The ‘dark matter’ of ubiquitin-mediated processes: using multiple approaches to investigate ubiquitin recognition

Elizabeth Radley BSc.

Thesis submitted to the University of Nottingham
for the Degree of Doctor of Philosophy

September 2019

Abstract

Post-translational ubiquitin modification of target proteins acts to regulate a diverse range of cellular processes, many of which are biomedically relevant. To allow this diversity ubiquitin modifications exhibit remarkable complexity, determined by a combination of polyubiquitin chain length, linkage type, numbers of ubiquitin chains per target, and decoration of ubiquitin with other small modifiers. This ubiquitination complexity is interpreted through selective recognition by ubiquitin-binding domains (UBDs). However, there is a notable mismatch between the number of ubiquitin-mediated processes, specific (poly)ubiquitin signal variations, and functionally-verified endogenous UBDs.

This led to the hypothesis that many endogenous UBDs have been overlooked and that there exist additional layers to specific ubiquitin recognition yet to be discovered. To test this hypothesis three main approaches are presented; bioinformatic review and analysis, high throughput Phage Display, and biochemical (proteomic) approaches, towards the identification of previously overlooked UBDs.

Firstly, a comprehensive bioinformatic overview of the variety of UBDs and their classifications was conducted resulting in up to date, human specific and, where possible, functionally-verified SLiMs for the identification of new UBDs. Secondly, Phage Display was used to pan a randomised peptide library against wildtype-, phosphomimetic- and phosphoubiquitin with Next-Generation Sequencing (NGS) employed to suggest novel artificial ubiquitin-binding peptides. Despite encouraging initial results candidate peptides could not be functionally validated as ubiquitin-binding through protein NMR spectroscopy. Finally, biochemical approaches with ubiquitin-affinity chromatography, LC-MS/MS identification and validation by protein NMR were used to uncover previously unidentified ubiquitin-binding proteins.

This multi-discipline approach led to the identification of a novel ubiquitin-binding protein, peflin. This protein did not contain any known UBD, suggesting the presence of a candidate novel UBD. As peflin is a calcium-binding penta-EF hand protein the effect of calcium on the peflin-ubiquitin interaction was investigated. Initial results indicated that calcium potentiated the interaction, providing evidence for the first known example of calcium and ubiquitin signal integration within a ubiquitin-recognition protein. Unexpectedly, ubiquitin was also found to respond to physiologically relevant calcium concentrations in the absence of any protein ligand. It was speculated that this could act to tailor ubiquitin signals in high calcium cell compartments for signal localisation. This preliminary data can be contextualised with recent reports of alternative ubiquitin conformational states and previous instances of ubiquitin binding to other divalent metal cations. Taken together, the concepts discussed here work towards the identification of new UBDs which may represent the 'dark matter' of the ubiquitin system.

"Stick kettle on and look out the window;
sun is rising, birds still singing, tis a beautiful day."

Prof. Rob Layfield,
helping me through lab-induced stress

"It is a truth universally acknowledged that every specific ubiquitin signal
is in want of a binding domain."

me,
drunk in the pub, enthusiastically trying to
explain my research to a complete stranger

Acknowledgements

First and foremost, I'd like to thank Mary, Niamh and my grandma; three women without whom I'd never have made it to this PhD, let alone finished it.

Formally, I'd like to thank Prof Rob Layfield, Prof Mark Searle, Dr Kevin Gough and Dr Jed Long for supervisory advice and support. You took a risk when you accepted an irreverent, caffeine-addicted PhD student all those years ago, I hope it paid off. Thanks to you I now know far more than I thought possible about ubiquitin and Phage Display (although I'm still irreverent and caffeine-addicted). Additionally, I'd like to thank Dr Huw Williams, Dr Ali Gray, and Dr Simon Dawson for providing assistance, support, technical advice and reagents. Separately, I'd like to thank the University of Nottingham high field NMR facility for equipment and assistance in NMR experiments.

Within the lab, special thanks have to be given to Alice, Barry, Jed and Dan for teaching me pretty much everything I know. To Barry, thanks for keeping us all in check; the lab would fall apart without you. A boat load of thanks to Jed, without your help and guidance I wouldn't have a PhD. Especial thanks to Sophie for checking up on me and noticing when I was more stressed than I appeared. Although no thanks to Dan and James for trying to give me a lab initiation hazing ceremony and (nearly) making me throw up in the JA (but thanks for everything else). Also, thanks to those in the lab who have since moved on, Charlie, Jess, Amy and Julius, for seeing me through late night experiments.

Kevin and the Gough lab/ADAS in general deserve a special acknowledgement for letting me crash in their lab for two years, for teaching me all about phage and for treating me as one of their own. You're a frankly insane bunch, but you absolutely made my time in Sutton Bonington and you kept me going through two years of failed experiments. For both professional and personal support I can't thank you all enough; I don't think I'll ever work in a lab as fun as yours.

Thank you to the D61 office in general, for humouring me when I shout contextless questions into the void and reminding me to eat on a regular schedule. Specifically, to Oli and Matt, for sharing write up stress, being the thesis support group and keeping me sane – we made it!

I'm particularly grateful to Emma, Raj and Al for taking me in as a waif and stray that time I was temporarily homeless, and for friendship and emotional support since.

A million thank yous to Georgia, Jenny, Lois and Beth. I've missed so many of your life milestones and celebrations due to this PhD but you've stuck with me throughout it all, and you can never understand how much that meant to me. I wouldn't have made it this far without you. You've kept me right since we met, and I hope our friendship will continue for many years to come.

Thanks to my family, for trying to be enthusiastic about my work despite not really getting it. Specifically to my dad, for being the Dr Radley I aspire to be (but biology is still better than chemistry), to my mother for teaching me to love learning and smash the patriarchy, and to my brother for being my partner in crime in long patience and meaningful eye rolls at family events - you'll never read this but thank you anyway. I'd also like to thank Teresa for being the sister I never had.

Last and not but least, I'd like to thank Rob for being the best supervisor a student could ask for, and for being 90% of the reason I made it to the end of this beast of a doctorate. I've genuinely enjoyed my PhD, and that's all down to you. Your dedication to your lab as both professionals and individuals had made a huge impact on my life, and I don't think I would have survived this PhD with any other supervisor. Also, thanks for replying to my 5am panic emails and for not firing me that time I told you to. It was the stress talking.

(P.S. thanks for caffeine and dry shampoo, for making me look like I got more than 5 hours sleep and had some form of reasonable life:work balance.)

Publications

Publications of relevance that arose from the investigations underpinning this thesis, from which some of the material presented here has been reproduced.

Radley, E.H., Long, J., Gough, K.C., and Layfield, R., 2019. The 'dark matter' of ubiquitin-mediated processes: opportunities and challenges in the identification of ubiquitin-binding domains. *Biochemical Society Transactions*: **47 (6)**: 1949–1962

Abbreviations

ACK	Activated CDC42-associated tyrosine kinase
AE	Acetic acid/EDTA
ALG-2	Apoptosis-Linked Gene 2 protein
ALS	Amyotrophic lateral sclerosis
APC/C	Anaphase-promoting complex/cyclosome
BC	Barcode
BCA	Bichinchoninic acid
Beta-Prp/WD40	Beta-Propeller/WD40
CARD	Caspase recruitment domain
CD	Circular dichroism
CNBr	Cyanogen Bromide-activated
CoCUN	Cousin of CUBAN
COPII	Coat protein complex II
CSB	Cockayne Syndrome B
CSP	Chemical shift perturbation
CUBAN	Cullin-binding domain associating with NEDD8
CUE	Coupling of ubiquitin conjugation to endoplasmic reticulum degradation
CUL3	Cullin-3
DUB	Deubiquitinase
DUIM	Double-sided ubiquitin-interacting motif
DUSP	Domain present in ubiquitin-specific proteases
EGF	Epidermal growth factor
EGFR	Epidermal growth factor receptor
ELISA	Enzyme-linked immunosorbent assay
ER	Endoplasmic reticulum
ERES	ER exit sites
ESCRT-1	Endosomal sorting complex required for transport-1
ESI	Electrospray ionisation
FPLC	Fast protein liquid chromatography
FTLD	Familial frontotemporal lobar dementia
GAT	GGA and TOM
GLB	Gel loading buffer
GLUE	Split-pleckstrin homology domain
GO	Gene Ontology
GST	Glutathione S-transferase
HECT	Homologous to the E6-AP Carboxyl Terminus
HRP	Horseradish peroxidase
HSQC	Heteronuclear Single Quantum Coherence
IPTG	Isopropyl β-d-1-thiogalactopyranoside
ITC	Isothermal Titration Calorimetry

Jab1/MPN	Jab1, Mpr1, Pad1 N-terminal
KHNYN	KH and NYN domain-containing protein
KLHL12	Kelch-like protein 12
LB	Luria Bertani Broth
LC-MS/MS	Liquid Chromatography with tandem mass spectrometry
MGRN1	Mahogunin Ring Finger 1
MINDY	Motif Interacting with Ubiquitin-containing Novel DUB Family
MIU	Motif interacting with ubiquitin
MyUb	Myosin VI Ubiquitin-binding domain
N4BP1	NEDD4 binding protein 1
N4BP2	NEDD4 binding protein 2
NEDD4	Neural precursor cell expressed developmentally down-regulated protein-4
NEDD8	Neural precursor cell expressed developmentally down-regulated protein-8
NEMO	NF-κB essential modulator
NF-κB	Nuclear factor-κB
NGPD	Next-Generation Phage Display
NGS	Next-Generation Sequencing
NMR	Nuclear Magnetic Resonance
NZF	Npl4 zinc finger
OD	Optical density
OPTN	Optineurin
OTU	Ovarian Tumour Protease
PBS	Phosphate buffered saline
PDB	Protein Data Base
PDCD6	Programmed Cell Death Protein 6
PEF	Penta-EF hand
PEF1	Peflin
PFU	PLAA family ubiquitin binding
PINK1	PTEN-induced putative kinase 1
PRU	Pleckstrin homology domain
PTEN	Phosphatase and tensin homolog
PTM	Post-translational modification
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SH3	Src homology 3
SILAC	Stable isotope labelling of amino acids in cell culture
SOUBA	Solenoid of overlapping UBAs
SPR	Surface Plasmon Resonance
SUMO	Small ubiquitin-like modifier
TAB2	TAK1-binding protein 2
TAB3	TAK1-binding protein 3
TAK1	Transforming growth factor-β-activated kinase 1
TAT	Tyrosine aminotransferase

TBK1	Tank-binding kinase 1
TBS	Tris buffered saline
TS101	Tumor susceptibility gene 101
tUBD	Tandem UBD
UBA	Ubiquitin associated
UBAN	Ubiquitin binding in ABIN and NEMO
UBAP1	Ubiquitin-associated protein 1
UBC	Ubiquitin conjugating enzyme
Ub-CR	Ubiquitin C-terminally retracted
UBD	Ubiquitin-binding domain
Ubl	Ubiquitin-like
UBL	Ubiquitin-like Domain
UBM	Ubiquitin-binding motif
UBP	Ubiquitin-specific processing protease
UBW	Ubiquitin-binding winged-helix domain
UBZ	Ubiquitin-binding zinc finger
UCH	Ubiquitin C-terminal Hydrolase
UEV	UBC E2 variant
UEVLD	Ubiquitin-conjugating enzyme E2 variant 3
UIM	Ubiquitin-interacting motif
UMI	UIM- and MIU-related UBD
USP	Ubiquitin-Specific Protease
VHS	VPS-27, Hrs and STAM
WHD	Winged-helix structural domain
WT	Wildtype
ZnF	Zinc finger
ZnF_A20	A20 type zinc finger
ZRF1-UBD	Zuotin-related factor 1 UBD

