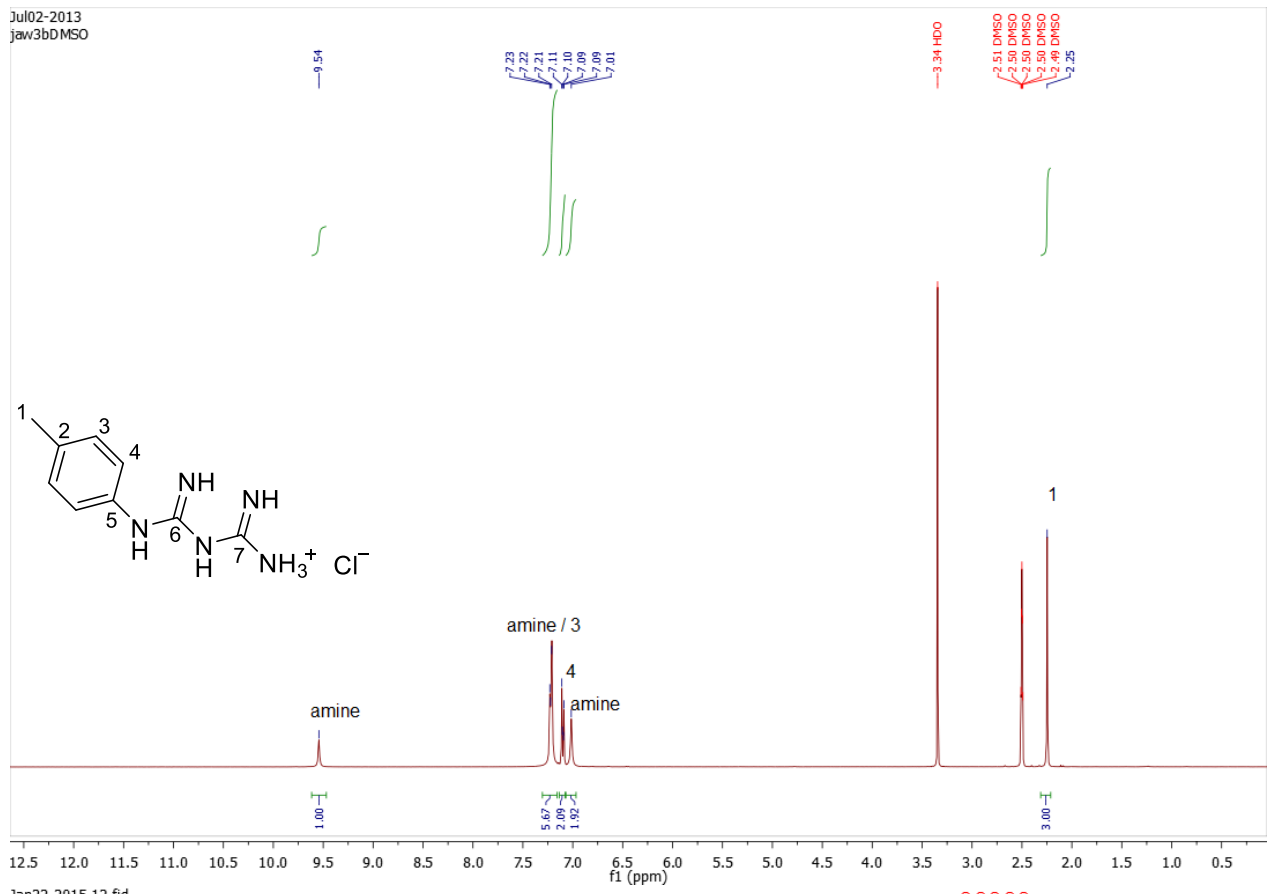
Mar08-2013/1  
 J Ward JAW128A d6-dmso capillary. 1H spectrum using DRX400 ; Mar08-2013/1