Contents

1 Introduction 1
1.1 Ubiquitin-mediated processes – why do they matter? 1
1.2 Creation and removal of the ubiquitin signals 3
1.3 Functional consequences of ubiquitin signalling 6
1.4 Ubiquitin-binding domains; driving ubiquitin recognition 9
1.5 The continuously expanding UBD family tree 12
1.5.1 The unique CUBAN UBD 13
1.5.2 The unique CoCUN UBD 14
1.5.3 The unique SOUBA UBD 15
1.5.4 The unique ZRF1-UBD 18
1.5.5 The unique UBW UBD 19
1.5.6 The unique MyUb UBD 21
1.5.7 Growing the ubiquitin-binding domain family tree 24
1.6 Mechanisms of specific (poly)ubiquitin recognition 24
1.6.1 Multiple UBDs and avidity 25
1.6.2 Polyubiquitin linkage selectivity 28
1.6.3 Polyubiquitin length selectivity 31
1.6.4 Selectivity for post-translationally modified ubiquitin 32
1.7 The potential for future selective UBDs 39
1.8 Bacteriophage as an experimental tool 40
1.8.1 Basic bacteriophage biology 40
1.8.2 Bacteriophage proteins and their uses 42
1.9 Next-Generation Phage Display 43
1.9.1 Development of Next-Generation Phage Display 43
1.9.2 Conducting Next-Generation Phage Display 45
1.10 Phage Display and the study of ubiquitin 47
1.11 The dark matter of ubiquitin recognition 51
2 Materials and Methods 55
2.1 Materials and reagents 55
2.2 Antibodies 55
2.3 Purchased peptides and proteins 56
2.4 Production of purified proteins 57
2.4.1 Ubiquitin production in bacterial culture 57

2.4.2 Ubiquitin purification strategies 57
2.4.3 Peflin growth in bacterial culture 59
2.4.4 GST-p62 growth in bacterial culture 63
2.5 General biochemistry methods 64
2.5.1 DNA electrophoresis 64
2.5.2 SDS-PAGE 64
2.5.3 Western blot 65
2.5.4 Coomassie staining 66
2.5.5 Bicinchoninic acid (BCA) assay 66
2.5.6 Obtaining porcine brain tissue lysates 66
2.5.7 Cyanogen Bromide-activated (CNBr)-Sepharose affinity chromatography 66
2.5.8 LC-MS/MS 69
2.6 Next-Generation Phage Display 70
2.6.1 ELISA 70
2.6.2 Preparation of TG1 stocks 71
2.6.3 Preparation of helper Exphage stocks 71
2.6.4 Preparation of the M13 bacteriophage library 73
2.6.5 Phage display and bio-panning method 78
2.6.6 Next-Generation Sequencing of Phage Display samples 86
2.7 Nuclear Magnetic Resonance 91
2.7.1 NMR experimental design 92
3 Defining consensus sequences for human UBDs – a bioinformatics approach 95
3.1 Introduction 95
3.2 Defining human UBD-proteins 97
3.3 Approaches to generating consensus sequences 103
3.4 Parameters in determining UBD consensuses 106
3.5 Consensus motifs for the α-helical UBD families 109
3.5.1 A consensus motif for the UBA family 110
3.5.2 A consensus motif for the UIM family 110
3.5.3 A consensus motif for the CUE family 111
3.5.4 A consensus motif for the VHS family 113
3.5.5 A consensus motif for the GAT family 115
3.5.6 A consensus motif for the MIU family 117
3.5.7 A consensus motif for the UBAN family 118
3.5.8 A consensus motif for the DUIM family 118
3.5.9 A consensus motif for the UMI family 120

3.5.10 A consensus motif for the UBM family 121
3.6 Consensus motifs for the ZnF UBD families 122
3.6.1 A consensus motif for the NZF family 122
3.6.2 A consensus motif for the UBP family 122
3.6.3 A consensus motif for the UBZ family 125
3.6.4 A consensus motif for the ZnF_A20 family 127
3.7 Consensus motifs for the UBC-like UBD families 127
3.7.1 A consensus motif for the Ubc family 127
3.7.2 A consensus motif for the UEV family 129
3.8 Consensus motif for the Other UBD families 131
3.8.1 A consensus motif for the beta-Prp family 131
3.8.2 A consensus motif for the SH3 family 134
3.8.3 A consensus motif for the CARD family 136
3.9 Consensus motifs for human UBDs 138
3.10 Application of the UBD consensus resource 141
3.11 Conclusions 144
4 Next-Generation Phage Display targeting monoubiquitin towards generating novel ubiquitin-binding candidate peptides 146
4.1 Introduction 146
4.2 Production of ubiquitin 150
4.3 Set 1: Biopanning against phosphomimetic ubiquitin 151
4.4 Protecting against variation in colony growth 156
4.5 Set 2: Biopanning targeting phosphoubiquitin 157
4.6 Next-Generation Sequencing of selected phage DNA 161
4.7 Selection of NGPD-derived candidate UBD sequences 162
4.8 NMR validation of NGPD-derived UBD peptides 171
4.9 Conclusions 175
5 Proteomic identification of ubiquitin-binding domains and ubiquitin-binding proteins 177
5.1 Introduction 177
5.2 Affinity chromatography and LC-MS/MS identification of mammalian ubiquitin-binding proteins 178
5.3 Trend analysis of affinity chromatography purified ubiquitin-binding proteins 182
5.4 Identifying peflin as a candidate novel ubiquitin-binding protein 189
5.5 Production of purified peflin 191
5.6 Peflin as a novel calcium-sensitive ubiquitin-binding protein 192
5.7 Characterising the peflin-ubiquitin binding interface 195
5.8 Investigating interactions of ubiquitin, peflin and calcium by NMR 199
5.9 The effect of calcium on global ubiquitin binding 212

5.10 Conclusions 216
6 General Discussion 218
6.1 The dark matter of ubiquitin-recognition – why does it matter? 218
6.2 Defining consensus sequences for human UBDs – a bioinformatics approach 219
6.3 Next-Generation Phage Display targeting monoubiquitin towards generating novel ubiquitin-binding candidate peptides 220
6.4 Proteomic identification of ubiquitin-binding domains and ubiquitin-binding proteins 225
6.4.1 Peflin, a little studied candidate ubiquitin-binding protein 227
6.4.2 Characterising the novel interaction between peflin and ubiquitin 232
6.4.3 The novel ability of ubiquitin to act as a calcium sensor? 234
6.5 Future work 242
6.6 Conclusions 247
7 References 249
8 Appendices 281

1 INTRODUCTION

1.1 UBIQUITIN-MEDIATED PROCESSES – WHY DO THEY MATTER?

Post-translational modification (PTM) of cellular proteins helps to specify and diversify interactions *in vivo* (Milo *et al.*, 2010). An additional layer of specificity can be achieved by modifications of some PTMs themselves. An estimated 5% of the human genome serves to carry out PTMs, including SUMOylation, acetylation and phosphorylation (Walsh, 2006). The transient nature of these modifications provides a rapid and dynamic response to the local environment, and their study can provide a wealth of information about regulatory mechanisms and biological processes.

Ubiquitination is one such reversible PTM, involving the covalent addition of one or more ubiquitin molecules to a target protein *via* an isopeptide bond, with several thousand human proteins undergoing ubiquitination (Chen *et al.*, 2014; Huang *et al.*, 2016). Ubiquitin is a highly conserved protein of 76 amino acids which are all contained within the globular structure apart from the flexible Leu-Arg-Gly-Gly C-terminal tail (Fig.1.1.1) (Vijay-Kumar *et al*., 1987).

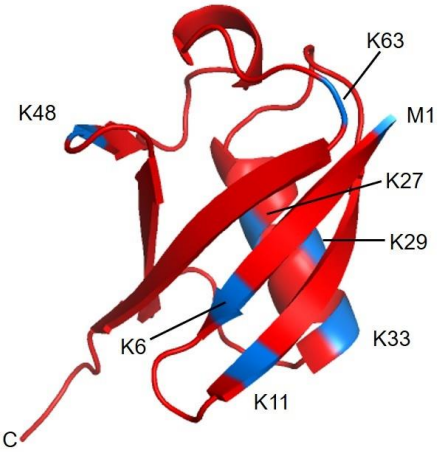


Fig.1.1.1. The general structure of a human ubiquitin monomer. The seven lysines and N-terminal methionine used for polyubiquitination are shown (blue). The human sequence is shown below with the same colour scheme. Figure created with PyMOL v2.3.2 from PDB 1UBQ (Vijay-Kumar *et al*., 1987).

Ubiquitin was first implicated in lymphocyte differentiation (Goldstein *et al.*, 1975) and has since earned its name by being ubiquitously expressed and linked to over 600 biological processes and molecular functions in the QuickGO database (Huntley *et al.*, 2014).

These processes are diverse, ranging from cell cycle control, autophagy, DNA repair and viral assembly, to plant immunity. Accordingly, dysfunction or dysregulation of ubiquitin-mediated processes has been implicated in a wide variety of pathologies, including numerous cancers, kidney disease, Paget's disease of bone, glaucoma, Alzheimer's disease, Parkinson's disease, Amyotrophic lateral sclerosis (ALS) and prion diseases (for comprehensive reviews, see Rape, 2018; Gilberto and Peter, 2017; Li *et al.*, 2016; Rittinger and Ikeda, 2017; Courtois *et al.*, 2018 Sharma *et al.*, 2016; Mansour *et al.*, 2018). Consequentially, the study of the specific formation, removal and recognition of polyubiquitin modifications is crucial for the understanding of both biological and pathological processes, for fundamental and therapeutic research. Despite extensive study, there remains much to be learnt about ubiquitin, especially its specific recognition by proteins which decode ubiquitin signals.

1.2 CREATION AND REMOVAL OF THE UBIQUITIN SIGNALS

E1/E2/E3 enzymes work in sequence to conjugate the carboxy-terminus of an ubiquitin to a selected nucleophilic lysine side-chain in a target protein. First, an E1 ubiquitin-activating enzyme, of which there are two in humans, binds ubiquitin in an ATP-dependent manner, which is then transferred to one of ~40 human E2 ubiquitin-conjugating enzymes. When in thioester-linked complexes with ubiquitin, an E2 enzyme is considered 'charged' or 'loaded'. Next, the charged E2 binds a target-bound E3 ligase which transfers the ubiquitin to the target (Kumar *et al*., 1997; Nuber and Scheffner, 1999). There are over 600 E3 ligases within the human genome, reflecting the specificity between ligase, ubiquitin modification and target protein (reviewed, Zheng and Shabek, 2017). The first round of E1/E2/E3 activity mono-ubiquitinates the target, and repeated rounds of this process can create polyubiquitin chains through the addition of a subsequent ubiquitin to one of seven lysines or the N-terminal methionine in the preceding ubiquitin moiety (Met1, Lys6, Lys11, Lys27, Lys29, Lys33, Lys48 or Lys63) (Chau *et al*., 1989; Bloom *et al*., 2003) (Fig.1.1.1 and Fig.1.2.1).

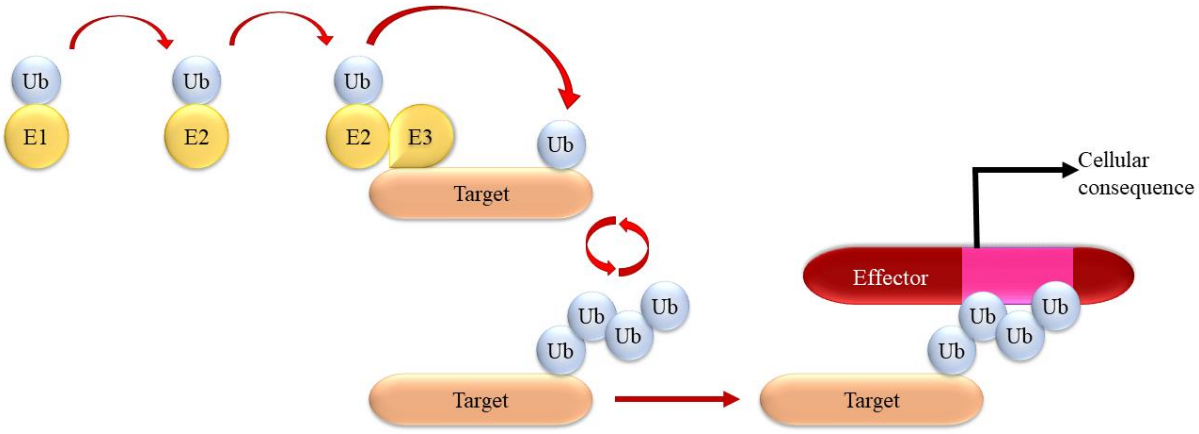


Fig.1.2.1. A summary of ubiquitination and ubiquitin recognition. E1/E2/E3 enzymes ubiquitinate target proteins with polyubiquitin chains of specific length and topology. These chains are recognised and 'decoded' by ubiquitin-binding domains (pink) within effector or ubiquitin-binding proteins. This directs the target protein towards the intended pathway or cell fate.

A single protein target can be ubiquitinated on more than one lysine, and the polyubiquitin chains can form in a homotypic, branched or ring structure. Each of these different

polyubiquitin structures adopts a distinct topology. For example, Lys48-polyubiquitin has been shown by protein nuclear magnetic resonance (NMR) to adopt a 'closed' conformation, whereas Lys63-polyubiquitin has been likened to 'beads on a string' (Varadan *et al.*, 2004; Varadan *et al.*, 2002). The specific topological structure of the ubiquitin modification can form a distinct signal and accurately determine the molecule's fate. Linkage-specific ubiquitination is central to a wide variety of cell processes and is a key mechanism in discriminating between different ubiquitin-related pathways. For example, Lys63-polyubiquitin acts to promote STAT3 signalling through IKKβ ubiquitination (Gallo *et al.*, 2014), is necessary for EGF receptor degradation (Huang *et al.*, 2013) and increase cell sensitivity to the oxidative stress response (Silva *et al.*, 2015), whereas Lys11-polyubiquitin is important for regulation of transcriptional silencing in the DNA damage response (Paul and Wang, 2017) and ubiquitination with Lys11-polyubiquitin chains is hijacked by Hepatitis C to promote successful host cell infection (Pham *et al.*, 2019).

Deubiquitinases (DUBs) are ubiquitin-binding proteins that remove the ubiquitin modifications from the ubiquitinated target. This helps fine tune the ubiquitin-mediated process and ensures that it does not persist. There are around 100 DUBs encoded by the human genome (Nijman *et al.*, 2005) divided into six families; Ubiquitin-Specific Proteases (USP), Ubiquitin C-terminal Hydrolases (UCH), Josephins, JAB1/MPN/MOV34 metalloenzymes, Ovarian Tumour Proteases (OTU), and the Motif Interacting with Ubiquitin-containing Novel DUB Family (MINDY) (Pickart and Rose, 1985; Balakirev *et al*., 2003; Burrow *et al*., 2005; Tzvetkov and Breuer, 2007; Maurer and Wertz, 2016; Rehman *et al.*, 2016). They can act at three different locations within a polyubiquitin modification; cleaving the distal ubiquitin moiety from a chain, cleaving within a chain to remove several moieties at once, or cleaving at the base of the chain to remove the entire polyubiquitin chain

from the target *en bloc*. These are referred to as exo-, endo- and base-cleavage or chain amputation, respectively (Maurer and Wertz, 2016; Heride *et al.*, 2014) (Fig.1.2.2).

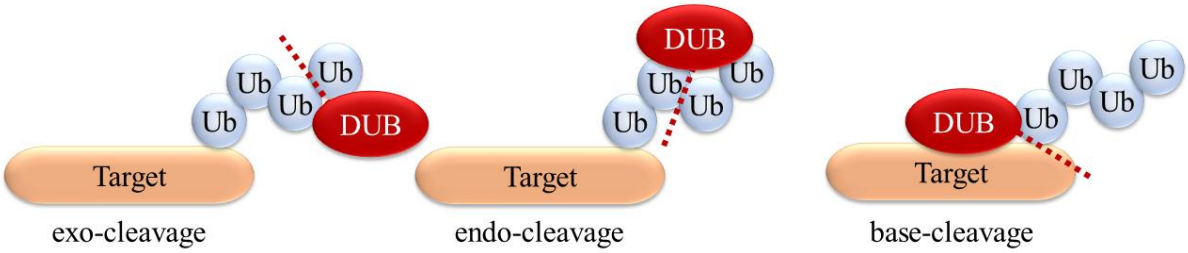


Fig.1.2.2. The three modes of deubiquitination. Deubiquitinases can cleave polyubiquitin chains by removing one ubiquitin at a time (exo-cleavage), by removing multiple moieties (endo-cleavage) or by removing the whole chain (base-cleavage) from the target protein. The site of deubiquitination is indicated by the red line.

Some DUBs show polyubiquitin linkage selectivity, such as Lys48-polyubiquitin cleavage by OTUB1 (Edelmann *et al.*, 2009; Wang *et al.*, 2009). Interestingly, it has been suggested that due to the linkage-specificity of some DUBs, they are unable to conduct base-cleavage and so can only reduce the length of a polyubiquitin chain on a target, rather than completely remove the chain (Bremm *et al.*, 2010). Other DUBs are target-specific, removing various chains appended to a certain target protein. This allows them to cleave multiple forms of ubiquitin from a specific target, as opposed to a single polyubiquitin from various targets (Sowa *et al.*, 2009; Ye *et al.*, 2011; Virdee *et al.*, 2010).

Specificity in ubiquitination is fundamental to its function and is maintained at every level of the ubiquitination process. Whilst the E3 ligases and DUBs work together to promote defined ubiquitin modification of targets and maintain the specificity of ubiquitination, it is the non-covalent recognition of the ubiquitin itself by other effector proteins that determines the consequences of ubiquitination. Specific recognition of the ubiquitin signal is therefore paramount to the coordination of a timely, appropriate and reliable ubiquitin-mediated process.

1.3 FUNCTIONAL CONSEQUENCES OF UBIQUITIN SIGNALLING

The functions and biological consequences of ubiquitin-mediated signalling are numerous, diverse and intertwined. Several have been mentioned above, and here the roles of ubiquitin in transcription factor nuclear factor-κB (NF-κB) signalling in the innate immune response are discussed as an example.

The well-conserved innate immune system is a major part of protection from infection and disease (Martinelli and Reichhart, 2005; Kasamatsu, 2013). The system involves the recruitment of immune cells, production of cytokines, activation of the complement cascade, recognition of infection bodies or substances, activation of the adaptive immune system and clearance of dead cells by autophagy to detect and remove pathological threats (Janeway *et al*., 1989; Martinez *et al*., 2011; Hansen *et al.*, 2016; Zarnegar *et al.*, 2018). It is formed from a multitude of complementary, intertwined and counter-balancing systems with many feedback and regulation mechanisms.

The innate immune system can be summarised as the interaction of microbial pathogens with the host immune system to stimulate signalling through toll-like receptors and interleukin-1 cytokines to increase transcription of inflammatory mediators as part of the normal innate immune response (Cao *et al.*, 1996; Lomanga *et al.*, 1999; Naito *et al.*, 1999). Ubiquitin plays many roles in this process, especially in the activation of transforming growth factor-β-activated kinase 1 (TAK1), NF-κB and IKK. The molecular pathways have been studied for some time and so provide a good example of the importance of specific ubiquitin recognition and ubiquitin-mediated processes (Wang *et al.*, 2001; Shim *et al.*, 2005).

Within this context the most straight forward function of ubiquitin is to activate the transcription factor IRF7. During interleukin-1 cytokine signalling the E3 ligase TRAF6 Lys63-polyubiquitinates target proteins such as the kinase IRAK1 (Deng *et al*., 2000).

Ubiquitinated IRAK1 phosphorylates IRF7, which translocates to the nucleus to promote transcription of type 1 interferons to counter infection (Honda *et al*., 2004; Honda *et al*., 2006; Conze *et al*., 2008).

TRAF6 polyubiquitination of target proteins is also recognised by the ubiquitin-binding domains of TAK1-binding proteins 2 and 3 (TAB2 and TAB3), the mechanism of which will be discussed later in this chapter. TAB2/3 act as adaptors to promote TAK1 activation. TAK1 in turns phosphorylates Ser177 of IKKβ, which along with IKKα and NF-κB essential modulator (NEMO) form the IKK complex. IKKβ-Ser177 phosphorylation leads to autophosphorylation at Ser181 (Xia *et al.*, 2009; Kanayama *et al.*, 2004 Kanayama *et al*., 2004; Ea *et al.*, 2006; Wu *et al*., 2006; Skaug *et al.*, 2009; Zhang *et al*., 2014). Activated IKK mediates Ser32 and Ser36 phosphorylation of the inhibitory component of NF-κB. This is then Lys48-polyubiquitinated by the E3 ligase $SCF\beta^{TRCP}$, targeting the inhibitory component for proteasomal degradation. The other components of NF-κB are then free to promote gene transcription (Chen *et al*., 1995; DiDonato *et al*., 1996; Winston et al., 1999; Nomura *et al.,* 2000; Zhang *et al.,* 2015).

The NEMO component of IKK also contains ubiquitin-binding domains. Their function is required to colocalise the TAK1 and IKK complex and so provide an extra layer of ubiquitin-mediated regulation (Yamaoka *et al.,* 1998; Rothwarf *et al*., 1998; Ea *et al*., 2006). The exact specificity of the NEMO ubiquitin-binding domains has not yet been fully decoded and will be discussed later in this chapter. However, NEMO is known to bind Met1-polyubiquitin signals catalysed by the linear ubiquitin chain assembly complex (LUBAC) E3 ligase, and this step is required for efficient IKK activation (Lo *et al*., 2009; Rahighi *et al*., 2009; Emmerich *et al.,* 2013; Zhang *et al*., 2014).

LUBAC is formed of the HOIL-1, Sharpin and HOIP components (Kirisako *et al*., 2006; Gerlach *et al.*, 2011; Ikeda *et al*., 2011; Tokunaga *et al.*, 2011). The HOIP component has shown specific recognition of Lys63-polyubiquitin chains *in vitro*, suggesting that initial Lys63-polyubiquitination recruits LUBAC for Met1-polyubiquitination. The HOIL-1 component has shown preference for Met1-polyubiquitin, however, suggesting a positive feedback loop that creates a ubiquitination cascade (Haas *et al.*, 2009; Sato et al., 2011; Emmerich *et al.*, 2013).

Additionally, IKKβ also phosphorylates the transcription factor IRF5 at Ser462, which causes IRF5 dimerisation and translocation to the nucleus. Together with NF-κB, these two transcription factors promote production of interleukin-12, pro-interleukin-1β, ASC, NLRP3 and interferon β. These proteins are required for production of inflammatory mediators, pro-inflammatory cytokine release and activation of immune cells (Takaoka *et al*., 2005; Dinarello, 2009; Ren *et al*., 2014; Lopez-Pelaez *et al.*, 2014).