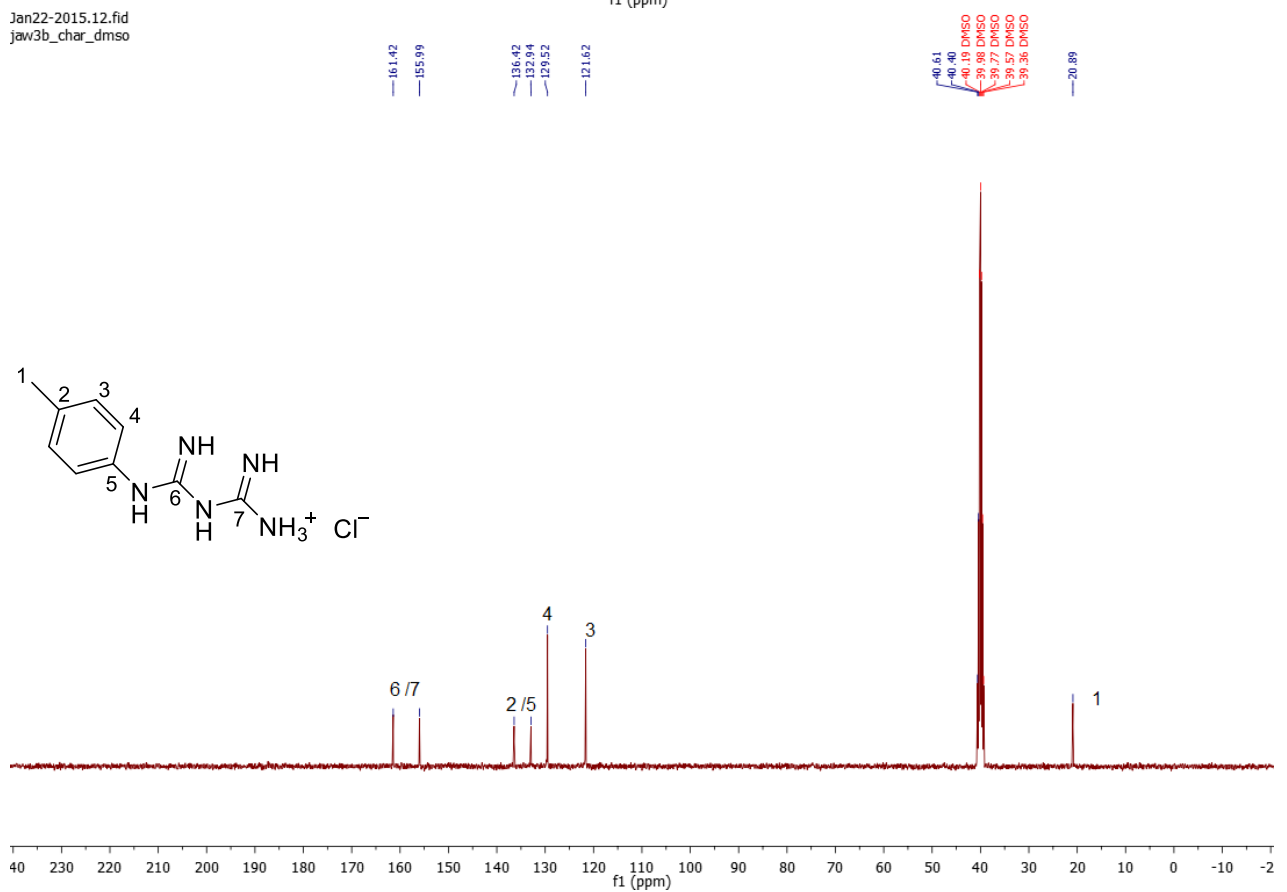


# NMR Spectra: Compound 11

Jul02-2013  
jaw3bDMSO

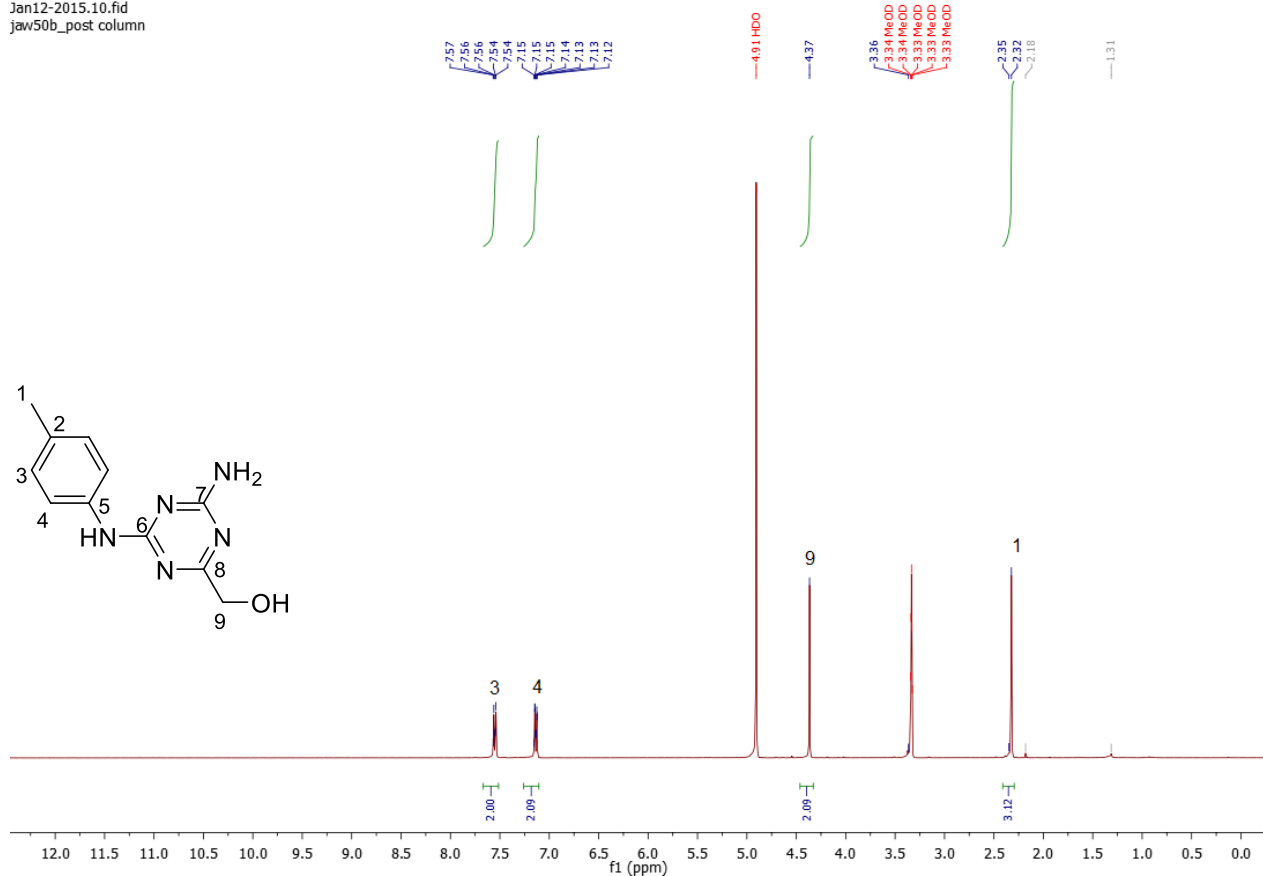


Jan22-2015.12.fid  
jaw3b\_char\_dmso

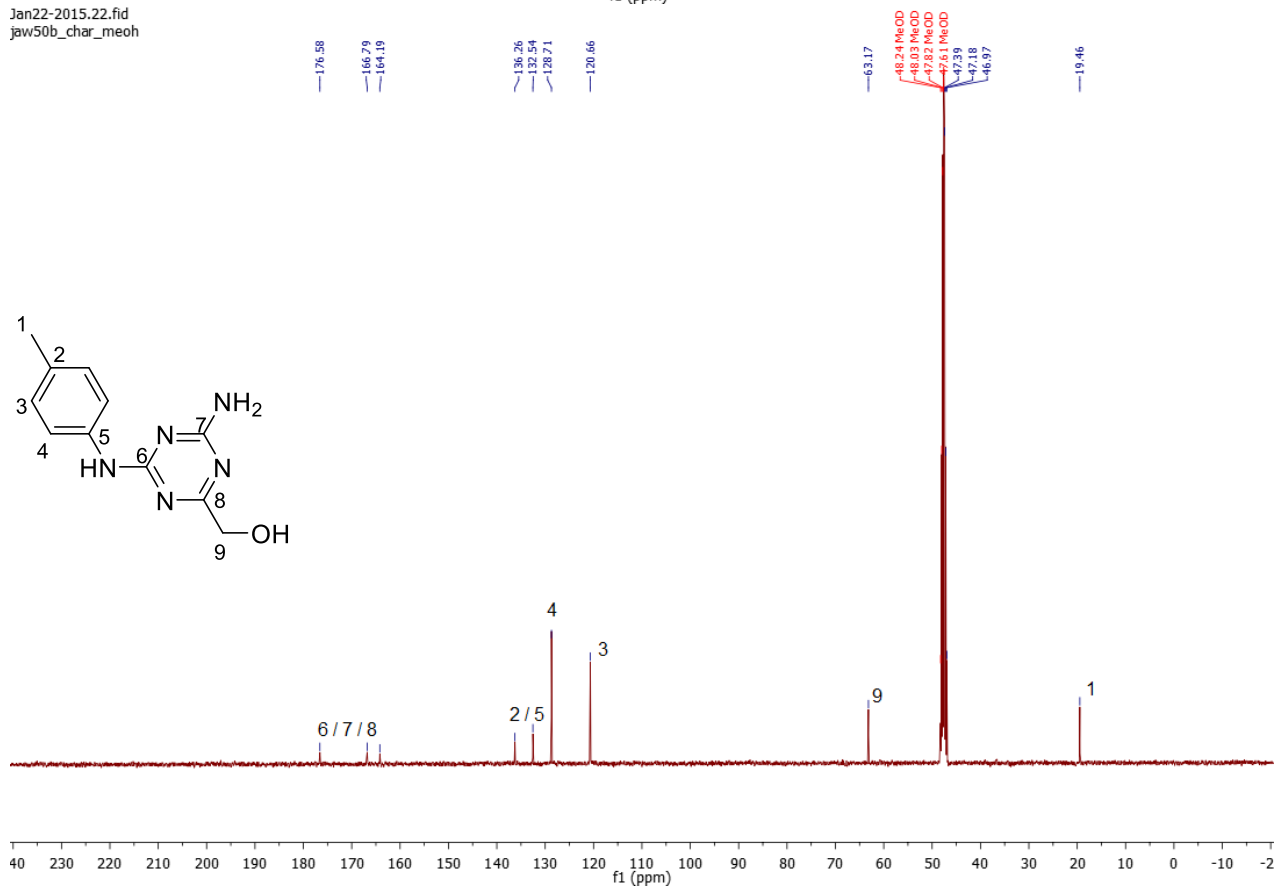


# NMR Spectra: Compound 10

Jan12-2015.10.fid  
jaw50b\_post column

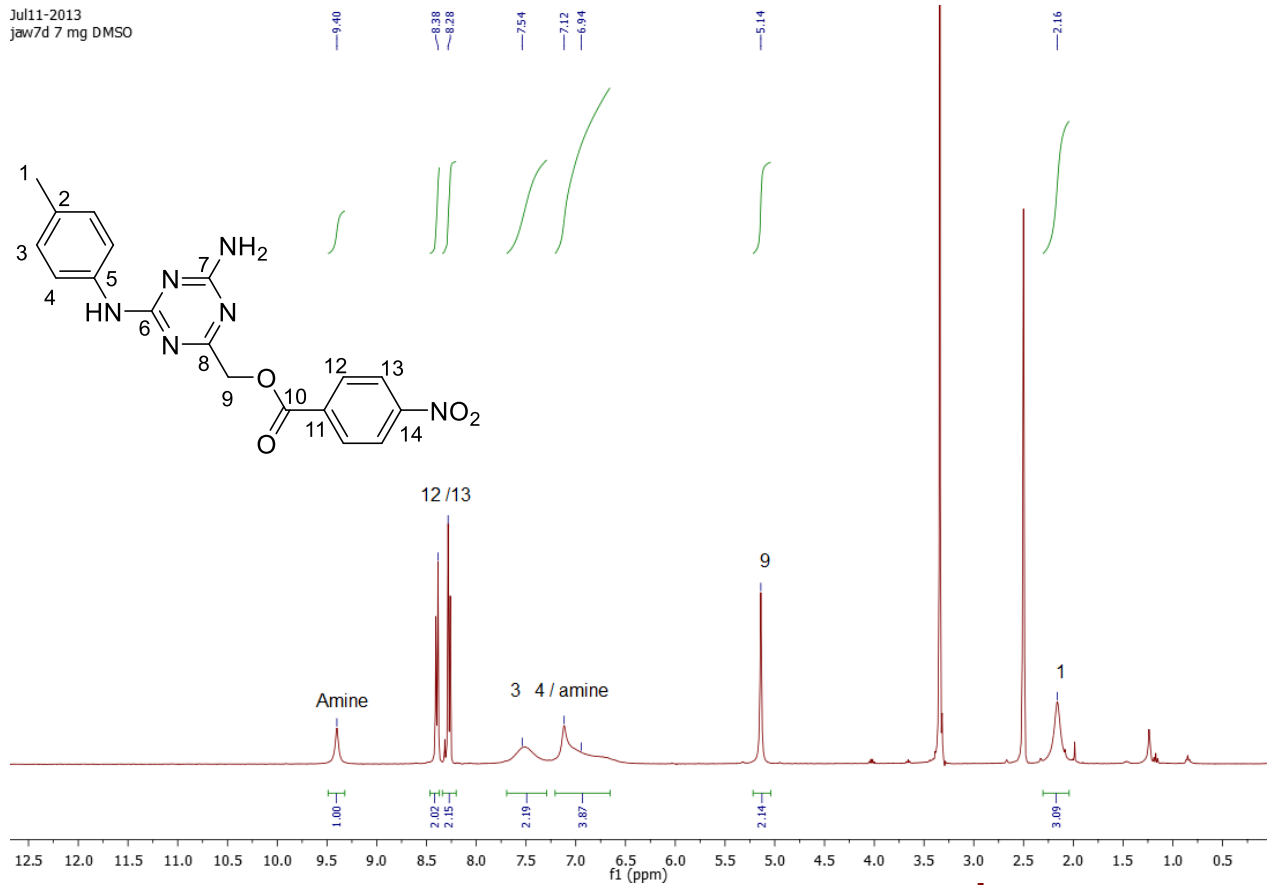


Jan22-2015.22.fid  
jaw50b\_char\_meoh

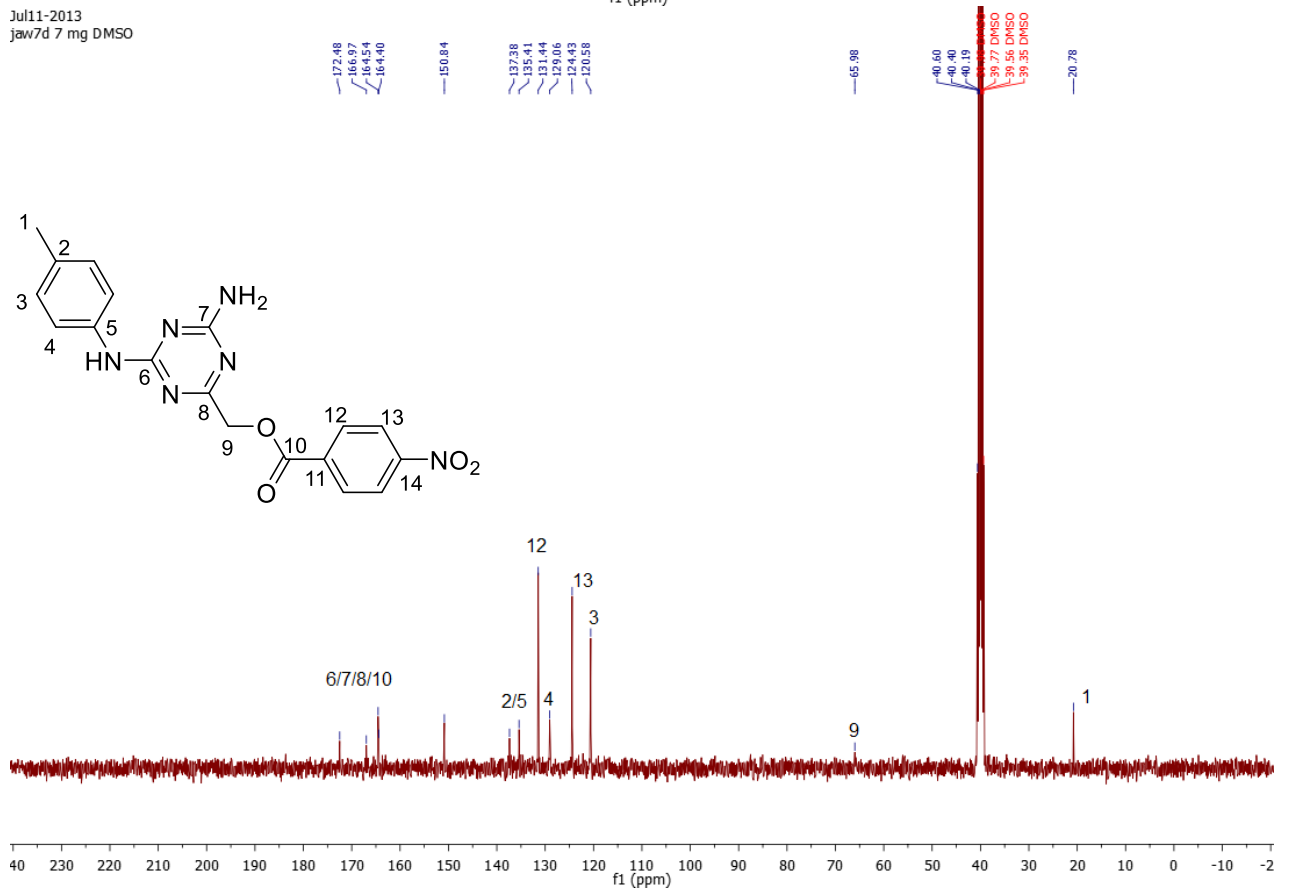


# NMR Spectra: Tz9

Jul11-2013  
jaw7d 7 mg DMSO

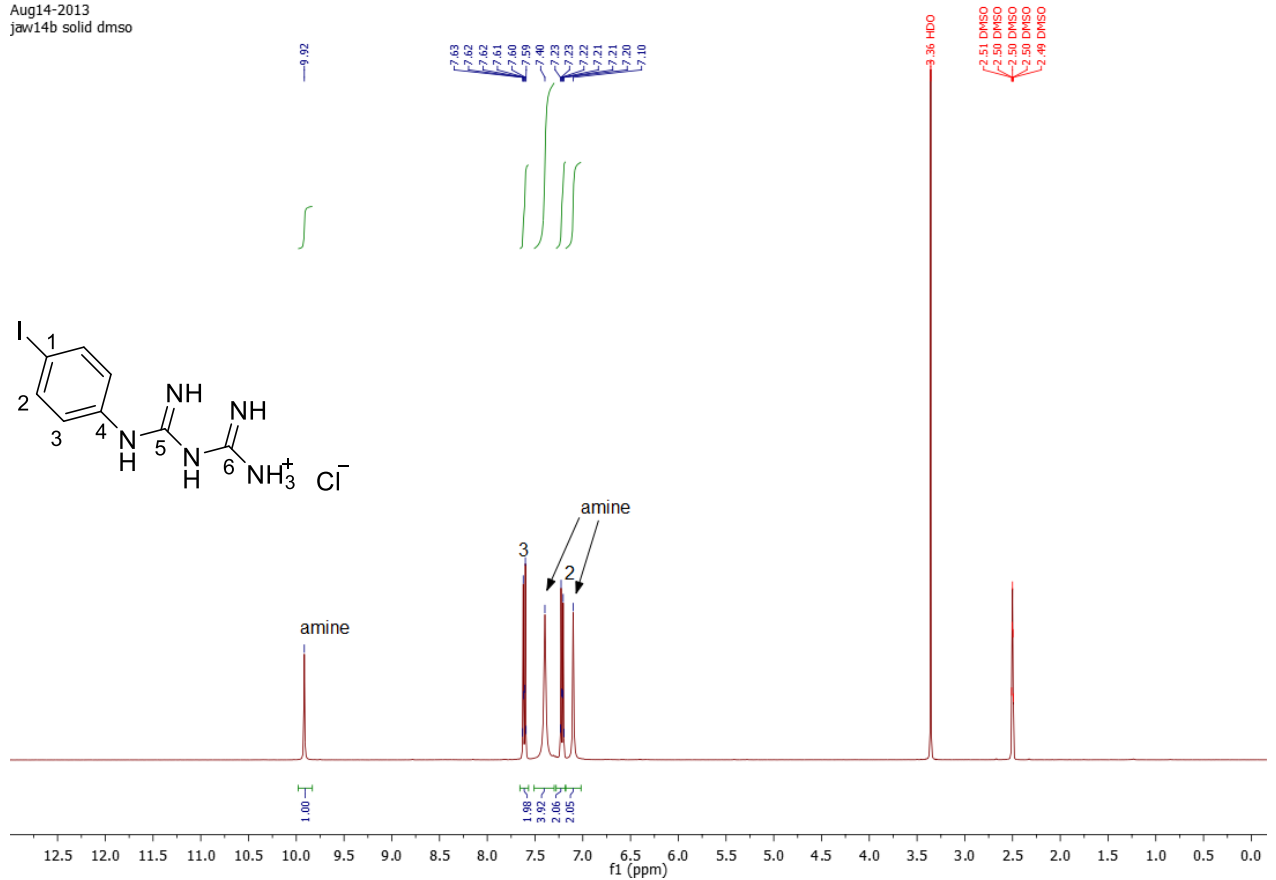


Jul11-2013  
jaw7d 7 mg DMSO

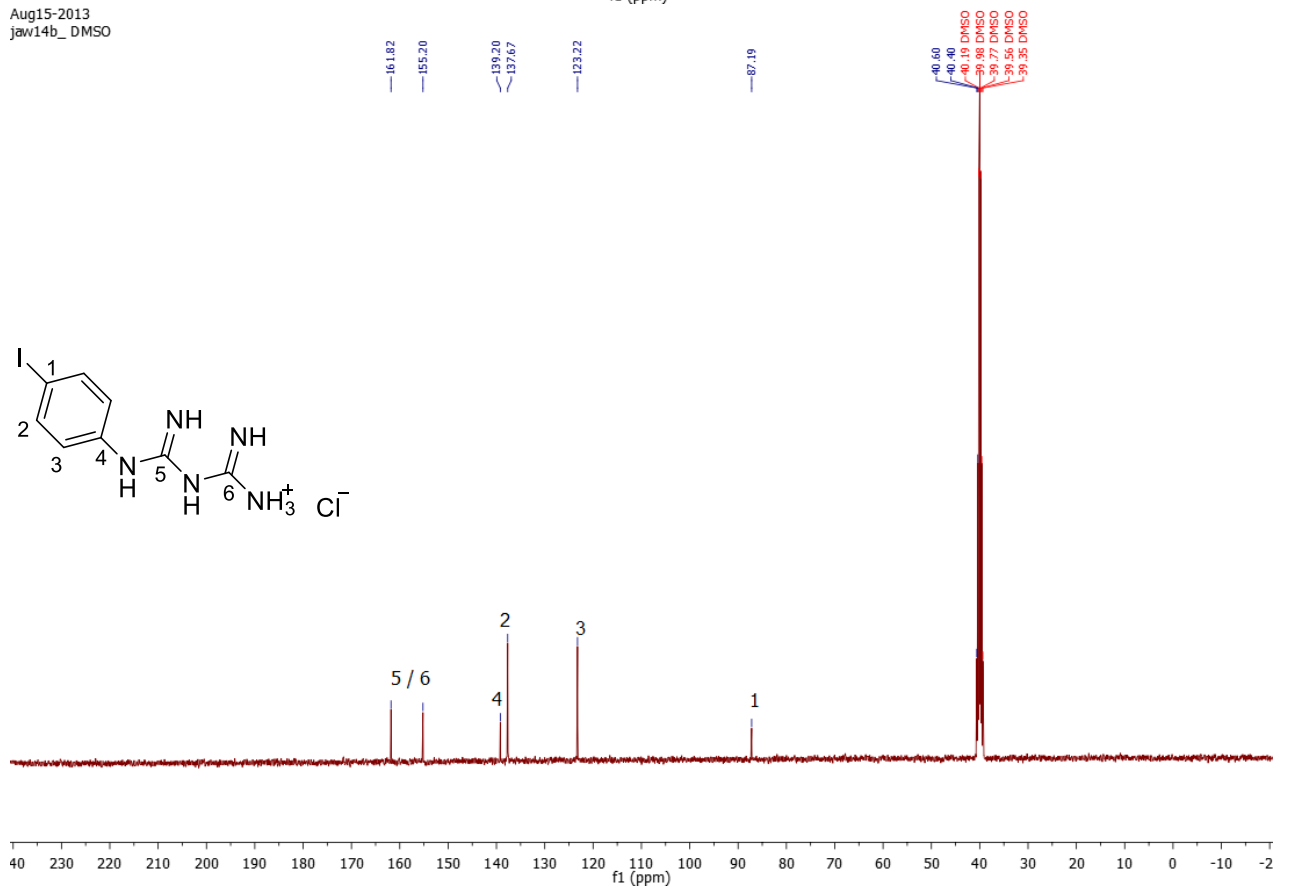


# NMR Spectra: Compound 20

Aug14-2013  
jaw14b solid dmso

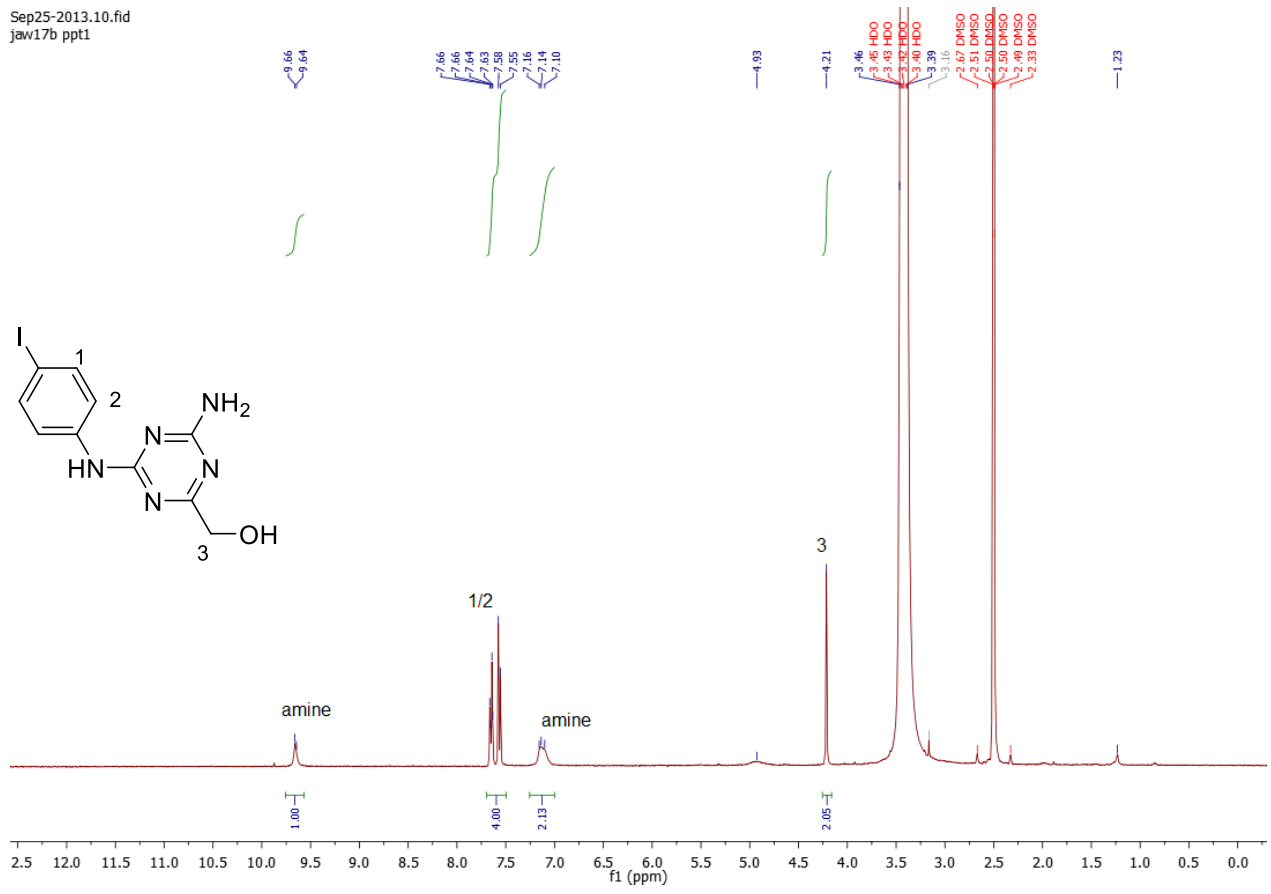


Aug15-2013  
jaw14b\_DMSO