Overall the signals described here in brief amount to a ubiquitin-dependent pathway from infection to innate immune system activation. This vital pathway is built upon multiple different ubiquitin signals and their specific recognition by ubiquitin-binding proteins. It neatly demonstrates the immense diversity and interconnectivity of ubiquitin-mediated signals which can lead to activation, inhibition, dimerization, recruitment, phosphorylation, degradation, transcription and more. Despite the complexity described here there are many more functions of ubiquitin within transcription during the innate immune response (examples; Schimmack *et al.*, 2017; Fletcher *et al.*, 2018; Lin *et al*., 2019). Incorrect ubiquitin recognition would result in a mistimed or inappropriately activated immune response. It is no surprise, therefore, that the regulation and recognition of ubiquitin has been highjacked by several pathogens and is of great interest to in clinical immunology (Xu *et al*., 2018; Bauer *et al*., 2019; Wan *et al*., 2019; Gibbons *et al*., 2020; Guo *et al.*, 2020).

1.4 UBIQUITIN-BINDING DOMAINS; DRIVING UBIQUITIN RECOGNITION

Ubiquitination is specifically bound and "decoded" by effector proteins. It is that interaction that dictates the fate of the ubiquitinated protein, although it can also be fine-tuned by non-covalent ubiquitin binding to E3 ligases (Kumar *et al.*, 2017). As well as the biological consequences outlined above, ubiquitin "decoding" can also lead to degradation of the protein (Mukhopadhyay *et al.*, 2007), recruitment of polymerases for DNA repair (Hoege *et al.*, 2002), downstream kinase activation (Wang *et al.*, 2001) or construction of ubiquitin-mediated multiprotein complexes (Xia *et al.*, 2009).

There are several binding patches on ubiquitin that are recognised by well-conserved ubiquitin-binding domains (UBDs) within the effector proteins (Sloper-Mould *et al.*, 2001) (Fig.1.4.1). The canonical hydrophobic Ile44 patch on ubiquitin is delineated by Leu8, Ile44, His68 and Val70 on a solvent-exposed β-sheet, surrounded by physiochemically diverse amino acids which permit binding variety and specificity (Sloper-Mould *et al.*, 2001). There are also two common alternative sites; a hydrophobic patch surrounding Ile36 and the diglycine C-terminal tail (Reyes-Turcu *et al.*, 2006) and a polar patch centred on Asp58 (Penengo *et al.*, 2006). The exact binding site varies between UBDs. For example, the Ubiquitin-binding motif (UBM) UBDs of murine Pol iota interact with an altered binding site covering the canonical Ile44 but shifted off-centre towards the C-terminal tail (Bienko *et al.*, 2005; Burschowksy *et al.*, 2011).

The binding affinity between a UBD and ubiquitin moiety is usually low, with average affinities of 10-500μM, although this can be increased to 1-15μM through the use of polyubiquitin chains, multiple UBDs in tandem, avidity effects or various additional post-translational modifications (Sims *et al.*, 2009; Sims and Cohen, 2009; Riedinger *et al.*, 2010; Swaney *et al.*, 2015).

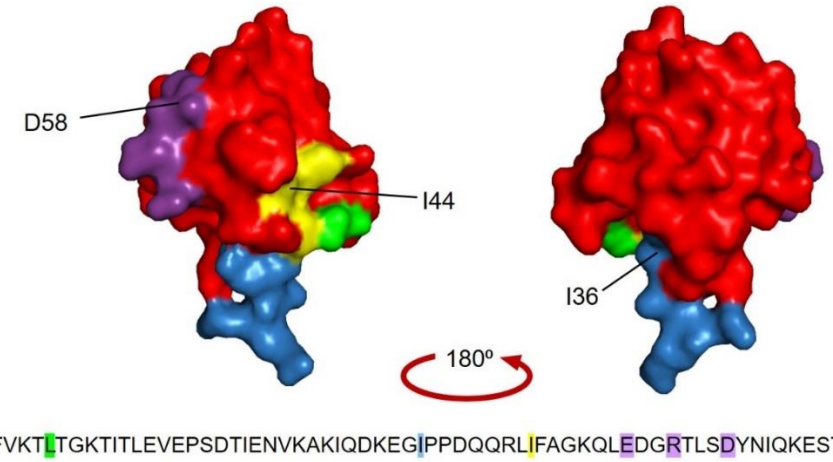


Fig.1.4.1. Three of the major binding surfaces of human ubiquitin. The canonical Ile44 hydrophobic patch is shown in yellow, the polar Asp58 patch in purple and the C-terminal Ile36/Gly75-Gly76 patch in blue. The Leu8 residue shared between the Ile44 and Ile36 patches is shown in green. The eponymous residues of each patch are labelled. The human sequence is shown below with the same colour scheme. Figure created with PyMOL v2.3.2 from PDB 1UBQ (Vijay-Kumar *et al*., 1987).

It is generally agreed that there are currently 19 known families of human UBDs (as of September 2019) which, through a comprehensive literature analysis conducted later in this thesis, was increased to 29 types of UBD including various subtypes and several additional unrelated UBDs (Appendices Table A1.1). The additional UBDs are, currently, only found in a single human protein. As they are unrelated to any family they are sometimes referred to as 'orphan' UBDs. The identification of so many novel UBD families indicates that there is likely still much more to discover about ubiquitin recognition. The recent discovery of novel UBDs may be due to advances in biophysical and high throughput methods enabling identification of the less common or more transiently bound domains. This suggests that further investigations using novel approaches or novel applications of established methods could identify further types or examples of UBDs.

The number of UBD families and members can vary slightly between databases or publications depending on inclusion. The range of species considered can also have an impact as UBDs discovered in plants, for example, may not be present in mammals. As such,

when viewing a list of UBD families or members, careful attention must be paid to the selection criteria and range used when the list was compiled.

The literature has arrived at the consensus of there being around 20 structural families (hedging their bets) (Randles and Walters, 2012). 19 families are well-established in the literature with multiple example proteins to be found, with ten 'orphan families' that only contain a single protein so far (Appendices Table A1.1, asterisk). These families are comparatively little studied, and for the most part are yet to be included in bioinformatic databases or review articles. The E3 ligase Parkin has also been found to contain a ubiquitin-binding site (Kumar *et al.*, 2017). Whilst the interacting residues have been identified, and considerable work undertaken investigating the structural consequences of ubiquitin-binding, no UBD has yet been defined. For this reason it was not included in the list mentioned above, although it will likely be included in future lists.

UBDs can be present in a wide variety of proteins and as noted previously their frequency can vary from single to multiple UBDs per protein (examples, Anindya *et al.*, 2010; Halidi et al., 2019). In some cases the domain may be present in multiple proteins, but only a few members can interact with ubiquitin (for examples see Hologne *et al.*, 2016; Teyra *et al.*, 2017). For example, the Src homology 3 (SH3) domain was originally defined as binding 'PXXP'-containing ligands and it was only later that a limited subset of SH3 domains were found to bind ubiquitin (Stamenova *et al.*, 2007). Similarly, even domains initially identified as UBDs can be re-examined to find alternative functions (Lim *et al.*, 2019; Ding *et al.*, 2018). It may be that several UBDs have multiple functions, providing layers to UBD-mediated processes. Whilst only human examples of UBD-containing proteins are included in the aforementioned table, it should be noted that each family may contain representatives from other species.

1.5 THE CONTINUOUSLY EXPANDING UBD FAMILY TREE

Ten of the 29 families of UBD are unique, in that only a single human protein has been found containing the UBD: Pleckstrin homology domain (PRU), Jab1, Mpr1, Pad1 N-terminal (Jab1/MPN), PLAA family ubiquitin binding (PFU), Split-pleckstrin homology domain (GLUE), Cullin-binding domain associating with NEDD8 (CUBAN), Cousin of CUBAN (CoCUN), Solenoid of overlapping UBAs (SOUBA), Zuotin-related factor 1 UBD (ZRF1-UBD), Ubiquitin-binding winged-helix domain (UBW) and Myosin VI Ubiquitin-binding domain (MyUb). These families are unrelated to the other 19 and are occasionally grouped together as the 'orphan' or 'other' UBD family (Zhou *et al.*, 2017). That these orphan domains continue to be identified (for example, Castagnoli *et al.*, 2019; Nepravishta *et al.*, 2019) suggests that there is still much left to discover about the ubiquitin system. This raises the question: are there still UBD sequences and families that have not been identified? And, most intriguingly, what new aspects of ubiquitination will be discovered alongside them? Will there be new ubiquitin-binding proteins, new ubiquitin-mediated processes or new methods of ubiquitin recognition? It seems that discovery of each new UBD family brings a further layer of interest and complexity to the study of ubiquitination. For example, the CUBAN domain shows stark preference between its very similar substrates, Neural precursor cell expressed developmentally down-regulated protein-8 (NEDD8) and ubiquitin (Santonico *et al.*, 2019), the SOUBA has overlapping Ubiquitin-associated (UBA) domains sharing structural and sequence areas (Agromayor *et al.*, 2012), and the UBW suggests the possibility that a yet undiscovered UBDs could exist in even well-characterised proteins (Takahashi *et al.*, 2019). To closer examine the possibilities opened up by the discovery of these diverse orphan domains, these and other select examples are discussed below.

1.5.1 The unique CUBAN UBD

The CUBAN domain was identified from an experiment aiming to identify proteins that bound both ubiquitin and the ubiquitin-like (Ubl) NEDD8. Ubl proteins are closely related to ubiquitin and share a similar structure, namely, a β-grasp superfold (Mayer *et al.*, 1998), with NEDD8 being the most closely related at 57% sequence identity and 76% similarity (Castagnoli *et al.*, 2019). The Ile44 patch is well conserved between ubiquitin and NEDD8, allowing some overlap in recognition by UBDs (for examples, see Singh *et al.*, 2012; Oved *et al.*, 2006; Den Besten *et al.*, 2012). Despite this the two are functionally distinct, as NEDD8 only modifies members of the Cullin protein family (Lammer *et al*., 1998; Liakopoulos *et al*., 1998).

The Phage Display approach used to identify CUBAN will be described in more detail in sections 1.9 and 1.10, but in brief, it consists of a library of randomised peptides panned against a target protein for identification of binding peptides. The library used by Castagnoli *et al.* (2019) was based on the human brain proteome and so contained peptides that were likely to align to endogenous proteins. This was panned against ubiquitin and NEDD8 to identify binding proteins, the majority of which were already known to bind ubiquitin.

However, one protein, KH and NYN domain-containing protein (KHNYN), was identified as binding both ubiquitin and NEDD8 but contained no annotated UBD. The C-terminal end of KHNYN had considerable homology to other Phage Display-derived proteins with known UBDs and was predicted to form a three α-helix bundle, similar to many known UBD structures. This 52 amino acid-long domain was named the Cullin-binding domain associating with NEDD8, or CUBAN, domain. It was found to prefer NEDD8 to monoubiquitin but displayed stronger binding to Met1- and Lys48-diubiquitin than to NEDD8. This indicated that the CUBAN domain was selective rather than specific: it was able to discriminate between the closely related NEDD8 and monoubiquitin but showed preference for longer polyubiquitin chains.

1.5.2 The unique CoCUN UBD

One of the alternative candidate proteins identified from the ubiquitin-targeting Phage Display experiment of Castagnoli *et al.* (2019) was NEDD4 binding protein 1 (N4BP1). This protein was related to the CUBAN-containing KHNYN, had previously been suggested as a ubiquitin-binding protein (Fenner *et al.*, 2009) and had been shown to inhibit both the HECT E3 ligase Itch (Oberst *et al.*, 2007) by substrate competition and NF-κB by binding to the DUB CEZANNE (Spel *et al.*, 2018). Initial alignments identified a possible Coupling of ubiquitin conjugation to endoplasmic reticulum degradation (CUE) domain: whilst this suggested the presence of an UBD the lack of overall homology suggested it was distinct from the CUE family. Instead, sequence and structural similarity to the previously identified CUBAN domain was noted. As such, this domain was named the Cousin of CUBAN (CoCUN) (Nepravishta *et al.*, 2019).

CUE domains contain a conserved Phe-Pro motif which binds the ubiquitin Ile44 patch and is essential for ubiquitin binding (Shih *et al.*, 2003). As this Phe-Pro motif was one of the areas of similarity with the CoCUN domain, it was tested for involvement of CoCUN binding to ubiquitin (Nepravishta *et al.*, 2019). HEK293-expressed GFP-N4BP1 successfully bound to glutathione S-transferase (GST)-ubiquitin beads, confirming the interaction. Modification of the CoCUN Phe-Pro motif by the engineered P866A mutant prevented interaction. This suggested that the N4BP1 CoCUN domain bound ubiquitin in a manner similar to other related UBDs.

NMR structural analysis identified a three α-helix bundle structure, analogous to that of the CUE (Kang *et al.*, 2003), GGA and TOM (GAT) (Prag *et al.*, 2005), UBA (Dieckmann *et al.*, 1998; Withers-Ward *et al.*, 2000) and CUBAN (Castagnoli *et al.*, 2019) domains. Complementary NMR experiments between ^{15}N-CoCUN and unlabelled ubiquitin, and then unlabelled CoCUN with ^{15}N-ubiquitin, identified the sites of both CoCUN and ubiquitin

involved in binding (Nepravishta *et al.*, 2019). Within CoCUN, Ala14, Leu15, Ile18, Ser22, Ile28, Gln30 and Asp40 contacted ubiquitin, and within ubiquitin the binding residues were Leu8, His68 and Val10. This localised the CoCUN binding site to the Ile44 patch of ubiquitin in a manner similar to that of existing UBDs. Isothermal Titration Calorimetry (ITC) experiments found that the isolated domain bound monoubiquitin in a 1:1 molar ratio with a K_D of approximately 50μM, slightly higher than that of the full length N4BP1 protein, which bound with a K_D of 61μM.

It is known that N4BP1 is endogenously ubiquitinated to alter its subcellular localisation, although the biological relevance of this is unknown (Sharma *et al.*, 2010). Transient HEK293 Phoenix cell expression of the N4BP1 CoCUN domain promoted its own ubiquitination, although whether this was due to a direct effect of the CoCUN domain or simply due to an artefact of the over-expression phenotype is also unknown.

The work did not explore the physiological significance of N4BP1-CoCUN recognition of or modification by ubiquitin. It is possible, however, that given N4BP1's ability to inhibit the E3 ligase Itch (Oberst *et al.*, 2007) ubiquitin recognition or ubiquitination of N4BP1-CoCUN could regulate Itch activity and consequently affect the transcriptional activity of Itch substrates. This may act within the DNA damage response, as N4BP1 also regulated NF-κB activity and is involved in the cellular response to UV. This work described not only the identification of a novel UBD but also demonstrated how Phage Display and NMR could be used to validate ubiquitin-binding activity and determine if a protein bound ubiquitin in a UBD-like manner.

1.5.3 The unique SOUBA UBD

The SOUBA domain, so named for its structure of solenoidal overlapping UBA domains, was identified in the ubiquitin-associated protein 1 (UBAP1) subunit of the endosomal sorting

complex required for transport-1 (ESCRT-1) (Agromayor *et al.*, 2012). UBAP1 has both fundamental and pathological interest within the context of ubiquitin-mediated processes as the ESCRT system is involved in transport of ubiquitinated cell surface receptors for lysosomal degradation (Hicke and Riezman., 1996; Strous *et al*., 1996; Katzmann *et al.*, 2001; Zhou *et al*., 2010). UBAP1 dysfunction has been implicated in nasopharyngeal cancers (Qian *et al.*, 2001), is embryonically lethal when deleted in mice (Agromayor *et al.*, 2012) and is a risk factor in familial frontotemporal lobar dementia (FTLD), which has notable links to ubiquitin and autophagy-related dysregulation (Rollinson *et al.*, 2009).

Several viruses highjack the ESCRT machinery during the sorting of ubiquitinated cargo for degradation and the cutting of thin membranes to facilitate viral release from infected cells for further viral infection (Baumgärtel *et al.*, 2011; Jouvenet *et al.*, 2011; Martin-Serrano and Neil, 2011; Morita and Sundquist, 2004; Morita *et al.*, 2011; Weissenhorn and Göttlinger, 2011). Specifically, the HIV-I virus commandeers ESCRT-1 to degrade the host antiviral protein tetherin in a UBAP1 and ubiquitin-dependent manner and so perpetuate viral infection (Agromayor *et al.*, 2012).

On a molecular level, UBAP1 had already been linked to ubiquitin binding, and contained two 'MGF' motifs, which are generally taken as conserved sequence markers for UBA domains (Stefani *et al.*, 2011). The discovery of the SOUBA domain resulted from crystallisation studies by Agromayor *et al.* (2012) that resolved the C-terminal, MGF-containing region to show seven α-helices (Fig.1.5.3).

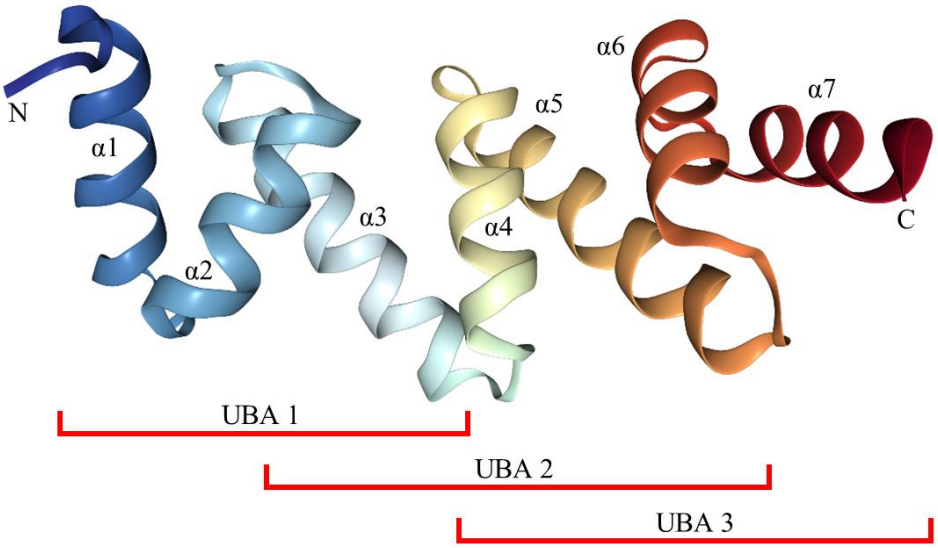


Fig.1.5.3. The helical arrangement of the UBAP1 SOUBA domain. The seven α-helices of the SOUBA domain were determined by crystallisation, spanning positions 389-502. This contained three overlapping UBA domains, each consisting of three α-helices and marked in red below. Adapted from PDB 4AE4 (Agromayor *et al.*, 2012).

The discrepancy between UBA domains canonically containing three α-helices and the presence of seven within the SOUBA was solved by the overlapping structure of the UBAs. The third helix of one UBA domain simultaneously acted as the first helix of the next UBA, creating the distinct overlapping solenoid that was structurally distinct and was not merely multiple UBA domains in a row. This arrangement resulted in a domain surface formed from helices one and three from each of the three UBA domains, equating to helices α1, α3, α5 and α7 in the diagram above. This correlated well with a previous study (Zhang *et al.*, 2008) which showed that interaction between UBA and monoubiquitin commonly used the end of helix α1 and helix α3.

As noted above, two MGF motifs had already been identified within UBAP1 (Stefani *et al.*, 2011). Sequence alignments localised these to UBA-1 and UBA-3 but showed that the methionine of UBA-2 had been replaced by a lysine, resulting in KGF (Agromayor *et al.*, 2012). The physiochemical properties of that position were not conserved as this replaced a bulky hydrophobic with a positive amino acid. Additionally, the dileucine motif that is also conserved within UBA domains was not well-conserved within the SOUBA; none of the

SOUBA UBA domains contained the second Leu. Instead this was replaced by phenylalanine, glutamine and methionine, respectively. Other than the bulky size of these residues, this too represented a physicochemical property change and suggested that the MGF and dileucine markers used to predict UBA domains were not fully appropriate.

ITC experiments provided a relatively high affinity of the SOUBA to both monoubiquitin and Lys48-diubiquitin, at about 70μM compared to the general 20-1,300μM range of most UBA domains (Agromayor *et al.*, 2012). Within this, each of the three UBA domains individually bound monoubiquitin, as per NMR of ^{15}N-SOUBA with monoubiquitin. This was found to involve the MGF motif and the corresponding KGF in UBA-2, which are commonly involved in ubiquitin-binding by UBA domains (Long *et al.*, 2008; Ohno *et al.*, 2005; Varadan *et al.*, 2005; Zhang *et al.*, 2008). The full length UBAP1 protein showed a slightly lower affinity of 140μM for monoubiquitin, although this was still comparatively high for a UBA (Agromayor *et al.*, 2012).

This described the first known example of overlapping UBA domains and provided a new concept in UBD structure; whilst it was previously known that domains could be arranged in tandem (see Section 1.6.1), this was the first instance of a UBA sharing structures or sequences between domains.

1.5.4 The unique ZRF1-UBD

Very little is known about the ZRF1-UBD. It was identified in a short but complex paper (Richly *et al.*, 2010) and appears to have been the subject of much speculation but comparatively little study (Richly and Croce, 2011).

The zuotin-related factor 1 (ZRF1) protein bound through affinity purification to histone 2A-associated ubiquitin and was found to promote its removal, possibly by USP16 or USP21 recruitment (Richly *et al.*, 2010). Different truncation mutants of ZRF1 were assessed for

binding to GST-ubiquitin, localising the UBD to the C-terminal end of the DnaJ domain of ZRF1. This encompassed amino acids from position 160 to 250, inclusive. The domain also showed very minor binding to control GST, but this was considerably lower than binding to GST-ubiquitin. The paper did not discuss the UBD itself and no examination of its structure or sequence was provided.

The paper did, however, suggest a physiological role for the UBD in either DNA damage response or progression of cell growth. Ubiquitin-recognition by ZRF1 protein has been linked to nucleolar DNA repair (Chitale and Richly, 2017) and the formation of E3 ligase complexes at DNA lesion sites (Gracheva *et al.*, 2016). Richly *et al.* (2010) suggested that during cell differentiation the ZRF1 protein is bound to Inhibitor of Differentiation (Id) proteins via the DnaJ domain (Inoue *et al.*, 2000). Then, after ZRF1 recruitment to chromatin, the H2A-monoubiquitin outcompetes Id proteins for binding to the ZRF1-UBD, removing inhibition and allowing cell differentiation to progress. This also tied into ZRF1 misregulation, which is commonly found in various cancers (Resto *et al.*, 2000). This would seem to be at odds with its potential function in promoting deubiquitination of H2A-monoubiquitin. However, as most of this was speculation the two functions of ZRF1 could exist in a feedback loop or could act entirely separately from one another.