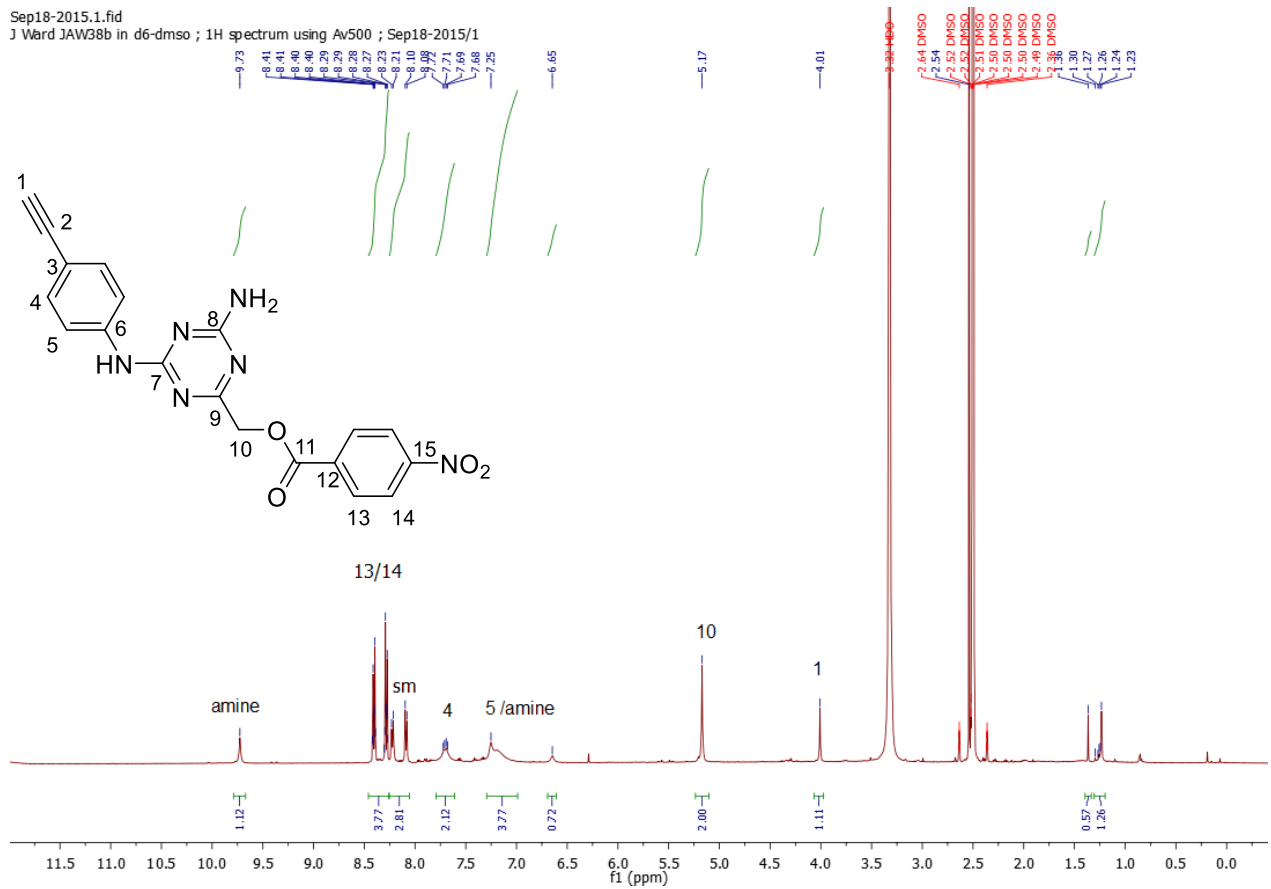
# NMR Spectra: Compound 21

Sep25-2013.10.fid  
jaw17b ppt1



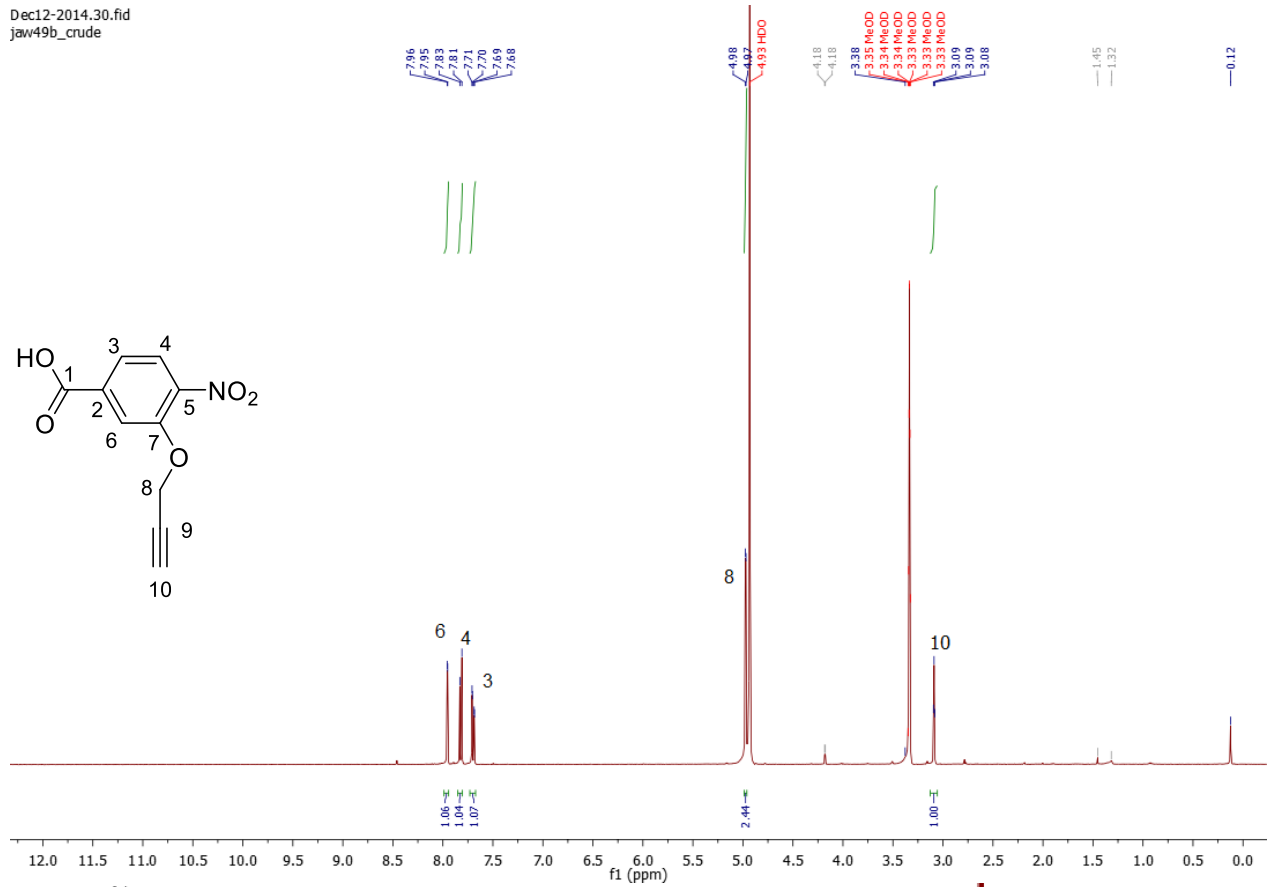
# NMR Spectra: T1

Sep18-2015.1.fid  
 J Ward JAW38b in d6-dms0 ; 1H spectrum using Av500 ; Sep18-2015/1

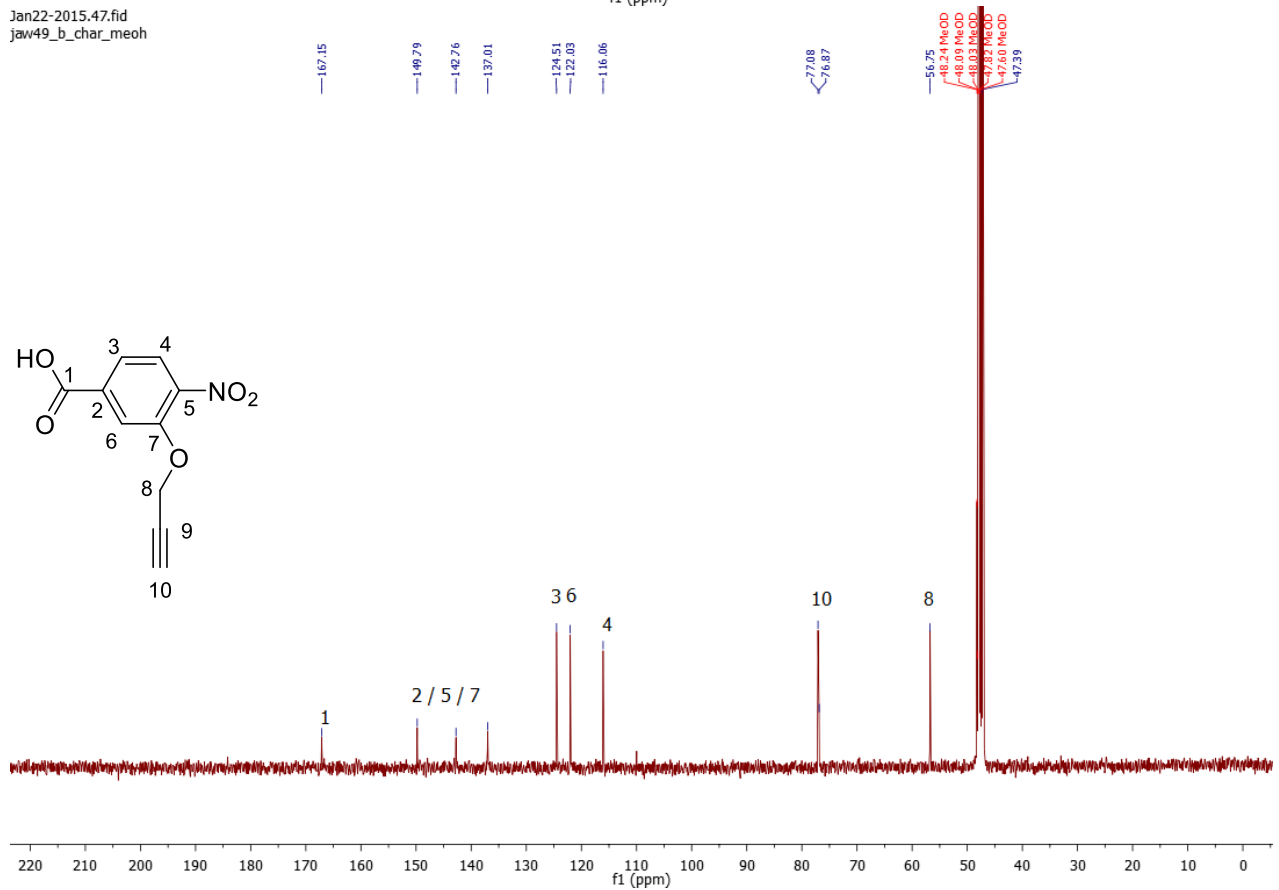


# NMR Spectra: Compound 12

Dec12-2014.30.fid  
jaw49b\_crude

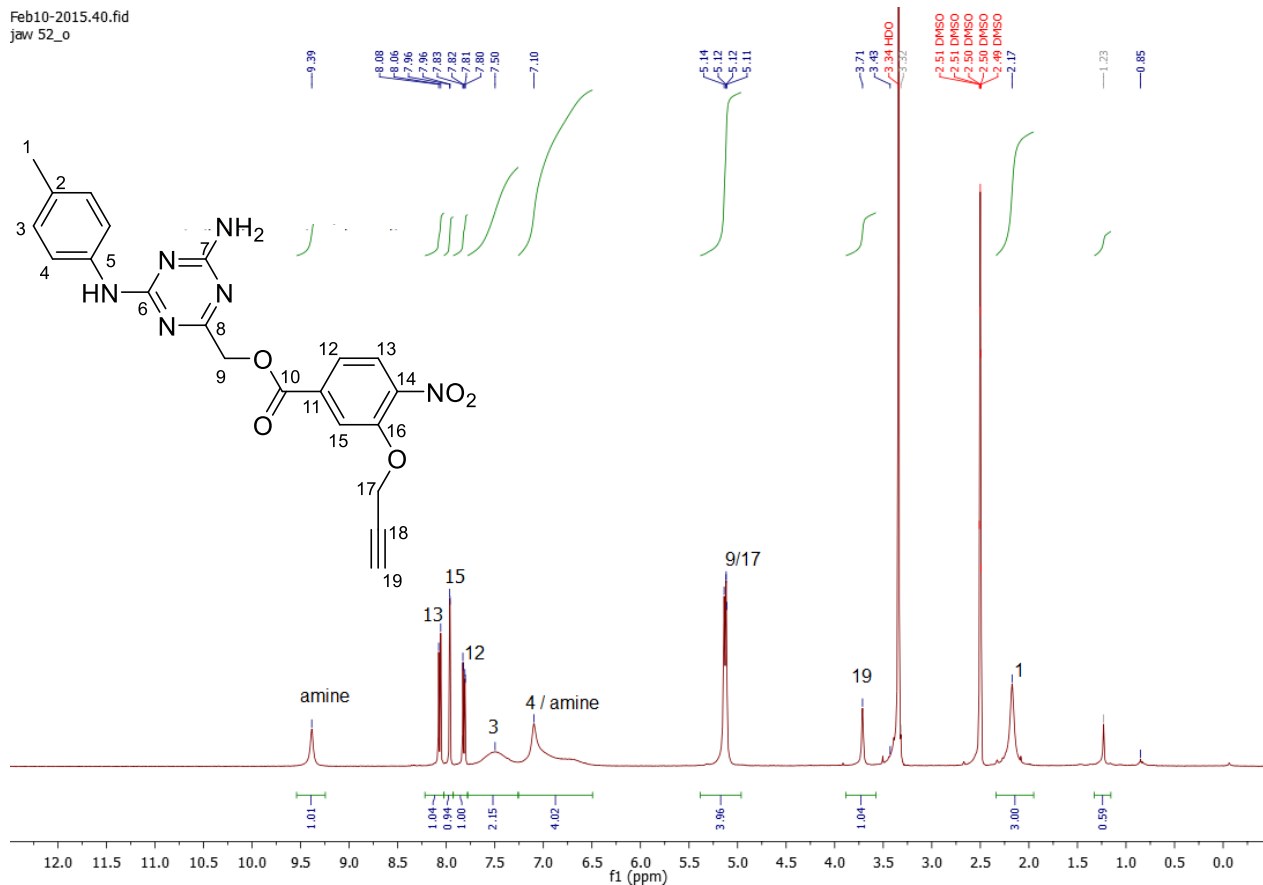


Jan22-2015.47.fid  
jaw49\_b\_char\_meoh

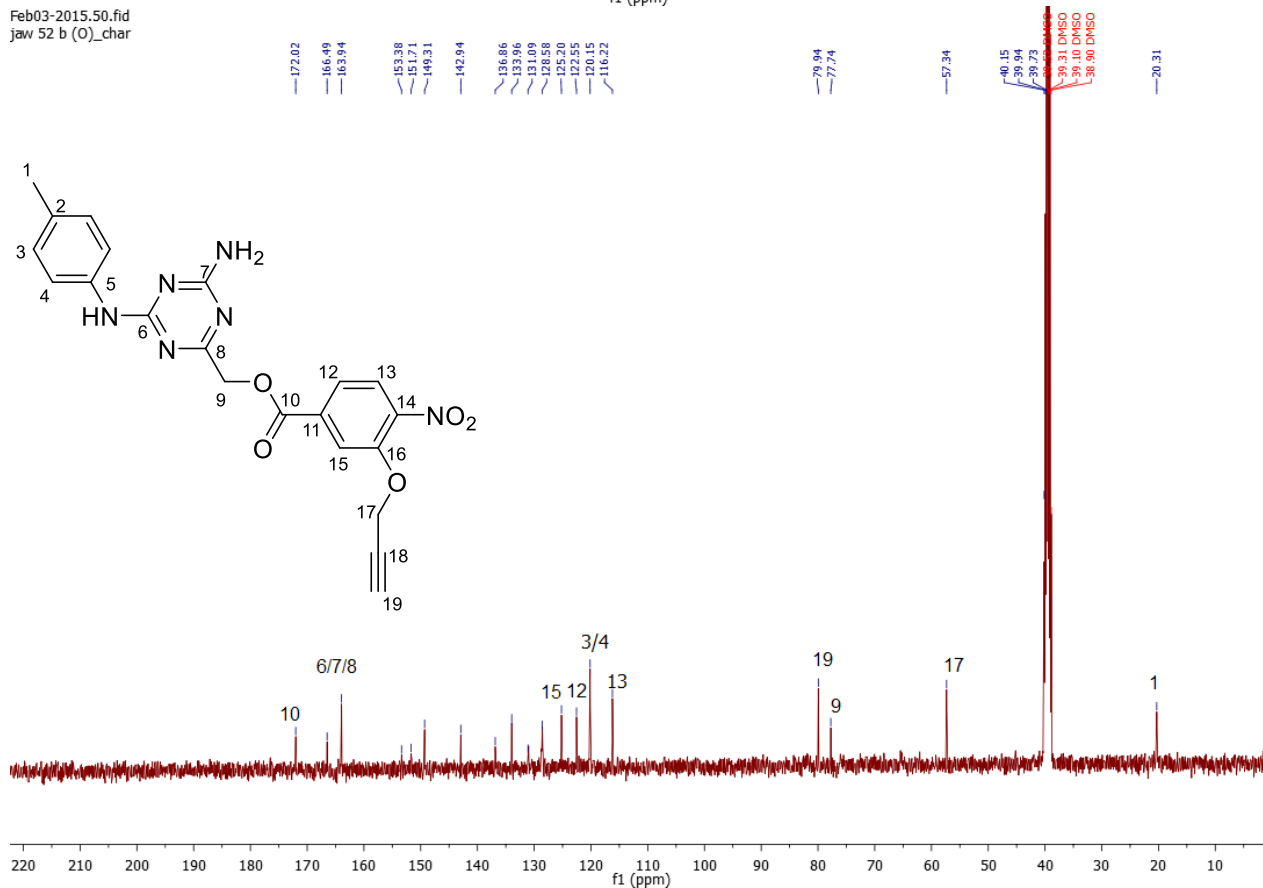


# NMR Spectra: T2

Feb10-2015.40.fid  
jaw 52\_o



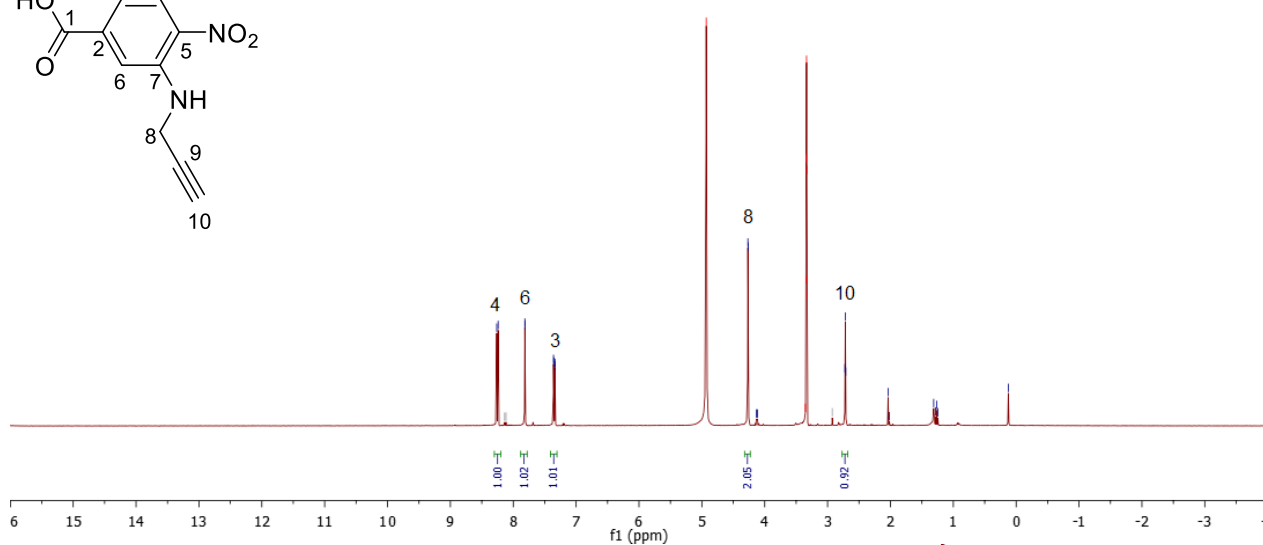
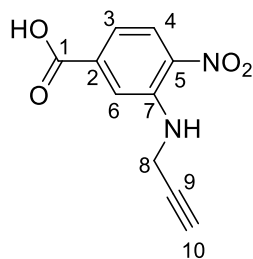
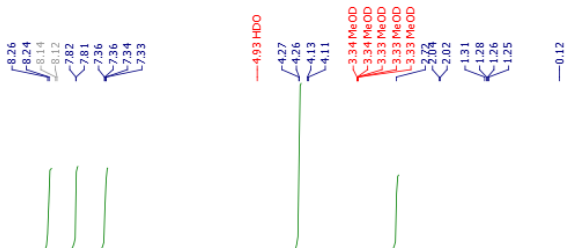
Feb03-2015.50.fid  
jaw 52 b (O)\_char