1.5.5 The unique UBW UBD

The UBD of Cockayne Syndrome B (CSB) remained undiscovered despite nearly 30 years of CSB study (Troelstra *et al.*, 1992). It wasn't until Anindya *et al.* (2010) conducted a sequence alignment to the CUE domain that the potential for an overlooked UBD within CSB was confirmed. The area of homology was located towards the C-terminus of the protein with a conserved Pro residue and di-leucine motif common to CUE and UBA. Despite this homology the sequences were too divergent to assign the domain as a CUE or UBA, suggesting it was a

novel UBD. Part of this region overlapped with the winged-helix structural domain (WHD) of CSB (Batenburg *et al.*, 2017).

To confirm that the region was functionally a UBD, the C-terminus of GST-CSB was incubated with HEK293 cell extracts, successfully pulling down endogenous ubiquitinated proteins (Takahashi *et al.*, 2019). No ubiquitin binding was seen to control GST, confirming that this interaction was specific for GST-CBS. The same domain also bound purified Lys48- and Lys63-polyubiquitin chains indicating that the interaction was direct.

The WHD of CSB was crystallised in a 1:1 complex with Lys48-diubiquitin, confirming that only the WHD domain, not the full C-terminal tail, was required for ubiquitin binding. In particular, the WHD second α-helix and the C-terminal extremity were critical for the interaction. Mutations of the WHD at D1425A, L1428A, R1432A, F1437D and Y1492A eliminated binding to ubiquitin, as predicted by structural models. Of these residues, CSB Asp1425 was predicted to interact with the Ala46-Gly47 loop of ubiquitin, CSB Leu1428 and Tyr1492 with the Ile44 patch, CSB Arg1432 with ubiquitin Leu71 and CSB Phe1437 with both ubiquitin Leu71 and Leu73 (Fig.1.5.5.1) (Takahashi *et al.*, 2019). The canonical Ile44 patch of ubiquitin was essential for binding, as CSB was unable to bind I44A-ubiquitin (Anindya *et al.* 2010).

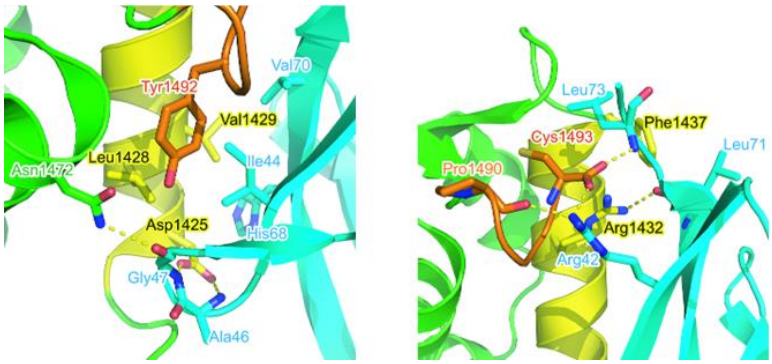


Fig.1.5.5.1. The interface between the CSB-WHD domain and ubiquitin. CSB is coloured in green, yellow and orange, and ubiquitin in blue. Dotted lines denote hydrogen bonds. Adapted from Takahashi *et al.*, 2019.

The ubiquitin-binding ability of WHD was found to be physiologically relevant, as removal of the CSB UBD left cells sensitive to UV irradiation and unable to recover RNA synthesis after irradiation, suggesting involvement in the DNA damage response (Anindya *et al.* 2010). As this is the only WHD found to bind ubiquitin, it was named the Ubiquitin-Binding WHD (UBW) to differentiate it from other WHDs (Takahashi *et al.*, 2019).

1.5.6 The unique MyUb UBD

Myosin VI had already been annotated with a Motif interacting with ubiquitin (MIU) domain ten years before the discovery of the Myosin VI Ubiquitin-binding domain (MyUb) (Penengo *et al.*, 2006). The protein was already linked to Lys63-polyubiquitin-mediated processes including membrane trafficking and the transport of endocytic vesicles (Buss *et al.*, 2004; Tumbarello *et al.*, 2013). It was also speculated that myosin could recognise the Lys11-ubiquitination of transcription regulators by the anaphase-promoting complex/cyclosome (APC/C), presumably for transcription regulation by substrate trafficking (He *et al.*, 2016). If this were the case then it would be of considerable interest, both fundamental and therapeutic, as the APC/C is an E3 ligase and has been a target for cancer research (McHugh *et al.*, 2019).

As well as the MIU, myosin also possesses an RRL motif that is known to mediate interactions with multiple myosin VI ligands (Morriswood *et al.*, 2007; Sahlender *et al.*, 2005; Bunn *et al.*, 1999; Spudich *et al.*, 2007), and it was across this promiscuous motif that the second UBD was found (He *et al.*, 2016).

The first indication of its existence came when deletion of the MIU failed to prevent GST-myosin interaction with ubiquitin from cell lysates. That the MyUb was sufficient for ubiquitin-binding even in the absence of the MIU indicated that MyUb bound ubiquitin in its own right and was not merely acting to stabilise the interaction. In fact, deletion of the MIU only caused a very slight reduction in the amount of ubiquitin bound, as per qualitative Western blot with

anti-ubiquitin. Further deletion analysis identified a 43 amino acid region critical for ubiquitin binding. This domain, named MyUb, was located C-terminally to the MIU domain in a compact helix-turn-helix-like fold as determined by NMR. The fold was stabilised by interactions between Tyr1084, Leu1086 and Trp1089 of the N-terminal region and Leu1094 and Leu1106 of the second helix.

Through affinity chromatography with Western blot and fluorescence polarisation assays GST-MyUb was seen to bind all linkages of diubiquitin, with a noted preference for Lys11-, Lys29- and Lys63-diubiquitin (He *et al.*, 2016). GST-MyUb interaction with Lys48-diubiquitin was only just detectable by Western blot. Correct folding of the domain through the stabilising residues mentioned above was responsible for the preference for Lys63 linkages, as mutations of L1086A or L1106A altered folding and reduced Lys63-polyubiquitin binding. Mutation of Trp1089 was tolerated, however, likely because it did not cause any major change in folding. Mutation analyses of Tyr1084 and Leu1094 were not reported, so it is unknown how they contribute to MyUb folding and specificity.

Size exclusion chromatography determined the binding stoichiometry of MyUb as 1:1 with Lys63-diubiquitin and 1:2 with Lys63-triubiquitin. The structure of the 1:1 interaction was examined by NMR, resulting in MyUb bound in between the two moieties of Lys63-diubiquitin (Fig.1.5.6.1). There were extensive contacts between MyUb and both moieties of diubiquitin, explaining the requirement for correct folding in determining selectivity.

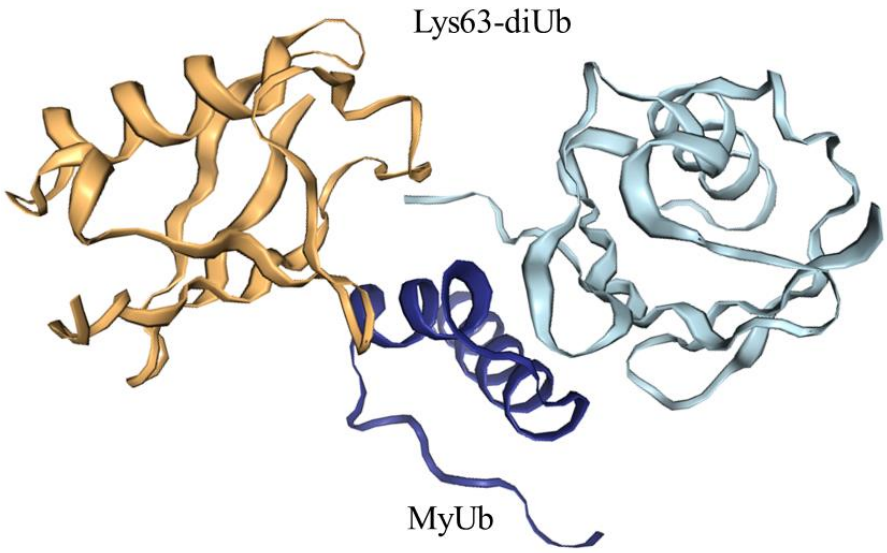


Fig.1.5.6.1. A ribbon diagram of myosin VI MyUb in complex with Lys63-diubiquitin. MyUb Gly1080-His1122 of myosin VI is shown in dark blue. The proximal ubiquitin moiety is shown to the left (gold) and the distal moiety to the right (cyan). Adapted from PDB 2N13 (He *et al.*, 2016).

Myosin VI is known to present as different isoforms according to its tissue and subcellular location (Au *et al.*, 2007; Buss *et al.*, 2001; Tomatis *et al.*, 2013). Specifically, the isoforms contain inserts of variable size between the MIU and MyUb domains. To examine the effect of the variable insert on ubiquitin recognition He *et al.* (2016) created protein fragments spanning the MIU and MyUb domains with different length linking regions. Unsurprisingly, these isoforms showed varying affinity and specificity towards ubiquitin binding. Those MIU-MyUb isoforms with the long insert, analogous to endogenous isoforms located at clathrin-coated pits/vesicles (Buss *et al.*, 2001), showed a four- to five-fold impairment in polyubiquitin binding regardless of the linkage.

The effect of MIU-MyUb linker length on binding specificity could easily have physiological relevance. The different isoforms are expressed in different regions of the cell and myosin VI is known to bind multiple interactors, suggesting a mechanism whereby the isoforms dictate the polyubiquitin linkage bound and therefore target myosin VI to the correct interactor and pathway.

1.5.7 Growing the ubiquitin-binding domain family tree

These novel UBD families have each brought something new to the understanding of ubiquitin recognition and UBD function. The discovery of CUBAN demonstrated the use of Phage Display in identifying novel UBDs (Castagnoli *et al.*, 2019). CoCUN provided a neat demonstration on how sequence alignments could be used to identify UBDs within a protein, and how protein NMR could be used to validate and delineate the binding motif (Nepravishta *et al.*, 2019). SOUBA described a new mode of arranging multiple UBDs within a single protein (Agromayor *et al.*, 2012). ZRF1-UBD demonstrated how much is yet to be learned, even about those UBDs that have already been identified (Richly *et al.*, 2010). UBW proved that even well-studied proteins could contain unidentified UBDs (Anindya *et al.*, 2010), and MyUb suggested that the ubiquitin-binding activity of a protein could be tailored by endogenously expressed isoforms (He *et al.*, 2016).

Their discovery not only hinted at the possibility of more unidentified UBDs but also outlined useful methods for their characterisation and opened up the possibility that there are entirely novel aspects of UBD biology yet to be discovered. For example, there could exist UBDs that show novel means of ensuring specificity, that function in biological processes not yet known to be regulated by ubiquitination or that work cooperatively with other methods of cell signalling to tailor the response to ubiquitination.

1.6 MECHANISMS OF SPECIFIC (POLY)UBIQUITIN RECOGNITION

Specificity in UBD-ubiquitin interactions, and therefore in UBD-containing proteins, is becoming well established and is crucial for discrimination between the many ubiquitin-mediated processes. Several proteins are now confirmed as polyubiquitin linkage-specific or selective (examples, Fu *et al.*, 2009; Sims and Cohen, 2009; Pesiri *et al.*, 2016). Selectivity is achieved by various mechanisms, including target sequence specificity and

crosstalk with other signalling mechanisms. For clarification, throughout this chapter the term tandem UBD (tUBD) shall refer to endogenous multiple UBDs rather than artificial or synthetic multiple UBD arrays.

1.6.1 Multiple UBDs and avidity

In the presence of multiple UBDs the combined affinity for polyubiquitin is often greater than that of the isolated UBDs themselves. This increased affinity from multiple binding events is termed avidity.