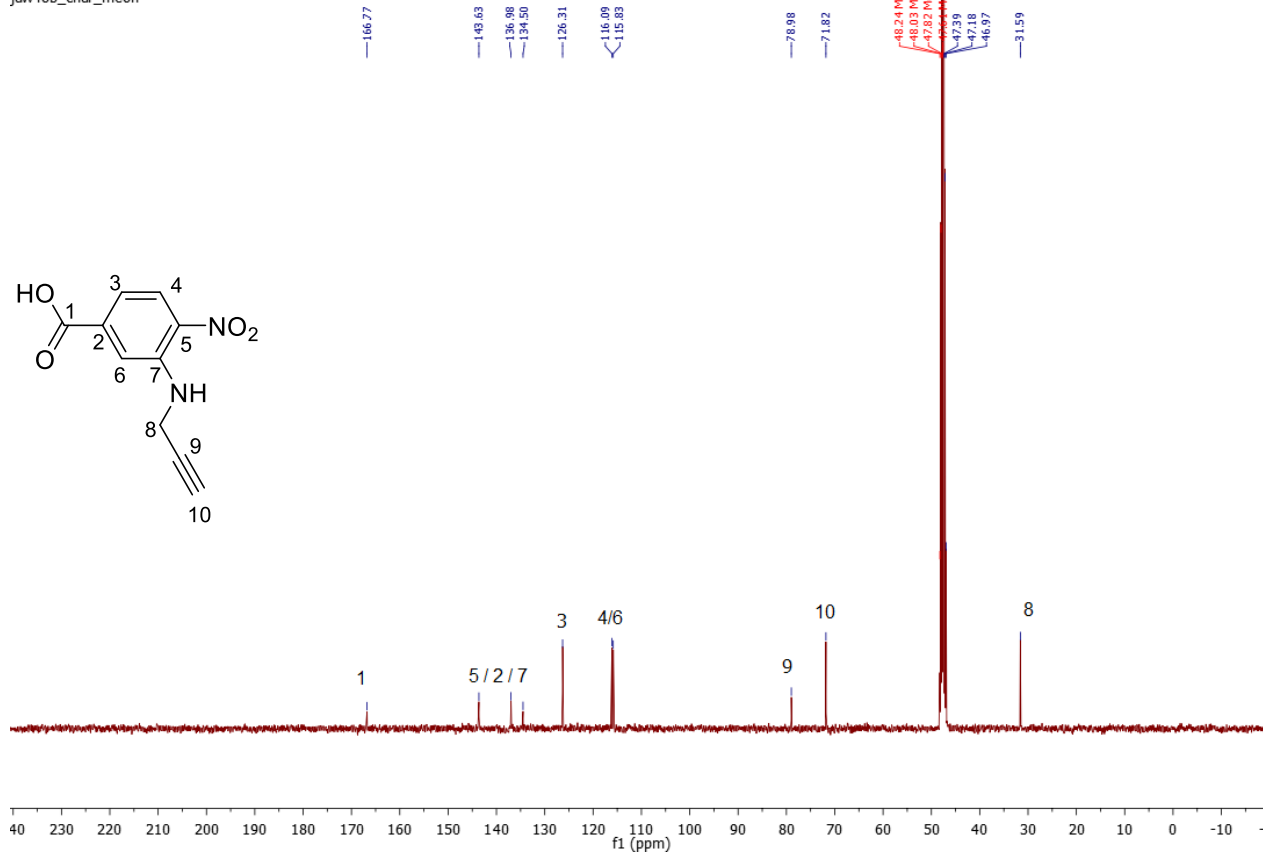
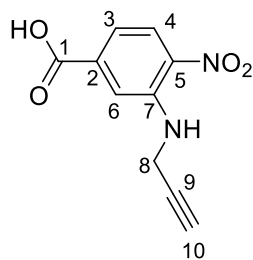


# NMR Spectra: Compound 13

Dec11-2014.10.fid  
jaw48b



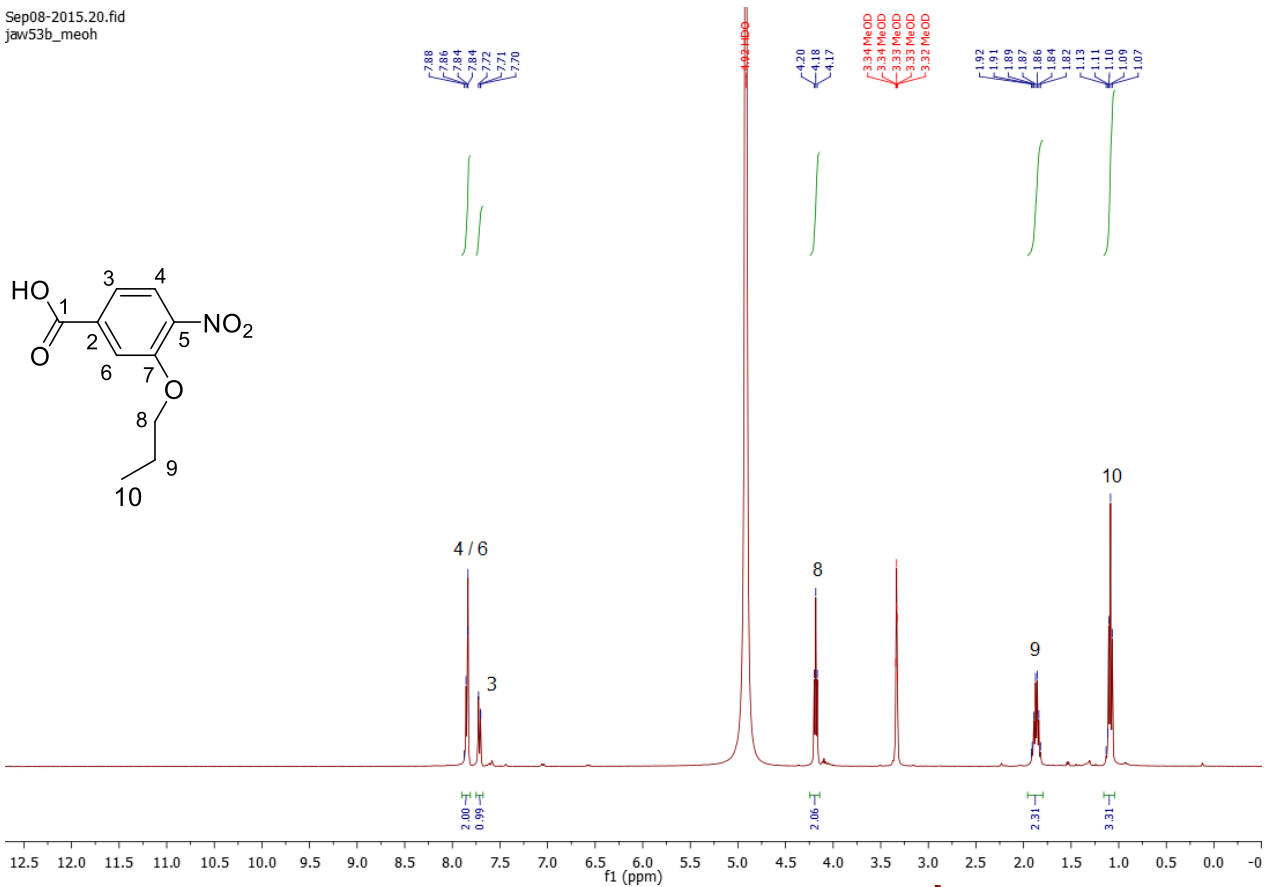
Jan22-2015.30.fid  
jaw48b\_char\_meoh



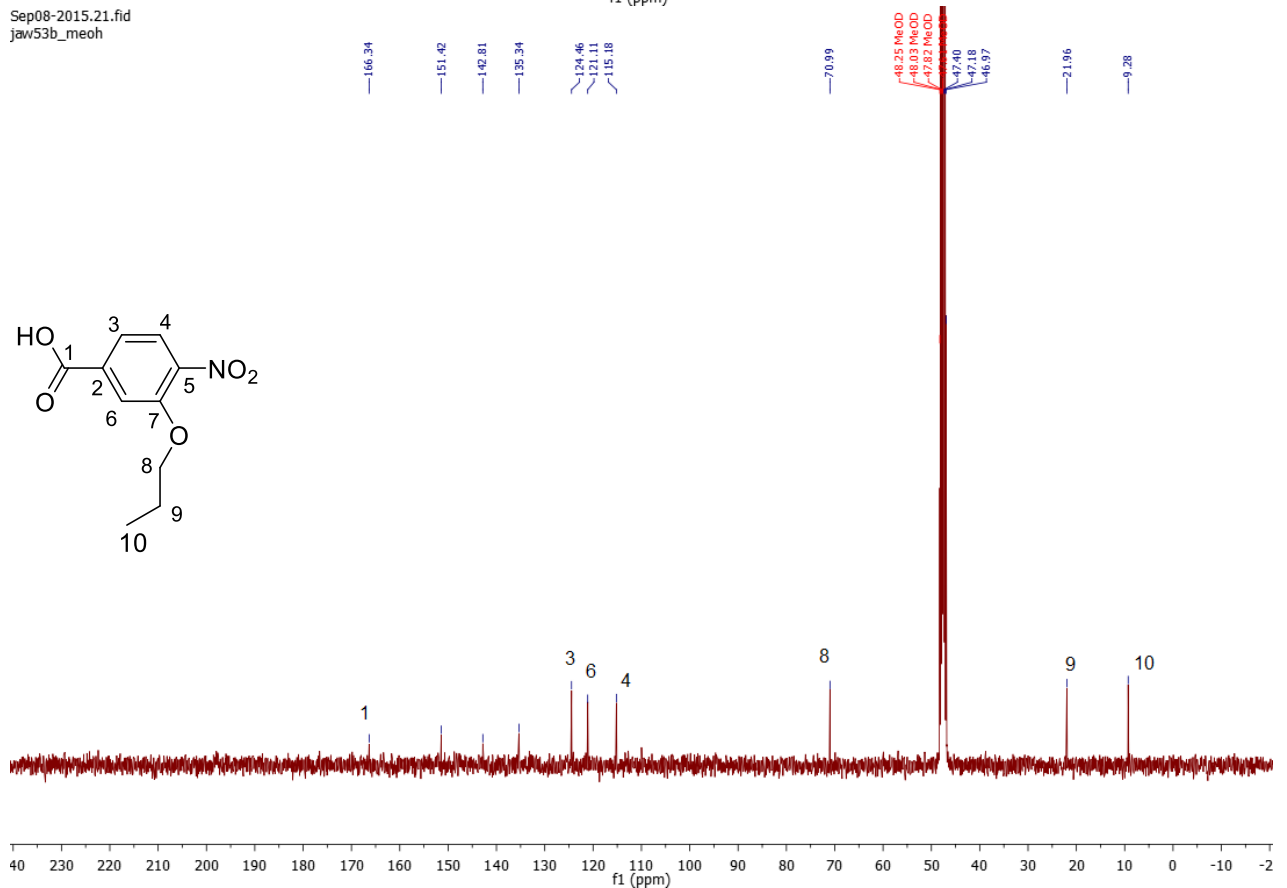


# NMR Spectra: Compound 23

Sep08-2015.20.fid  
jaw53b\_meoh

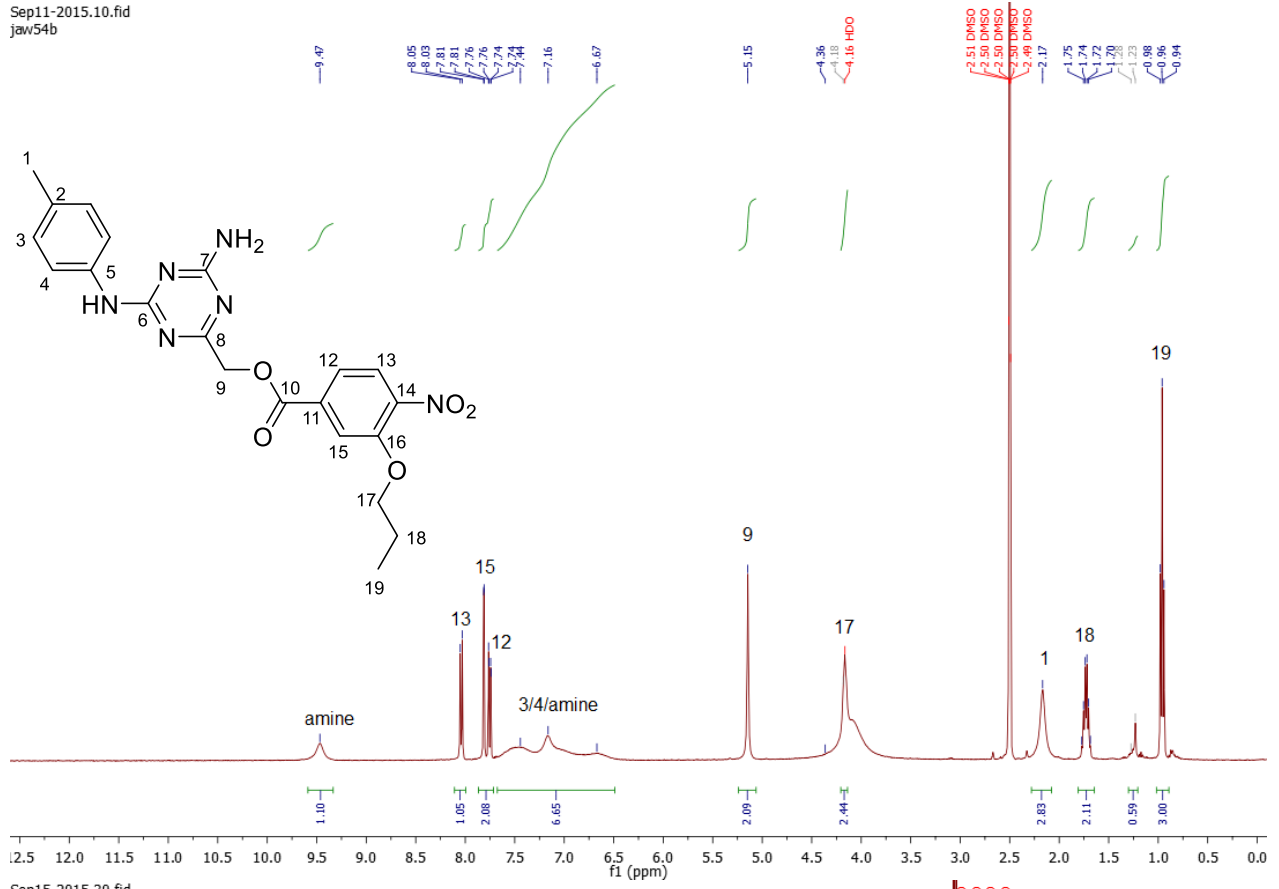


Sep08-2015.21.fid  
jaw53b\_meoh

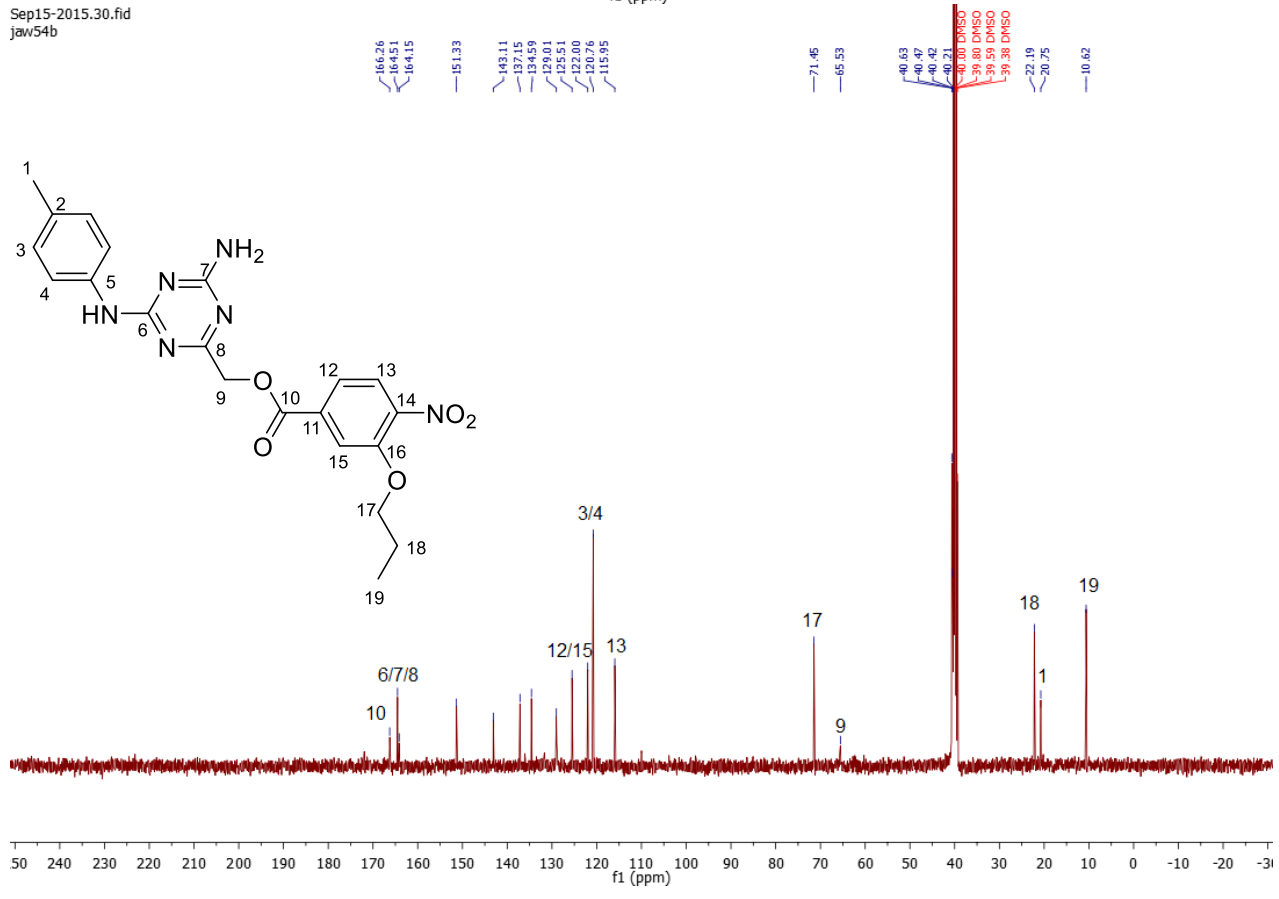


# NMR Spectra: T2c

Sep11-2015.10.fid  
jaw54b

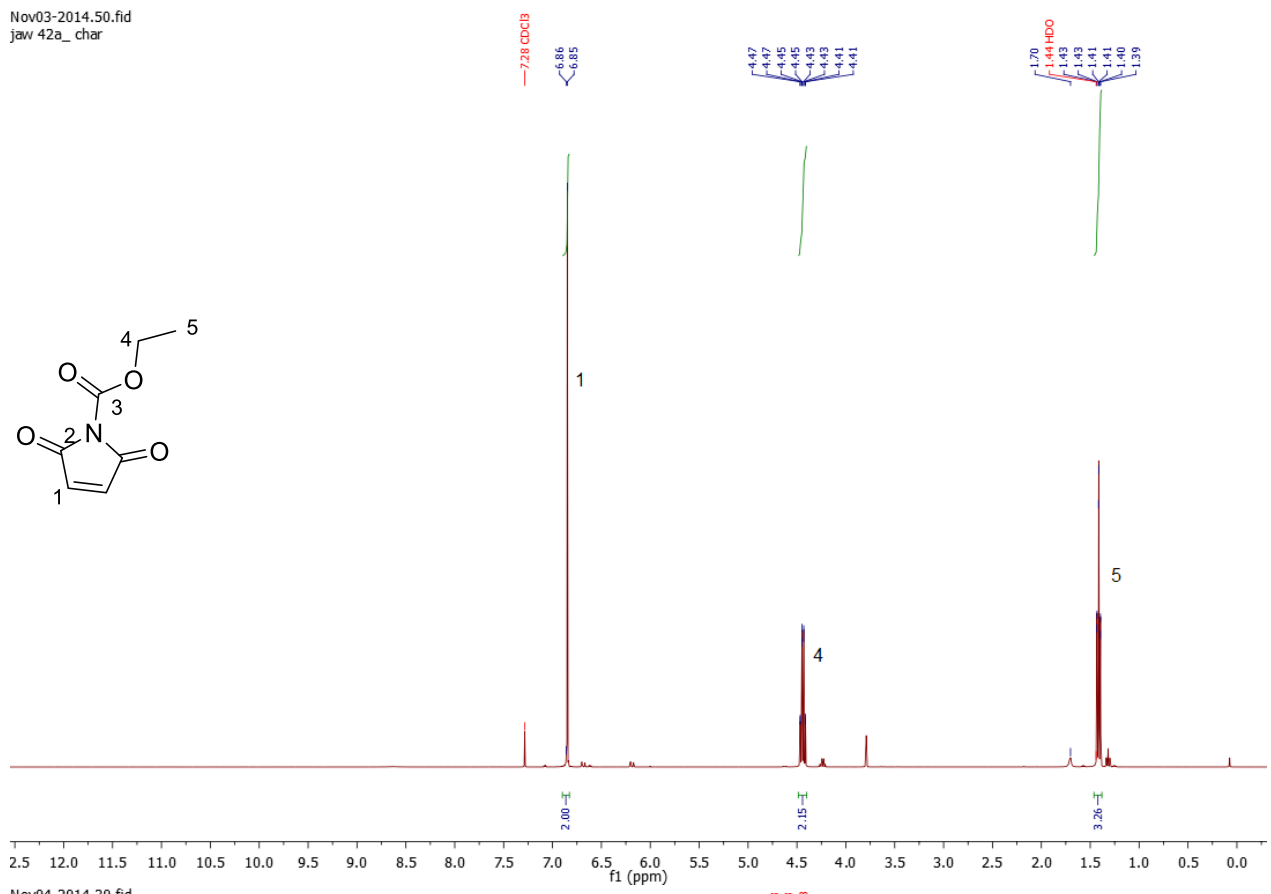


Sep15-2015.30.fid  
jaw54b

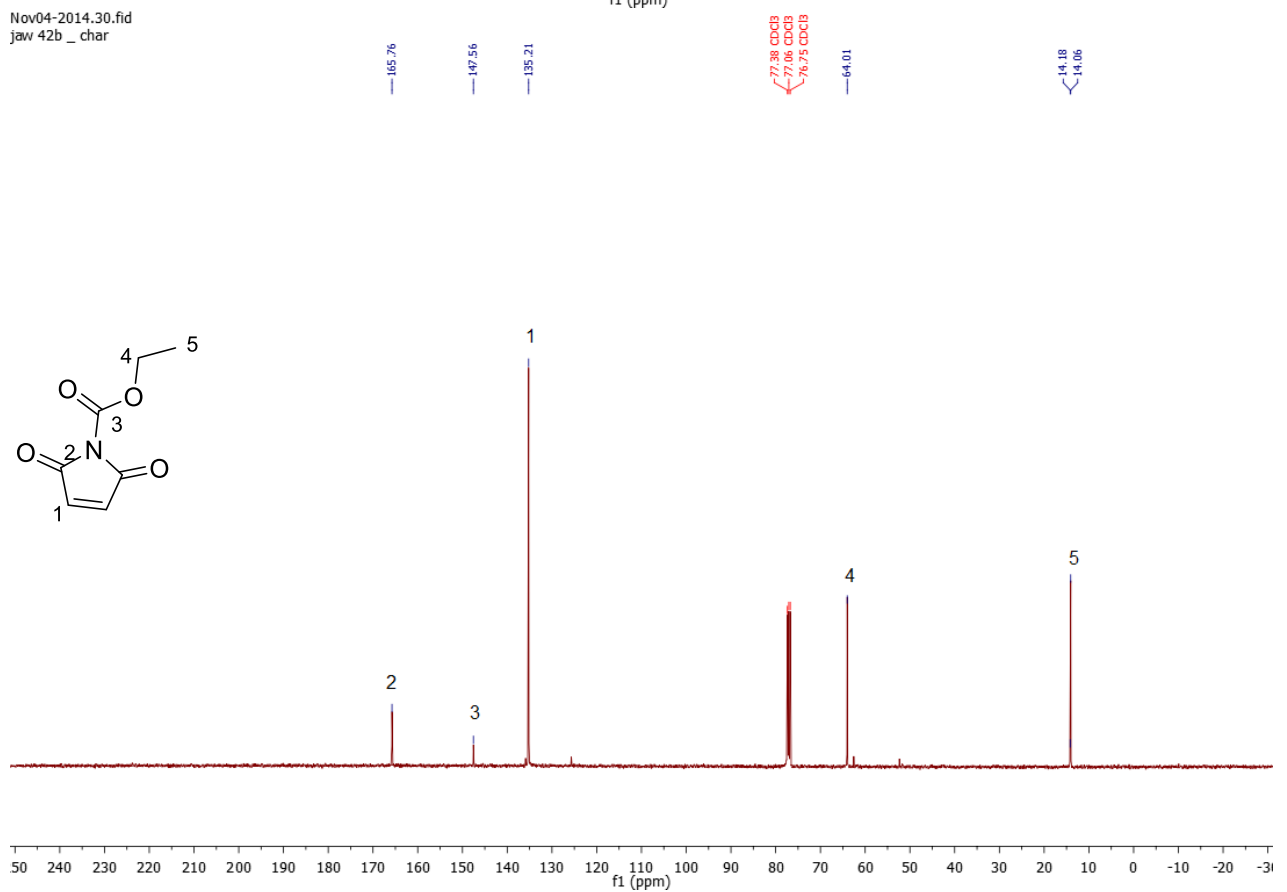


# NMR Spectra: Compound 29

Nov03-2014.50.fid  