Sims and Cohen (2009) suggested that ubiquitin binding by one UBD within a tUBD would bring the second UBD into alignment for a second binding event. This effectively enriches binding partners in the local environment, making the second UBD:ubiquitin interaction more favourable. To test this, Sims and Cohen (2009) led the work into the structural basis of ubiquitin selectivity by the tandem UBDs of Rap80.

The ubiquitin-binding protein Rap80 is central to one of the 'classical' ubiquitin-mediated processes; DNA double-strand break repair. The tandem UIMs of Rap80 recognise ubiquitinated proteins at the break site and act as an adaptor to localise proteins involved in DNA repair, such as BRCC6, Abraxas and BRCA1, to the damage site (Kim *et al.*, 2007; Sobhian *et al.*, 2007; Wang *et al.*, 2007; Wang and Elledge, 2007; Yan *et al.*, 2007). The full length Rap80 has shown a well-documented preference for Lys63- over Lys48-polyubiquitin. The Ubiquitin-interacting motif (UIM) domains in isolation, however, showed no particular linkage selectivity (Kim *et al.*, 2007; Sobhian *et al.*, 2007; Yan *et al.*, 2007), suggesting that it was the combination of two UIMs that conferred selectivity.

Structural models predicted that both the UIM domains and the inter-UIM linker region were α-helical, forming one continuous helix (Sims and Cohen, 2009). When in complex with polyubiquitin this orientated the UBDs such that they faced the same direction. This was

optimal for Lys63-diubiquitin binding. If a polyubiquitin chain was presented that contained an alternative linkage, for example Lys11-polyubiquitin, then the binding faces of ubiquitin would not align with those of the tandem UIMs (Fig.1.6.1.1). For the Rap80 model, correct orientation was dependent on the helix containing an integer number of helical turns, although this is not expected to be a requirement of other tUBD-containing proteins.

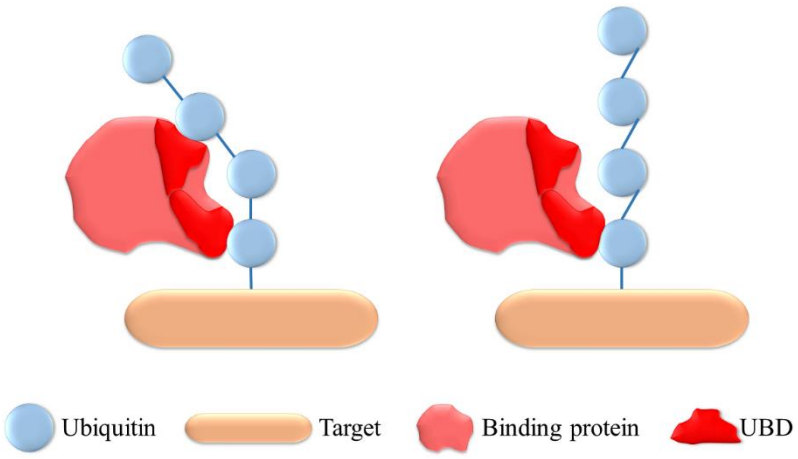


Fig.1.6.1.1. The structural basis of polyubiquitin linkage specificity as determined by tandem UBDs. For efficient binding by a tandem UBD the polyubiquitin must be orientated so that the binding surface of each ubiquitin and each UBD align (left). If another linkage is presented then the binding sites will be misaligned and binding will be inhibited (right).

These models have been supported by ITC, circular dichroism (CD) and NMR data, demonstrating that the avid high affinity to Lys63-diubiquitin was conferred by the alignment of the two UBD domains (Anamika *et al.*, 2014).

A caveat to this concept, however, is that some UBDs could recognise the linker sequence itself. Whilst Sims and Cohen (2009) conscientiously examined a variety of linker compositions, other experiments modifying linker composition may not have taken this into account.

A related idea has been explored by Garner *et al.* (2011), whereby one ubiquitin moiety binds two UBDs from separate proteins to form a ubiquitin-mediated ternary complex (Fig.1.6.1.2-A). Modelling data was in agreement with the logical assumption that smaller binding proteins

are more likely to successfully form ternary complexes, as the side chains of many larger proteins would cause steric hindrance. This was modelled on the UBA domain of p62/SQSTM1 and the A20 type zinc finger (A20-ZnF) of ZNF216, binding to the canonical ubiquitin Ile44 patch and the polar Asp58 patch, respectively. There was no hint of interaction between the binding partners through NMR, however.

It was also possible that ubiquitin could act as a scaffold or recruitment protein (Fig.1.6.1.2-B) or that the binding protein could be the protagonist, oligomerising to create linkage specificity (Fig.1.6.1.2-C). Unfortunately, there has since been very little work building upon this intriguing idea (Wurzer *et al.*, 2015; Kirkin *et al.* 2009). There are of course inherent difficulties in such an experiment: confirming that the two receptor proteins interact directly and confirming that their association allows ubiquitin binding as opposed to interaction as a consequence of ubiquitin binding.

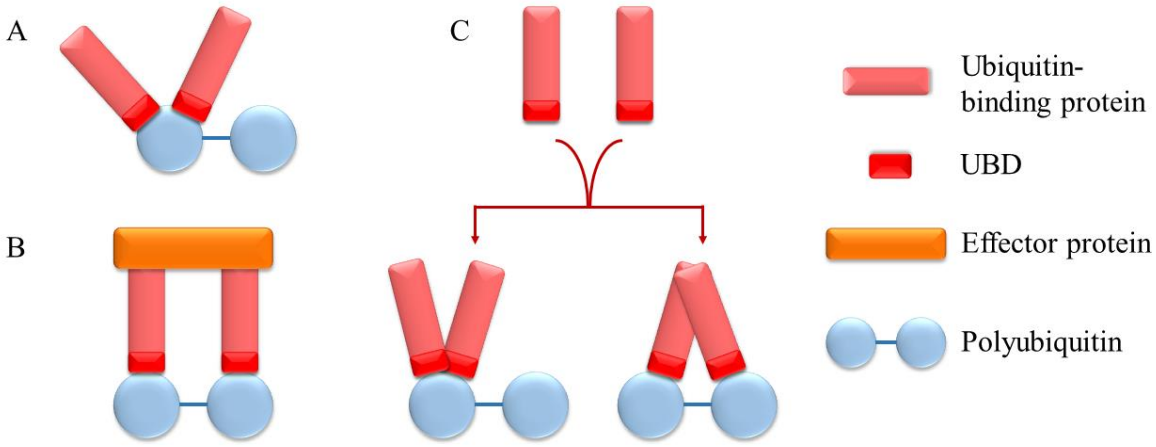


Fig.1.6.1.2. Theoretical binding arrangements of polyubiquitin with multiple UBD-containing proteins. A) One ubiquitin moiety mediates a ternary complex of non-interacting proteins. B) Ubiquitin acts as a scaffold to recruit proteins to indirectly interacting UBD-containing proteins. C) Two UBD-containing proteins oligomerise to specifically bind polyubiquitin.

The possibility of avidity arising from the binding of two separate UBD-containing proteins to one ubiquitin chain has been suggested but is still comparatively little studied (Sims and Cohen, 2009; Komander *et al.*, 2009; Wurzer *et al.* 2015). p62, a ubiquitin-binding protein linked to autophagy and several related diseases, contains a C-terminal UBA and has been shown to

oligomerise endogenously (Wurzer *et al.*, 2015). Kirkin *et al.* (2009) suggested that p62 oligomerisation might also contribute to ubiquitin binding as deletion of the Phox and Bem 1 (PB1) oligomerisation domain caused a decrease in ubiquitin binding. Wurzer *et al.* (2015) built upon this, investigating the effects of PB1-dependent oligomerisation on binding to Met1-, Lys48- and Lys63-tetraubiquitin. The PB1 domain proved important in binding to Met1- and Lys63-linked chains, but did not particularly affect Lys48-polyubiquitin binding, as determined by PB1-deletion mutants, suggesting that oligomerisation permitted the avidity that underpinned binding to Met1- and Lys63-polyubiquitin.

Intriguingly, oligomerisation showed avid selectivity towards Lys63-polyubiquitin, compared to the broader-binding mono-p62. Oligomerisation was found to stabilise the binding partners and protect the survival of the p62-ubiquitin complex. Data suggested that oligomerisation brought the UBA domains of each p62 into close proximity, which possibly underpinned this change in binding kinetics and selectivity. The p62-oligomer has been shown to bind multiple ubiquitin molecules simultaneously, which is potentially related to its function in autophagy, as proteins destined for degradation are frequently polyubiquitinated (Shabek *et al.*, 2012).

1.6.2 Polyubiquitin linkage selectivity

The mechanism underpinning the selectivity of a UBD to a certain polyubiquitin linkage is closely related to that of the selectivity induced by tUBDs. Whilst many proteins determine linkage selectivity through the arrangement of multiple UBDs, some UBDs are linkage-selective even in isolation. This demonstrates another method of ensuring specificity in the recognition of ubiquitin.

One example has already been mentioned in another context; the MyUb domain of myosin VI shows Lys63-polyubiquitin linkage selectivity even in the absence of other UBDs (He *et*

al., 2016). Another, better-studied, example is that of the Npl4 zinc finger (NZF) domain of TAB2.

Along with TAB3, TAB2 acts as an adaptor for TAK1 during NF-κB signalling (Xia *et al.*, 2009; Kanayama *et al.*, 2004). The adaptors specifically bind Lys63-polyubiquitin to promote autophosphorylation, oligomerisation and activation of TAK1. This enhances NF-κB signalling, which in turn regulates processes such as apoptosis, inflammation and immunity as discussed previously (Kanayama *et al*., 2004; Ea *et al.*, 2006; Wu *et al*., 2006; Skaug *et al.*, 2009).

The TAB2-NZF domain in general seems well structured for selectivity, as it is comparatively small and could easily fit between the different ubiquitin moieties within a polyubiquitin chain to make multiple contacts across the linkage (Moritsugu *et al.*, 2018). This was corroborated by pulldowns of a GST-fused C-terminal TAB2 region, containing the NZF, to Lys63-polyubiquitin but not to Lys48- or Met1-polyubiquitin (Komander *et al.*, 2009). This preference for Lys63-polyubiquitin has been demonstrated multiple times through different methods (Sato *et al.*, 2009; Mortisugu *et al.*, 2018; Kulathu *et al.*, 2009).

Structural analysis of this preference was conducted by Kulathu *et al.* (2009) by crystallisation of the TAB2-NZF in complex with Lys63-diubiquitin. The NZF sat between two ubiquitin moieties of a chain and simultaneously interacted with both, resulting in two NZF binding faces (Fig.1.6.2.1). The first binding face was situated on the side of NZF, centred around His678, Leu681 and Glu685 contacting the Ile44 patch and the Phe45-Gly47 loop of the proximal ubiquitin. The second binding face was slightly larger and contacted the Ile44 distal ubiquitin through NZF-Phe675. As a member of the ZnF UBD group, NZF domains also bind zinc across the second binding face, holding it in close proximity to the

Val70 and Leu8 of ubiquitin. The crystal structure also showed that the NZF does not interact with the diubiquitin linker region itself.