jaw 42a\_char

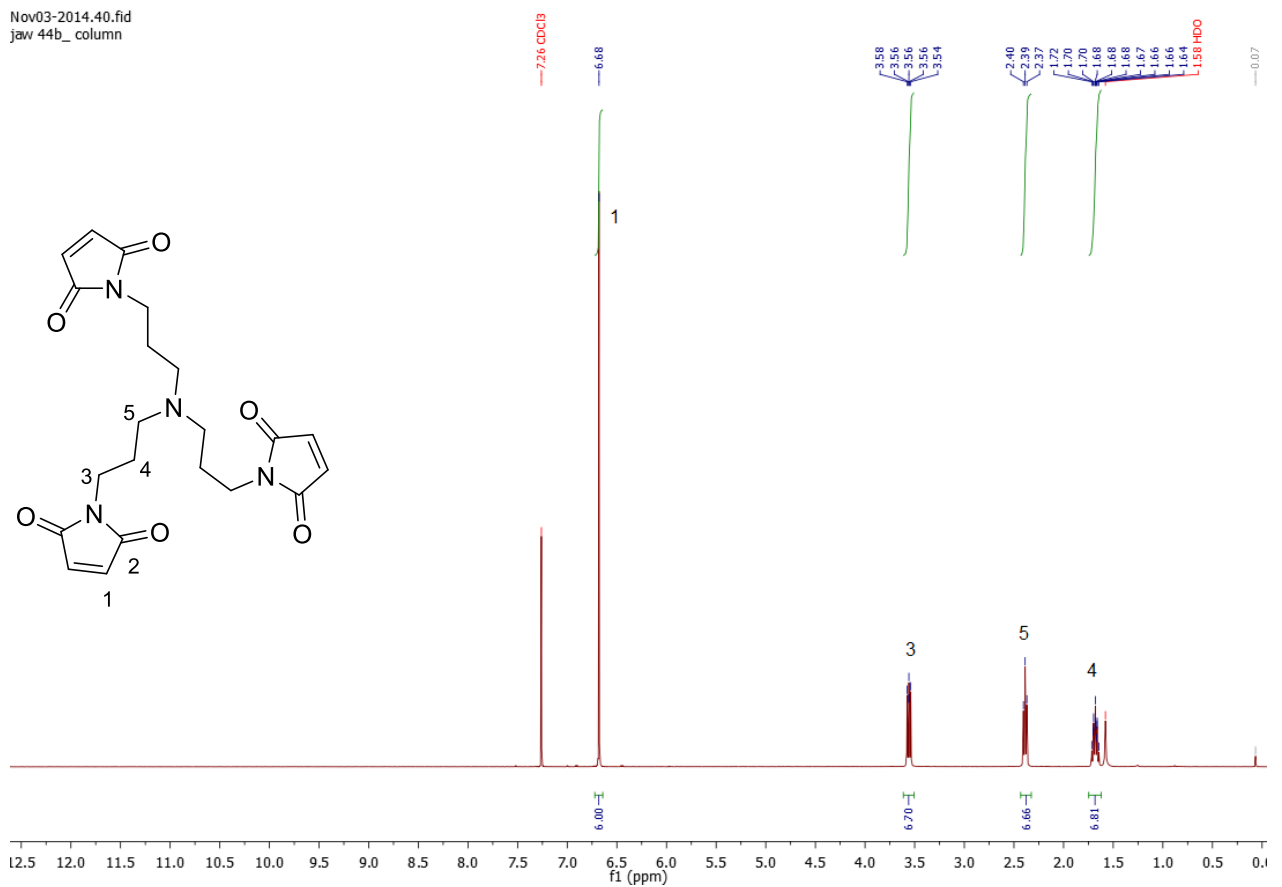


Nov04-2014.30.fid  
jaw 42b\_char

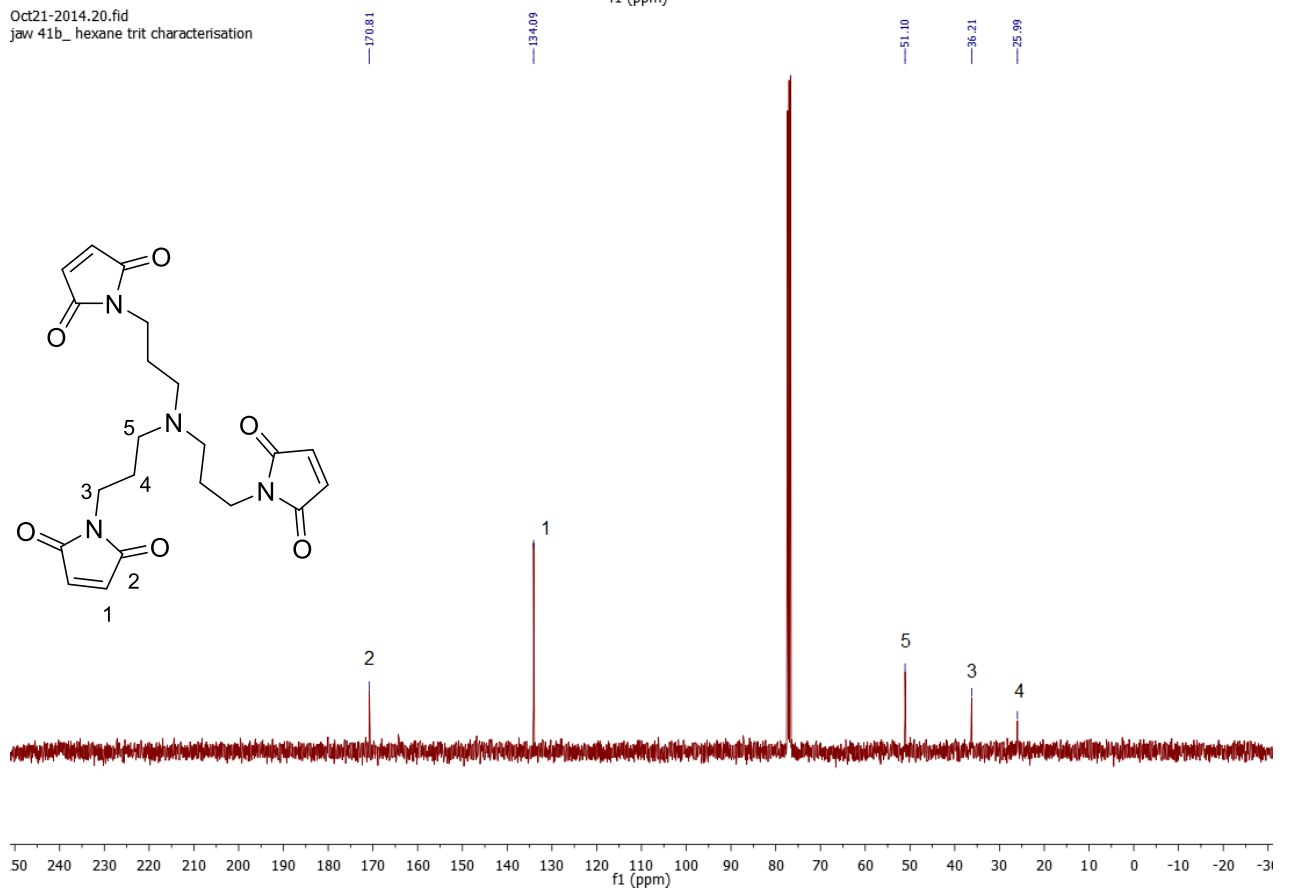


# NMR Spectra: Trap C

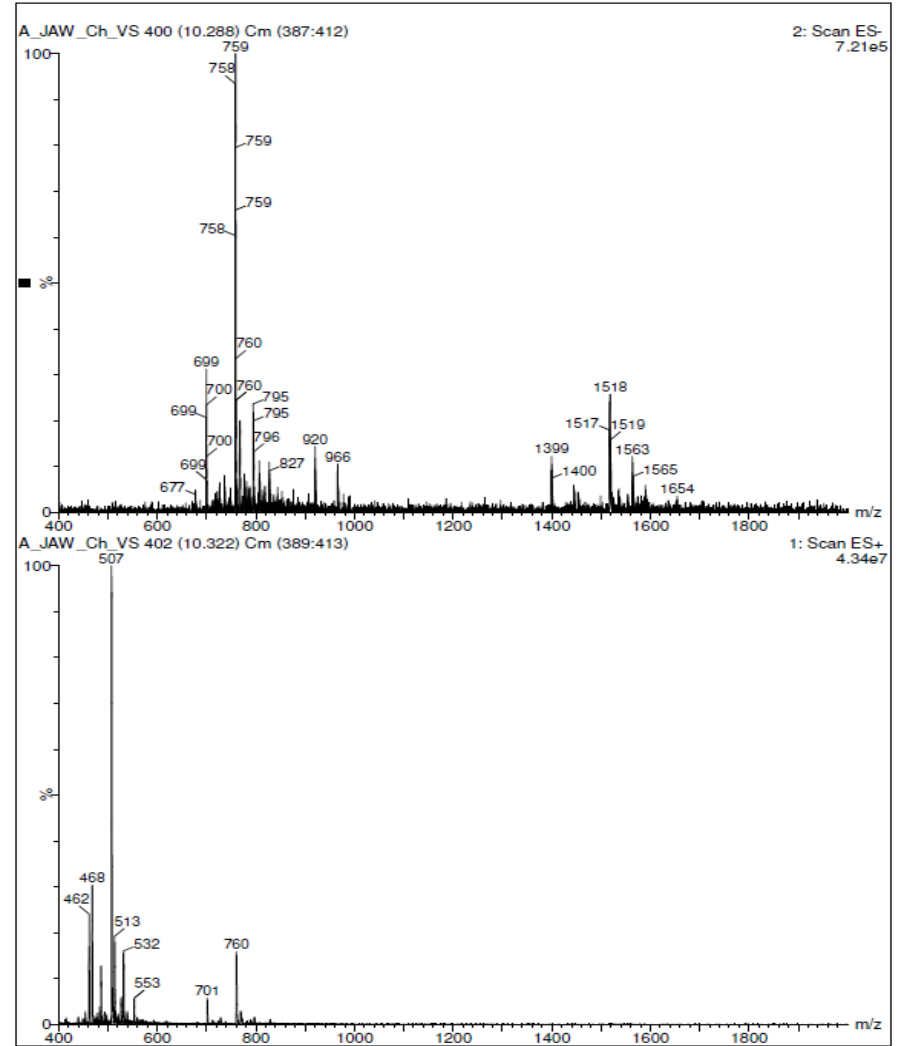
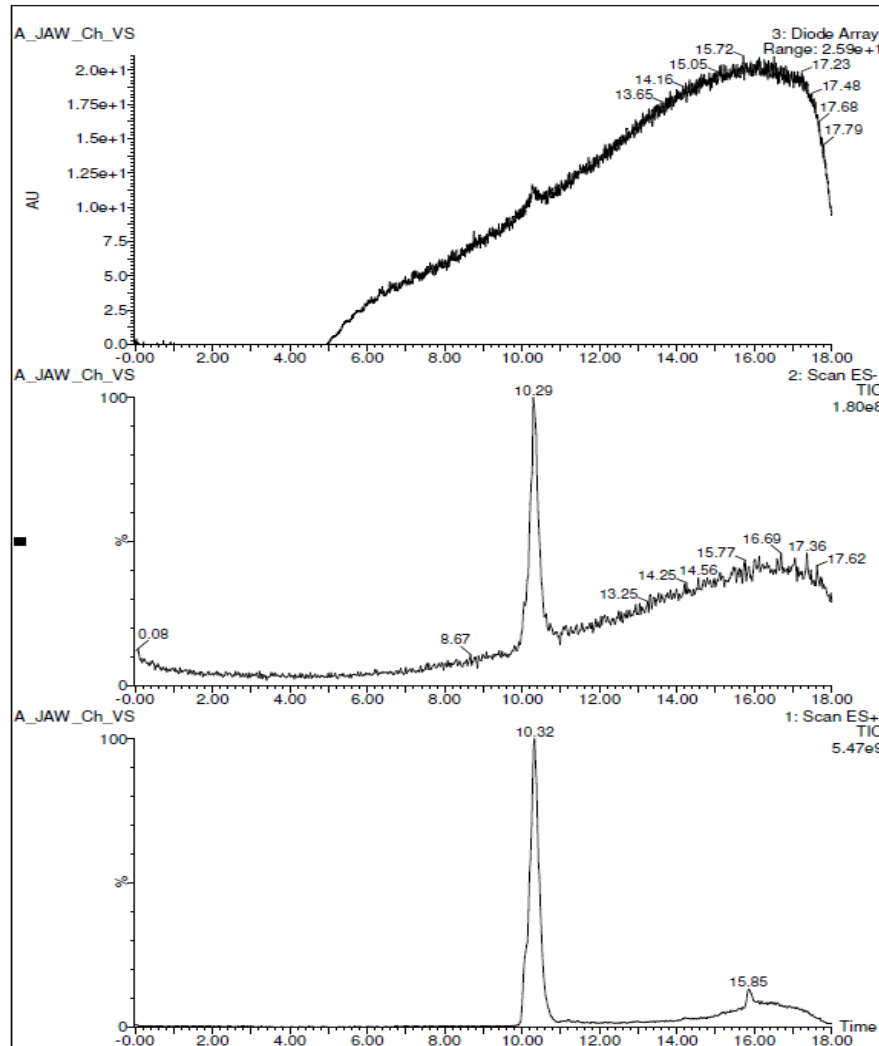
Nov03-2014.40.fid  
jaw 44b\_column



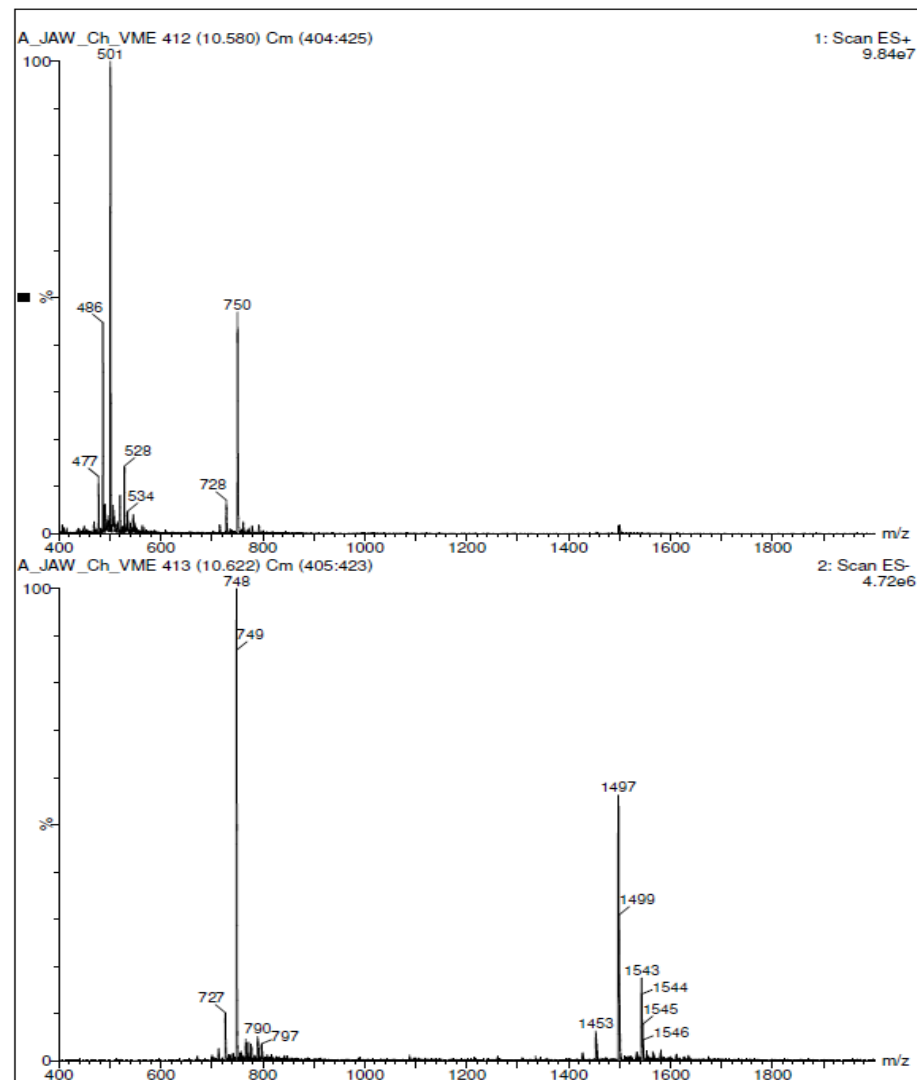
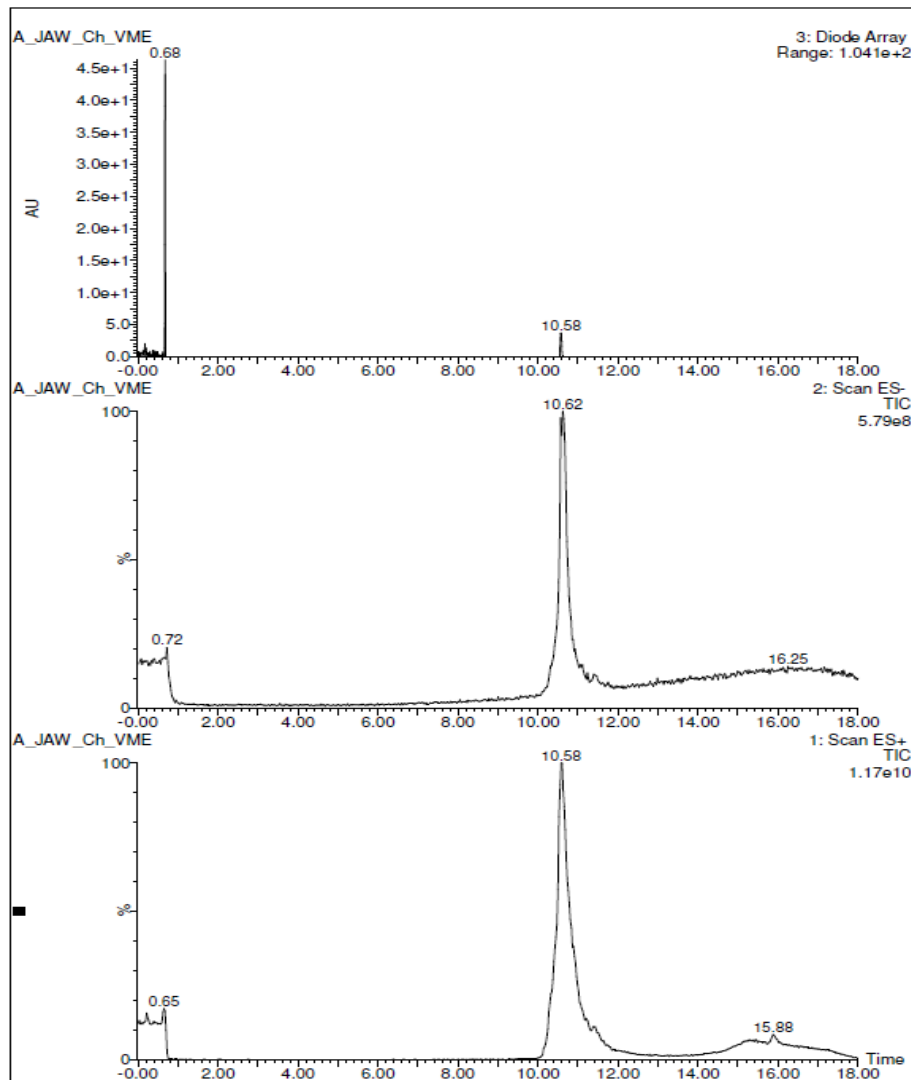
Oct21-2014.20.fid  
jaw 41b\_hexane trit characterisation



LCMS Chromatograms and Spectra: Pg-12-mer-VS

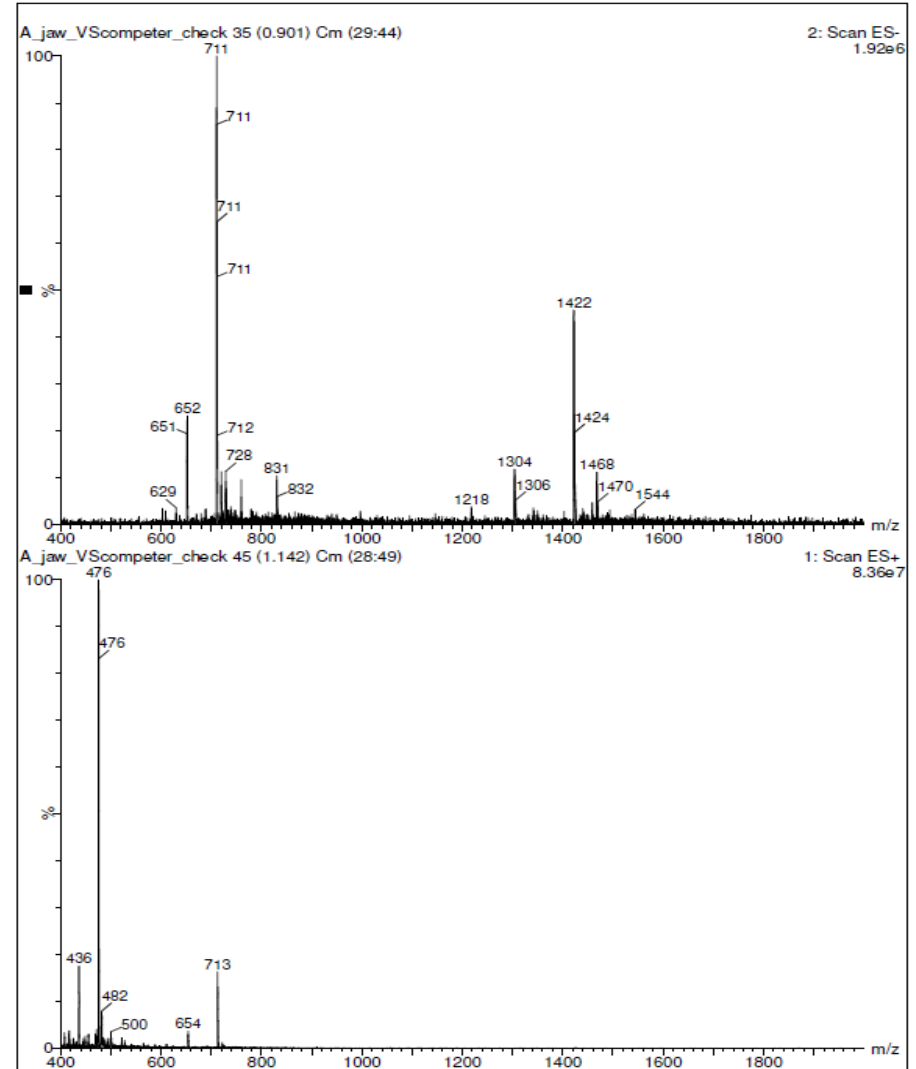
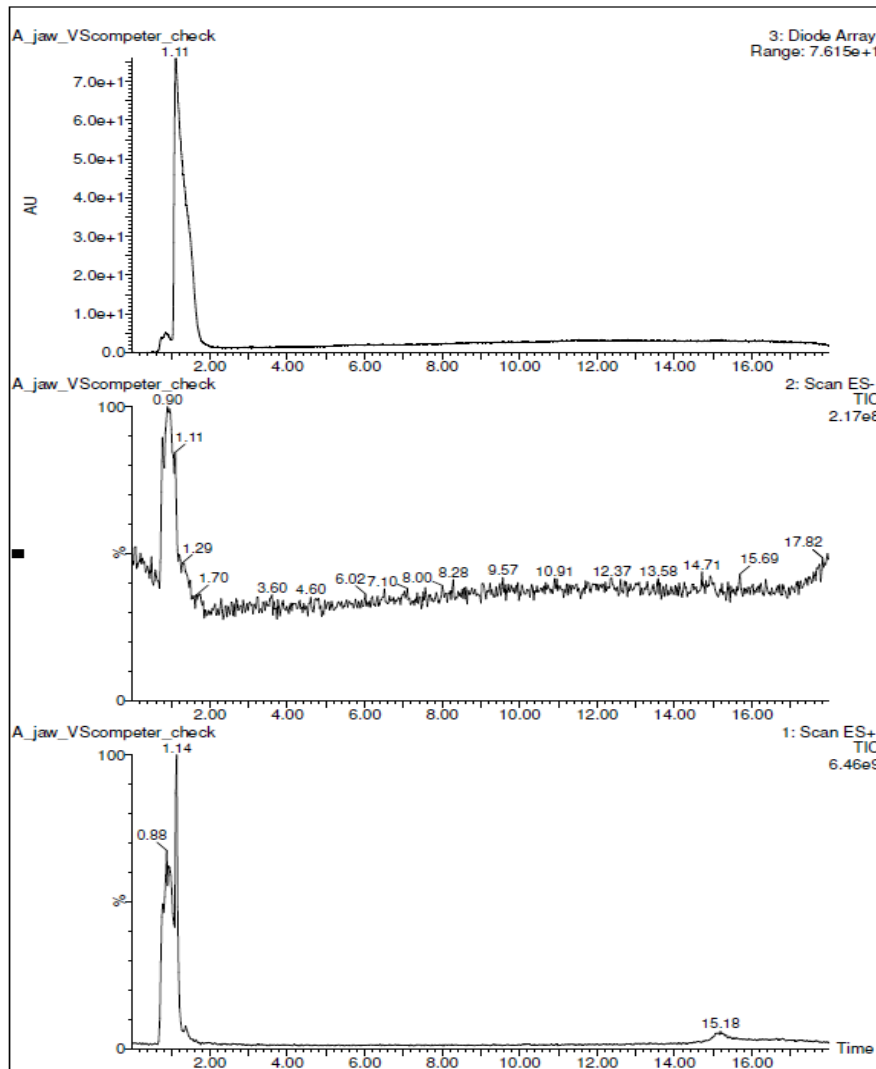


LCMS Chromatograms and Spectra: Pg-12-mer-VME

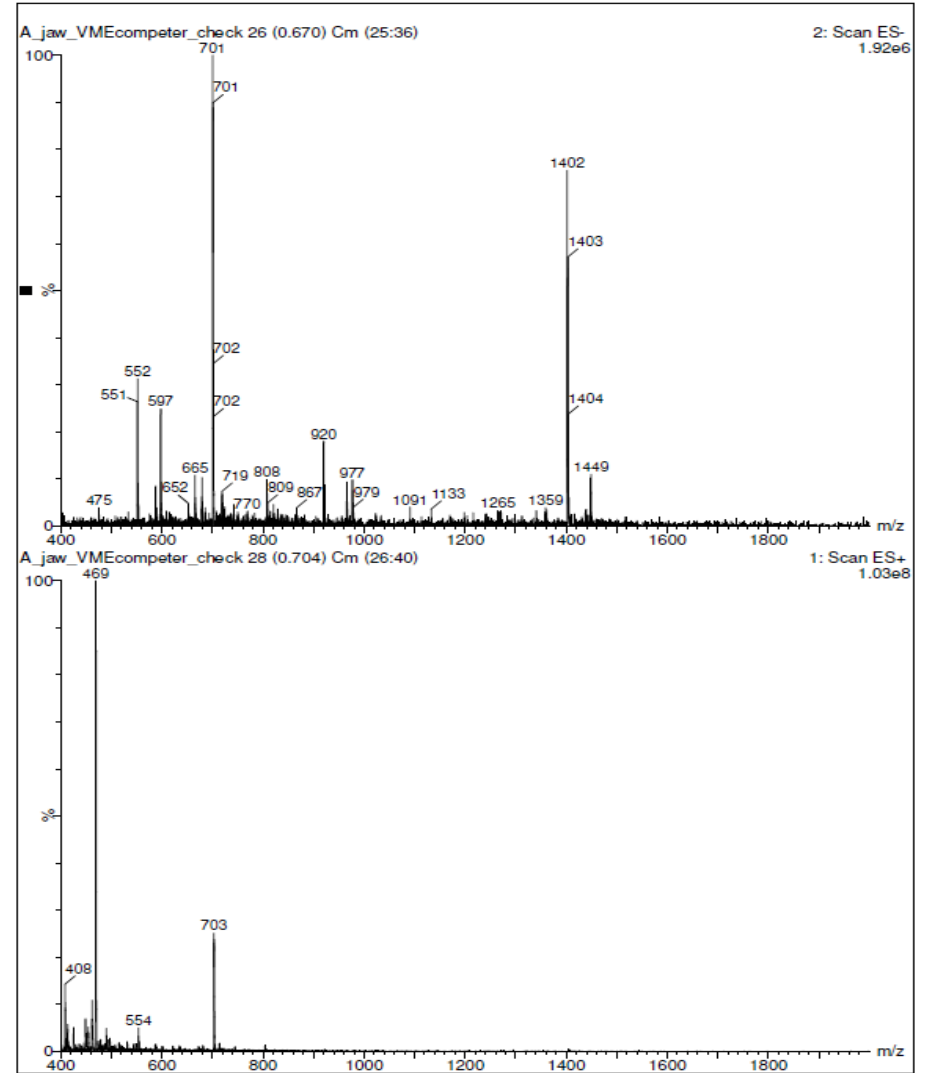
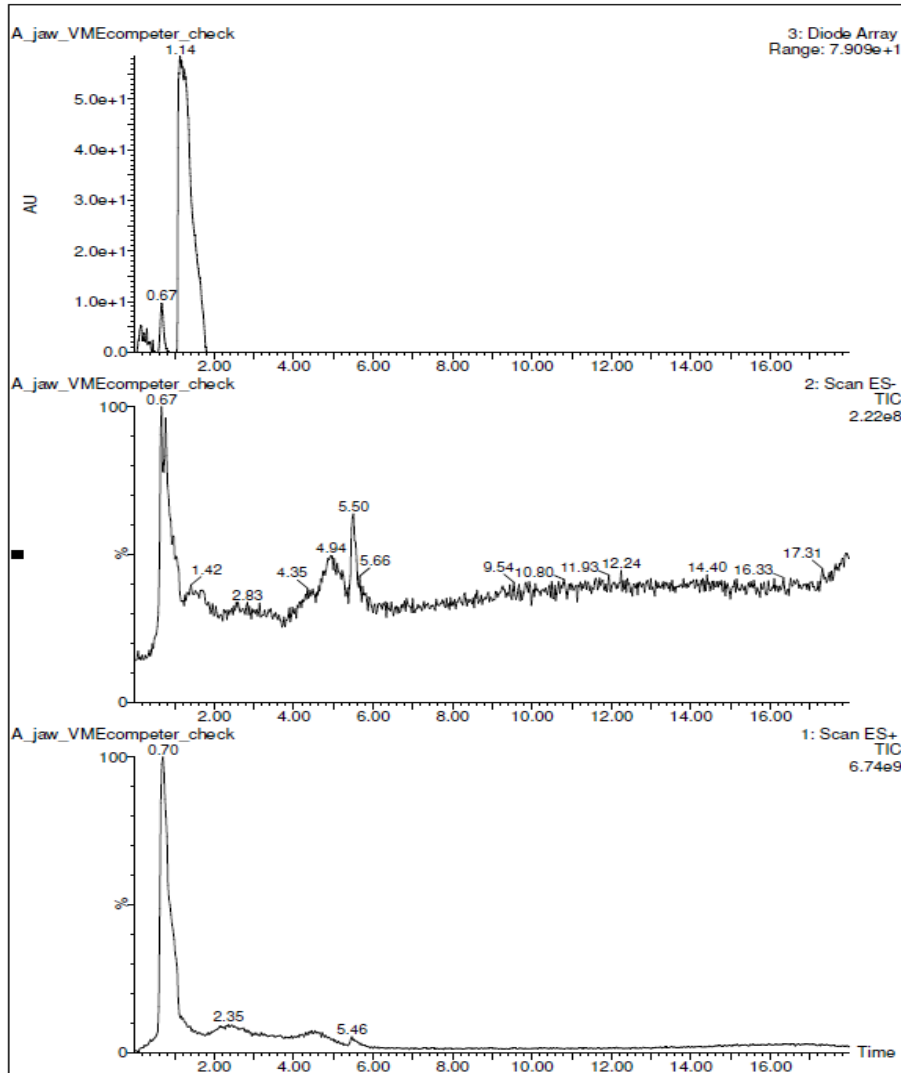




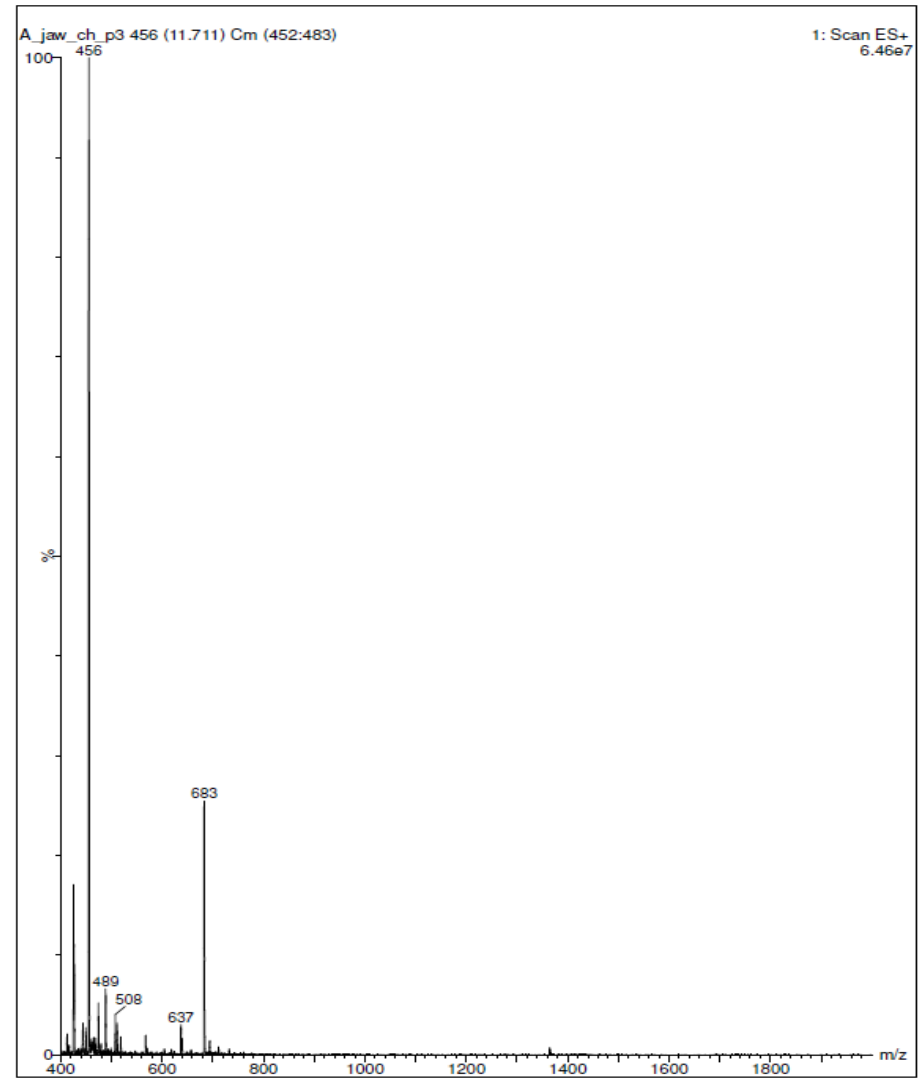
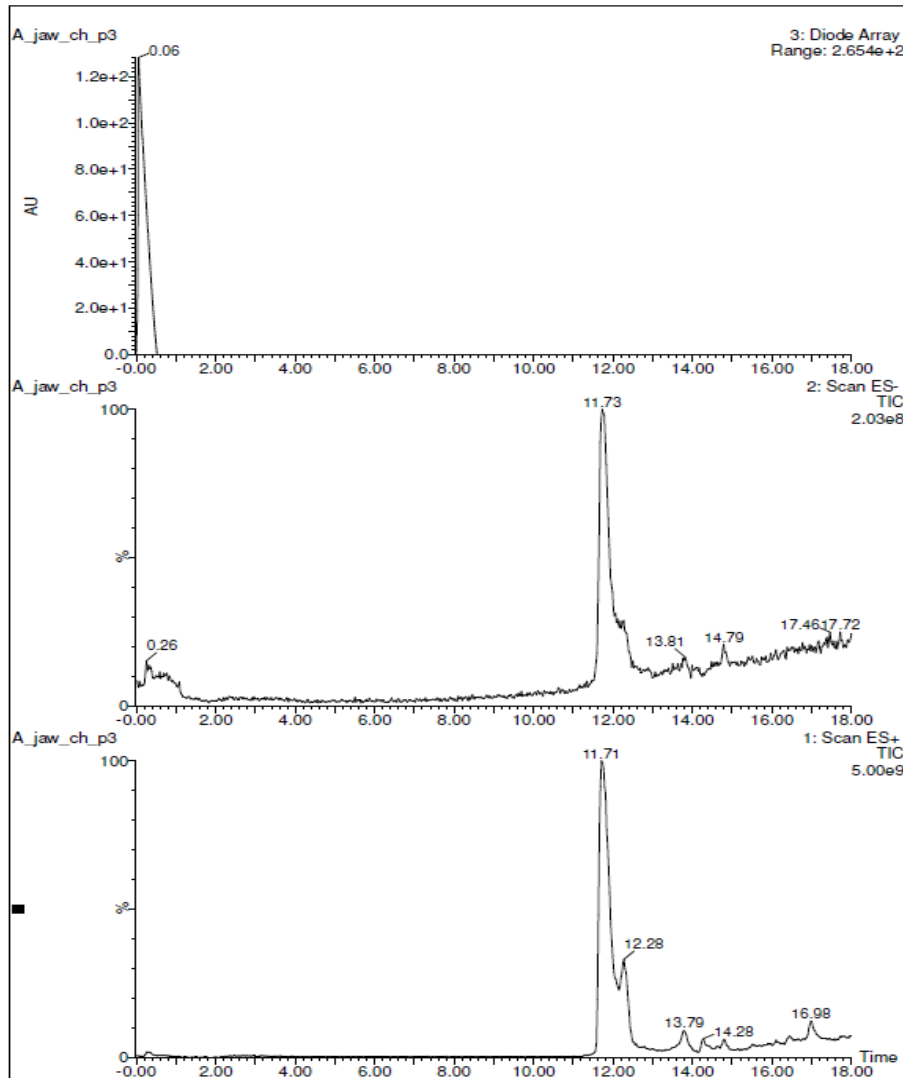
# LCMS Chromatograms and Spectra: 12-mer-VS



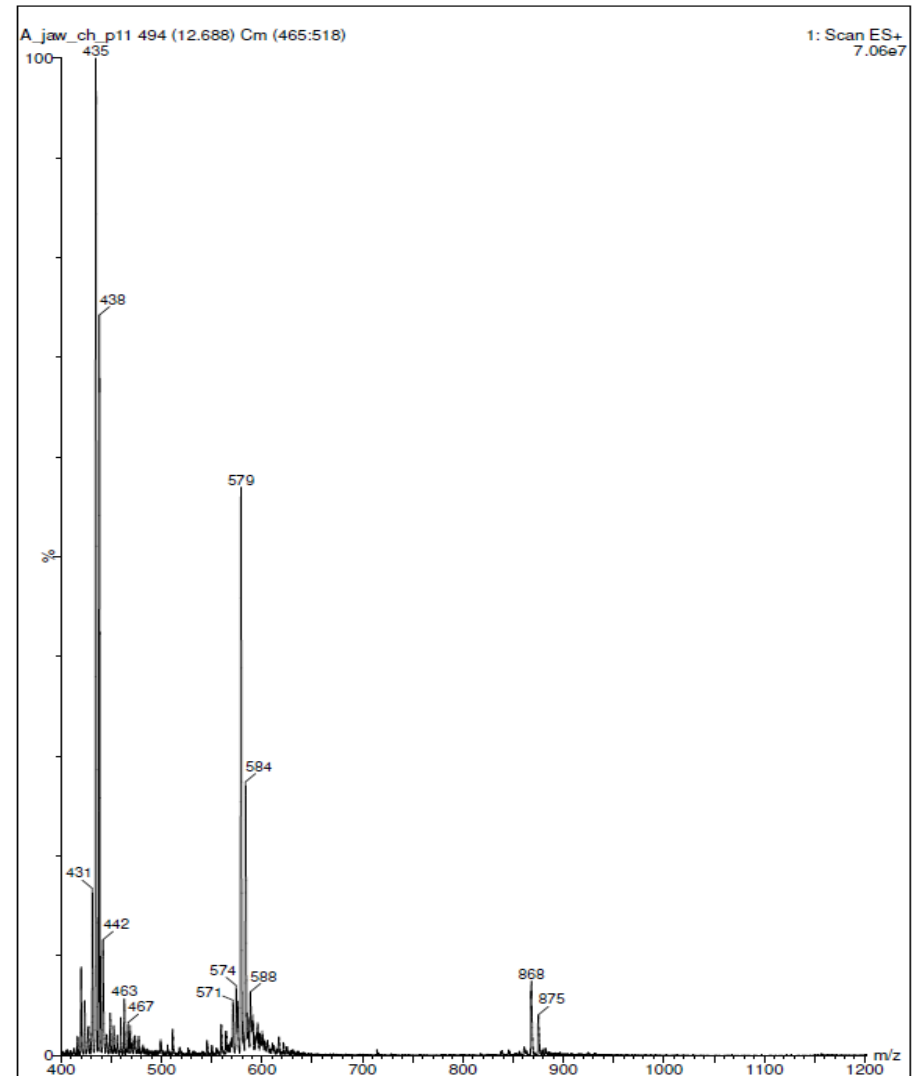
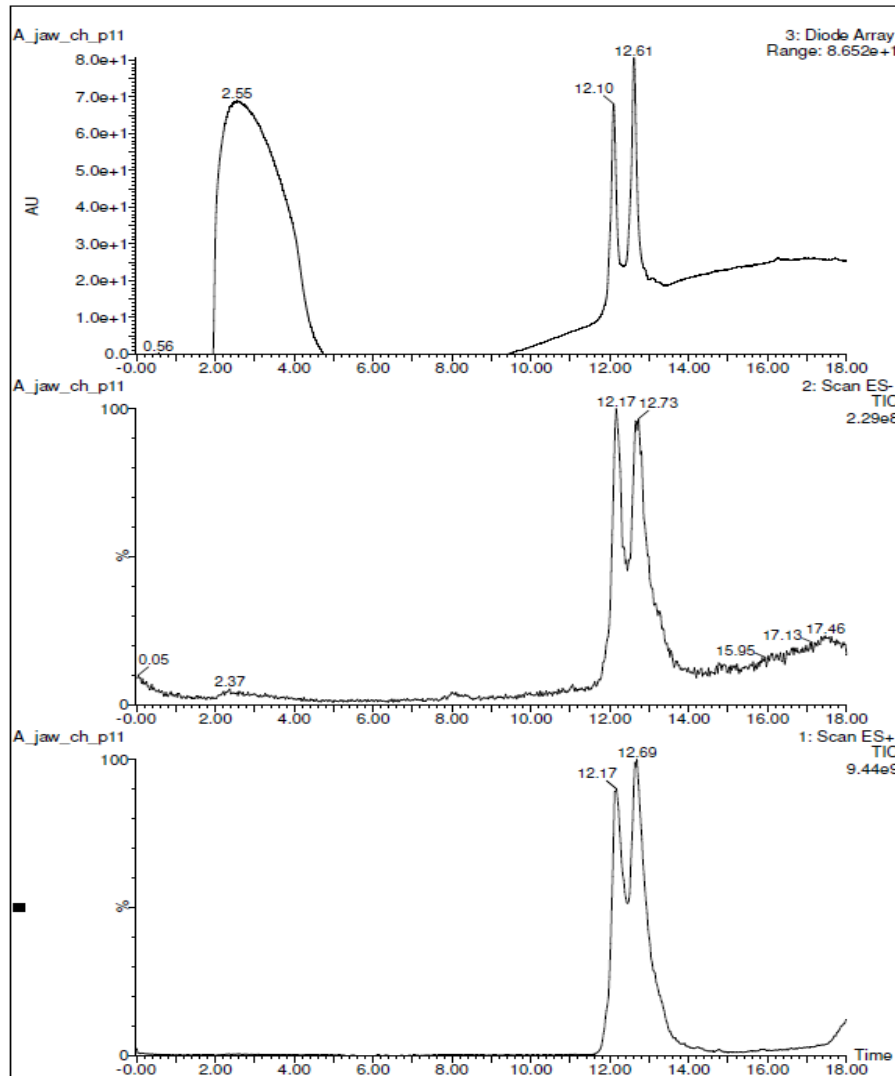
# LCMS Chromatograms and Spectra: 12-mer-VME



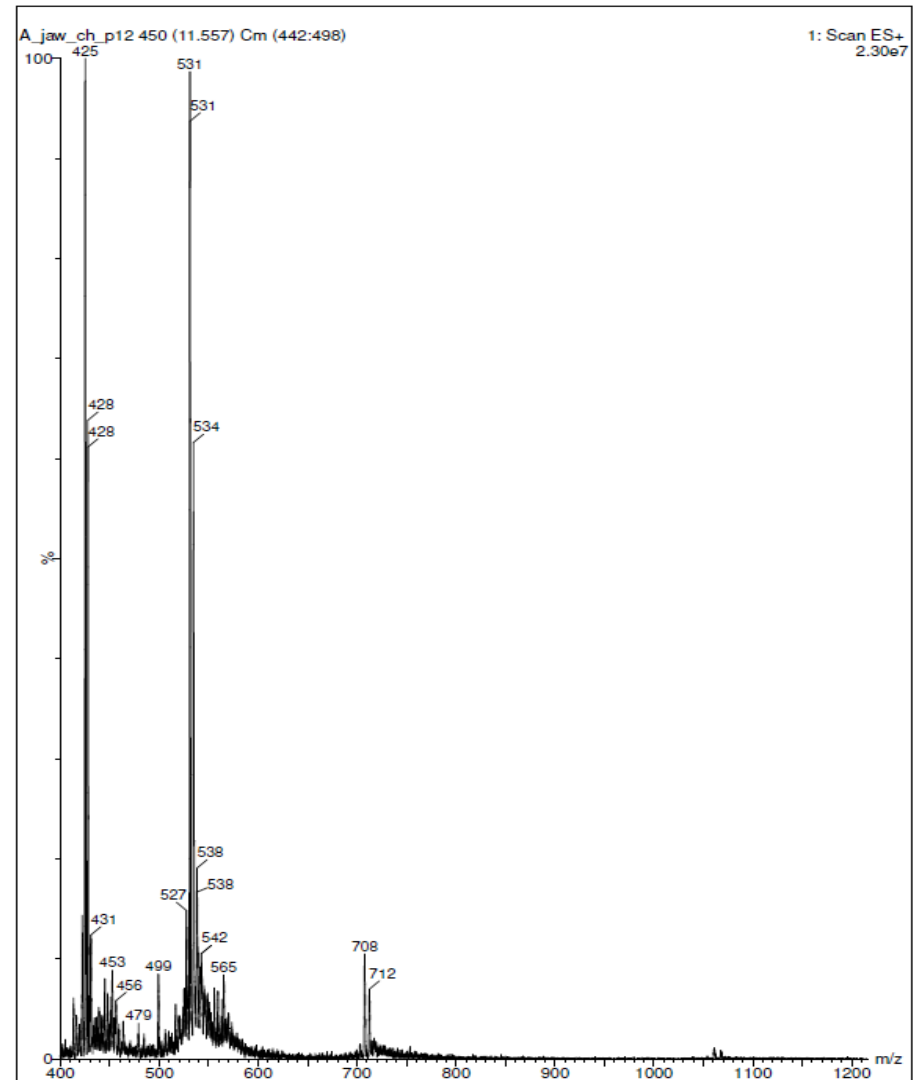
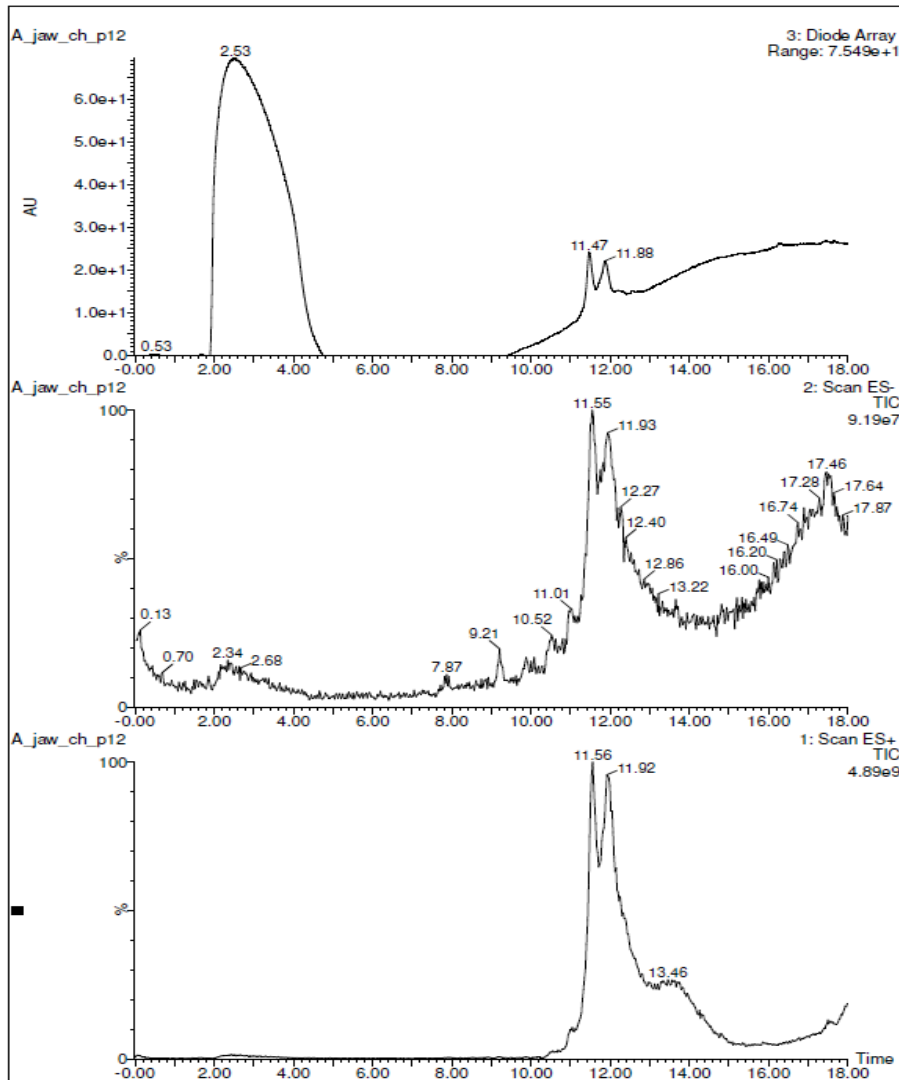
# LCMS Chromatograms and Spectra: 12-mer



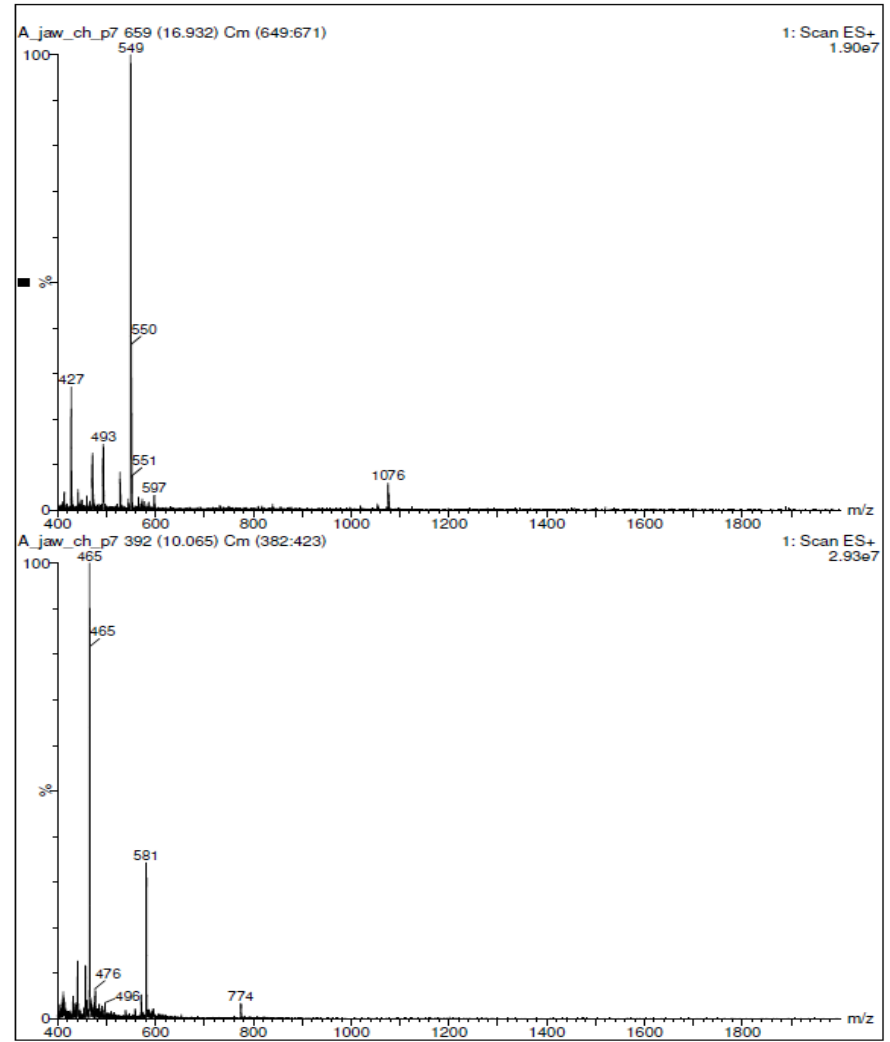
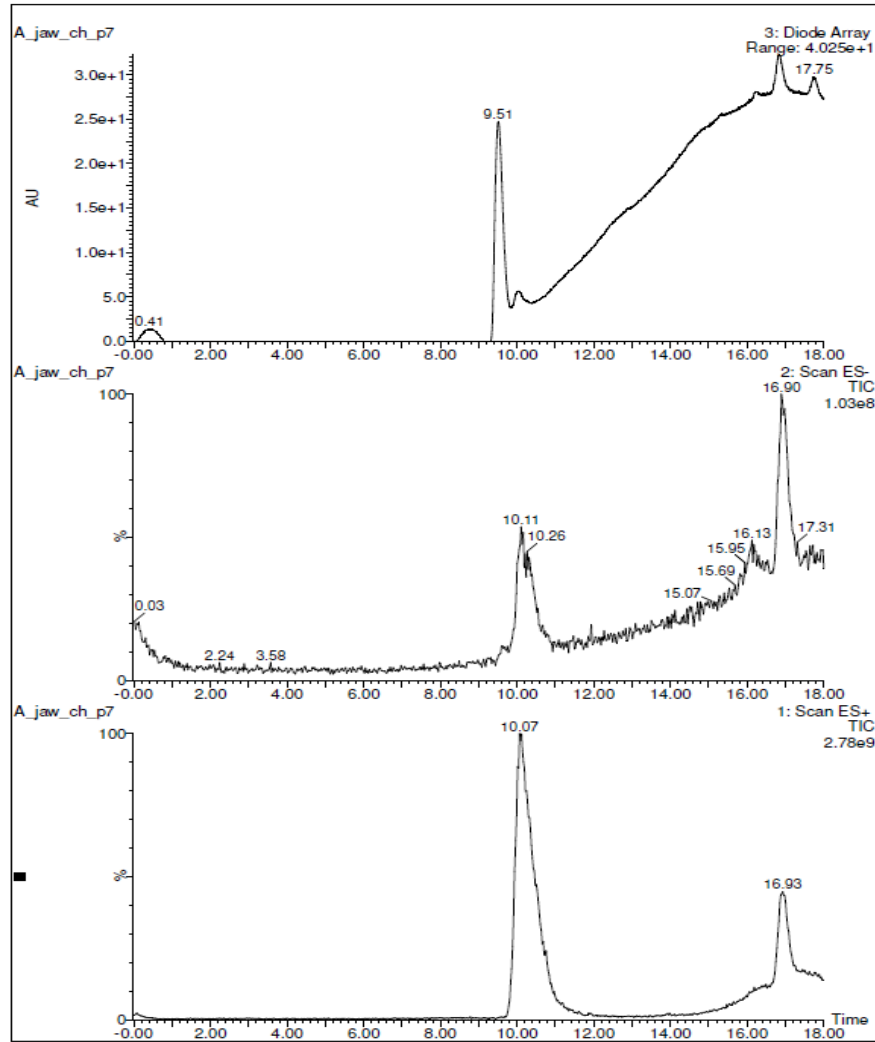
# LCMS Chromatograms and Spectra: T-12-mer



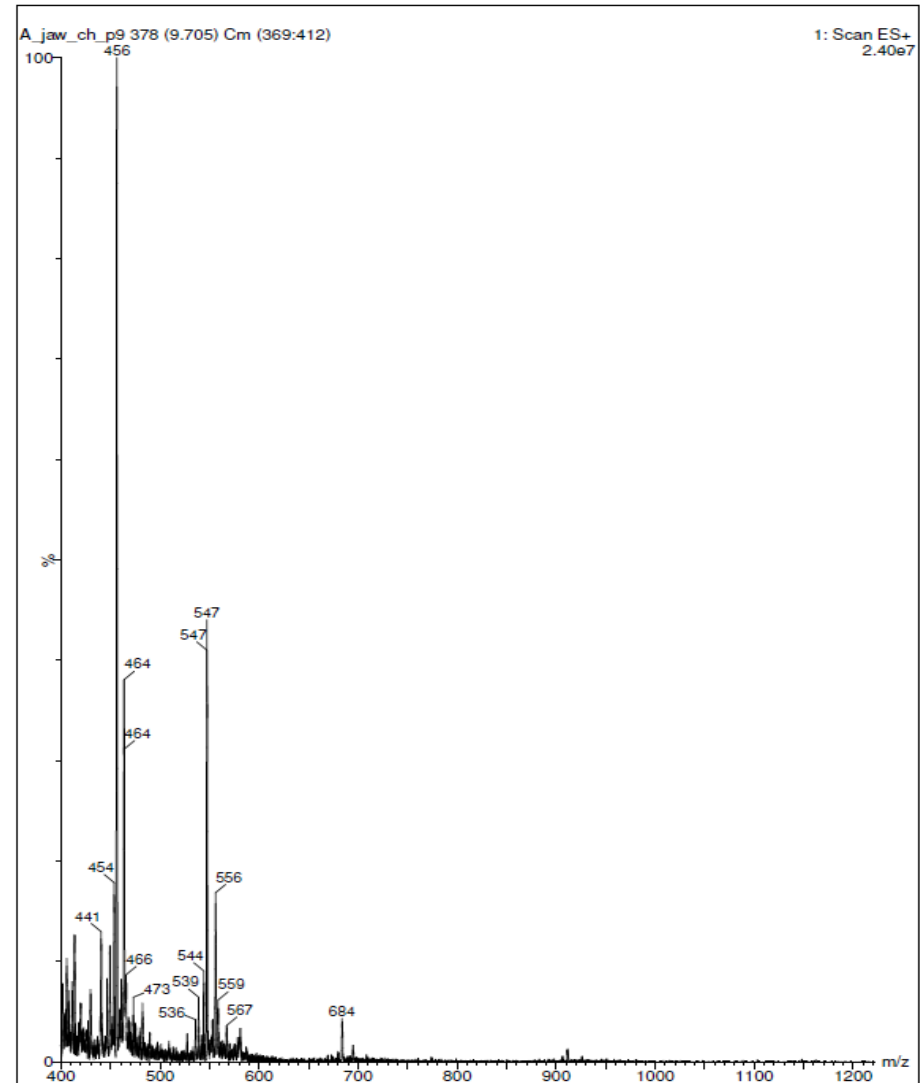
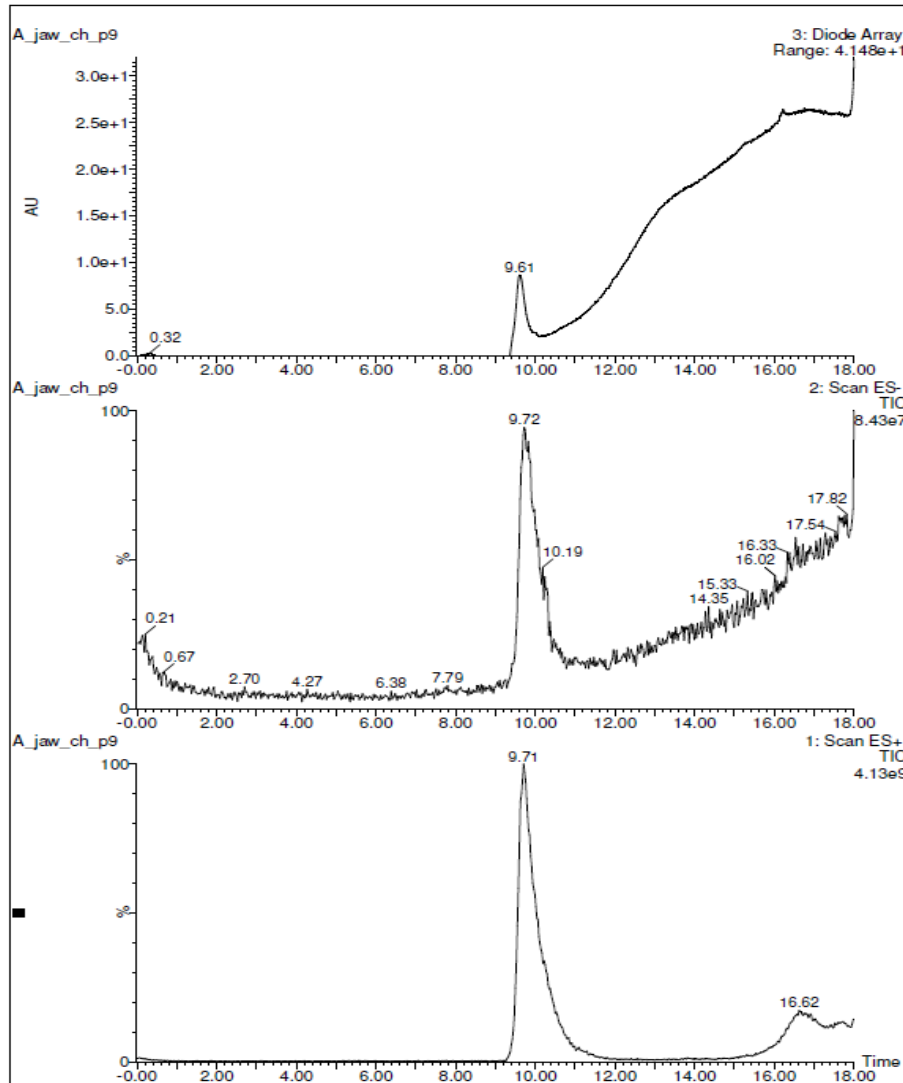
# LCMS Chromatograms and Spectra: T-15-mer



# LCMS Chromatograms and Spectra: NLS



LCMS Chromatograms and Spectra: CPP1



LCMS Chromatograms and Spectra: CPP2