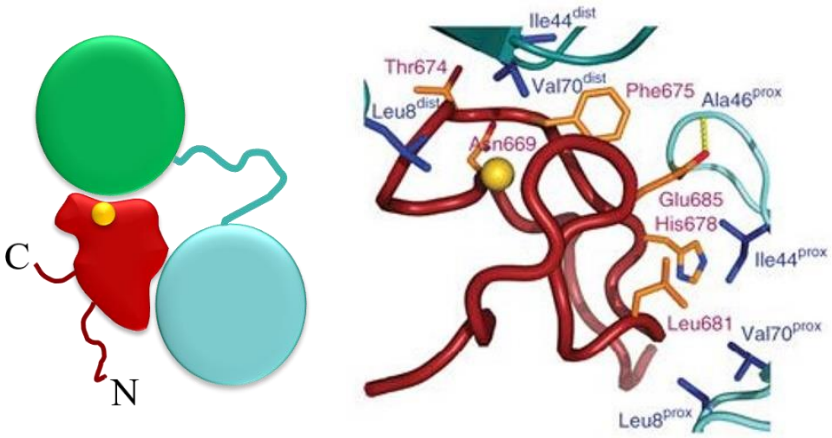


Fig.1.6.2.1. TAB2-NZF in complex with Lys63-diubiquitin. The NZF is shown in red, with zinc marked in yellow. To the left is a bubble diagram showing the orientation of the distal (green) and proximal (cyan) ubiquitin with NZF and zinc. To the right is the crystal structure and key interacting residues at the binding faces, adapted from Kulathu *et al.* (2009).

The interactions predicted by the crystal structure were confirmed through mutational analysis. Independent mutations of Phe675, His678, Leu681 and Glu685 to alanine were tolerated in binding to monoubiquitin. This supported the existence of two separate ubiquitin-binding faces, as each mutation only affected a single face leaving the other free to bind monoubiquitin. The mutations each had a dramatic effect on diubiquitin binding, however. The binding affinity of mutant TAB2-NZF was five- to eight-fold lower as determined by fluorescence anisotropy. This represented a K_D change from 8μM ±1.7 for wildtype-NZF to 40μM ±10 F675A, 34μM ±7 H678A, 35μM ±3 L681A, and 36μM ±1 E685A. As can be seen from the dissociation constants, the diubiquitin binding affinity was similar across all mutants. This suggested that even a single mutation was poorly tolerated by NZF binding to diubiquitin due to a loss of avidity effects. It is likely that these highly important residues form the basis of the structural selectivity of TAB2-NZF.

TAB2-NZF has also been linked to Lys48-diubiquitin binding in solution binding experiments (Kulathu *et al.*, 2009). It was suggested that this was made possible by the

flexibility of the diubiquitin linker, allowing a rotation of the proximal ubiquitin to bind both TAB2-NZF binding faces. However, these interactions could not be validated from cell lysate pulldowns and were of considerably lower affinity compared to Lys63-diubiquitin when assessed by *in vitro* methods, suggesting a low-level or physiologically less relevant interaction.

The correct binding of TAB2-NZF and Lys63-diubiquitin created a stable 1:1 complex where the NZF was held in place by the Ile44 binding patches of two linked ubiquitin moieties (Kulathu *et al.*, 2009). The Lys63-polyubiquitin chain is extended and open when compared to other linkages (Varadan *et al.*, 2004), permitting the NZF to access the binding pocket formed between the two Ile44 patches. Presentation of an alternative linkage with a more closed or folded structure, such as Lys33-polyubiquitin (Michel *et al.*, 2015; Kulathu *et al.*, 2009), would not permit stable binding, underpinning the linkage selectivity of the TAB2-NZF.

The minor binding to Lys48-diubiquitin prevents the TAB2-NZF from being entirely linkage specific but there was a clear linkage selectivity for Lys63-diubiquitin over mono- or Lys48-diubiquitin confers a linkage selectivity. It is possible that this preference fine tunes the binding of TAB2 to certain ubiquitination signals, and so localises or regulates the activation of TAK1 and NF-κB.

1.6.3 Polyubiquitin length selectivity

The mechanisms behind length-specific UBDs are closely tied to those behind linkage-specificity. Rather than the linkage positioning the polyubiquitin chain so that it can bind the UBD, in this scenario the chain must make multiple contacts with the UBD to ensure stable binding. Many ubiquitin-mediated systems have a preference for chains of a certain length, for example, efficient degradation by the 26S proteasome prefers the recognition of a polyubiquitin

chain of at least tetramer length (Thrower *et al.*, 2000), with the Rpn10-UIM showing an increased affinity for Lys48-tetraubiquitin compared to Lys48-diubiquitin as determined by Surface Plasmon Resonance (SPR) (0.20μM ±0.01 and 13.30 μM ±0.6 respectively) (Zhang *et al.*, 2009). Polyubiquitin length selectivity for most UBDs remains untested, or, if it has been investigated, only considers a few possible chain lengths. This is in part due to the relative unavailability of large quantities of defined length chains. As such, there may exist UBDs with high chain length specificity, but with the current tools available they remain undiscovered.

1.6.4 Selectivity for post-translationally modified ubiquitin

Whilst many UBD-containing proteins produce specificity through substrate assisted catalysis and specific sequence recognition, several UBD-containing proteins regulate selectivity by detecting crosstalk with other PTMs (Keusekotten *et al.*, 2013; Mevissen *et al.*, 2013). Crosstalk has been observed with phosphorylation (Cook and Chock, 1995), acetylation (Zong *et al.*, 2014) and SUMOylation (Uzunova *et al.*, 2007), and it is tempting to imagine that other examples may exist. Of these examples, the most characterised is crosstalk with phosphorylation, both of the UBD and of ubiquitin itself.

There are 11 possible phosphorylation sites on ubiquitin, nine of which have been observed experimentally in mammals; Thr7, Thr12, Thr14, Ser20, Thr22, Ser57, Tyr59, Ser65, and Thr66 (Ohtake *et al.*, 2015; Lee *et al.*, 2009; Shiromizu *et al.*, 2013; Bian *et al.*, 2014; Zhou *et al.*, 2012; Sharma *et al.*, 2014; Tsai *et al.*, 2015) (Fig.1.6.3.1). These are mostly uncharacterised, but pSer57 is linked to oxidative stress and has been identified as the most endogenously abundant (Peng *et al.*, 2003; Swaney *et al.*, 2015).

Separately, pSer65 has been estimated at 1.5% of all mammalian ubiquitin with links to mitophagy (Ordureau *et al.*, 2014; Kane *et al.*, 2014; Swaney *et al.*, 2015). pSer65-ubiquitin is further enriched in different cell compartments, with up to 20% of ubiquitin on the surface

of damaged mitochondria being phosphorylated at Ser65 (Ordureau *et al.*, 2014). The PTEN-induced putative kinase 1 (PINK1) from multiple species can phosphorylate Ser65 (Wauer *et al.*, 2015), suggesting endogenous importance and evolutionary conservation. It is also interesting to note that PINK1 can phosphorylate multiple Ser65 in a polyubiquitin chain, making it a polyubiquitin kinase and suggesting specific roles for distinct phosphorylation patterns across the polyubiquitin surface (Wauer *et al.*, 2015).

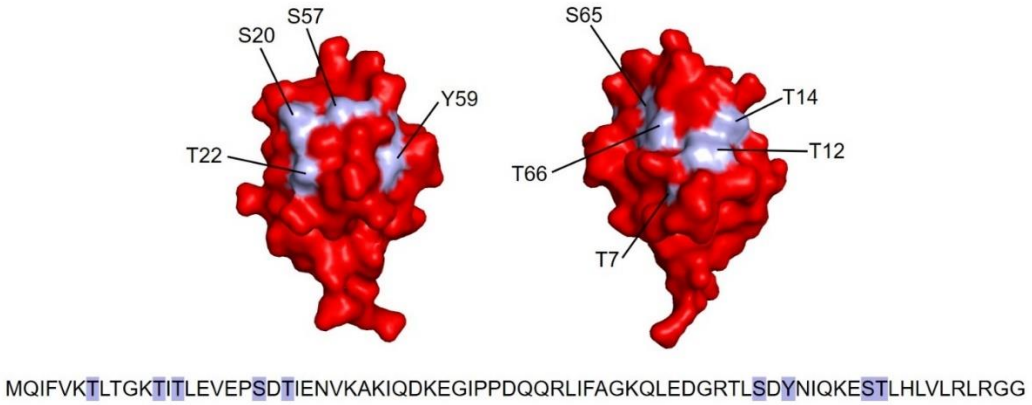


Fig.1.6.3.1. The phosphorylation sites of human ubiquitin. The nine observed sites (Ohtake *et al.*, 2015; Lee *et al.*, 2009; Shiromizu *et al.*, 2013; Bian *et al.*, 2014; Zhou *et al.*, 2012; Sharma *et al.*, 2014; Tsai *et al.*, 2015) were marked in blue. The human ubiquitin sequence is shown below with the same colour scheme. Figure created with PyMOL v2.3.2 from PDB 1UBQ (Vijay-Kumar *et al.*, 1987).

Crosstalk of phosphorylation and ubiquitination has been shown to make ubiquitin resistant to hydrolysis by DUBs (Wauer *et al.*, 2015), to be involved in mitochondrial clearance (Kondapalli *et al.*, 2012; Koyano *et al.*, 2014), and to be able to change the polyubiquitin selectivity of autophagic receptor UBDs, presumably in relation to mitochondrial clearance (Richter *et al.* 2016). Despite these hints at endogenous relevance phosphoubiquitin-selective UBDs are few and far between: the Ubiquitin binding in ABIN and NEMO (UBAN) domain of optineurin (OPTN) has been suggested to be sensitive to the phosphorylation state of ubiquitin, although this is contested (Richter *et al.*, 2016), and yeast Rad23 has been shown to prefer phosphomimetic over wildtype ubiquitin (Swaney *et al.*, 2015).

The first indication of Rad23 acting as a phosphoubiquitin-selective ubiquitin-binding protein came from a quantitative affinity chromatography experiment of yeast whole-cell extract (Swaney *et al.*, 2015). His-tagged yeast ubiquitin (yUb) and phosphomimetic S65E-yUb were used to pulldown UBD-containing proteins. Despite being conducted in yeast this was easily comparable to mammalian systems as yUb only differed from human ubiquitin by three amino acids (Sharp and Li, 1987). The majority of proteins bound in both conditions and were deemed to be unaffected by the phosphomimetic mutation (Swaney *et al.*, 2015). Several proteins, including Tom1, Cue5 and Vps27 showed slightly reduced binding to S65E-yUb. However, Rad23, which contained two UBA domains, preferred the phosphorylated ubiquitin substrate with a four-fold increase in unique peptides bound from yUb to S65E-yUb.

The same system was also used to suggest a physiological context for phosphoubiquitin. Various pharmacological stresses were applied to yeast cells expressing either wildtype or phosphomimetic ubiquitin and cell viability assessed. Cells expressing S65E-yUb were more susceptible to canavanine-induced protein misfolding and hydrogen peroxide-induced oxidative stress (Swaney *et al.*, 2015). The neighbouring S57E-yUb did not affect tolerance to either of these stresses, indicating that it was not generalised phosphorylation or ubiquitin mutation that caused the susceptibility, but rather that it was a specific effect of the phosphomimetic Ser65. Whilst no in-depth analysis of the mechanisms behind this susceptibility was conducted, it did provide a suggestion that pS65-yUb could be important for clearance of misfolded proteins and response to oxidative stress. This could feasibly be connected to the increased interaction of Rad23 to S65E-yUb, as Rad23 acts to shuttle ubiquitinated proteins for degradation by the proteasome and to excise areas of DNA damage (Schauber *et al.*, 1998; Dantuma *et al.*, 2009). This work provided proof of concept

of a pUb-selective UBD and also demonstrated the use of affinity chromatography in identifying previously unknown aspects of ubiquitin recognition.

The first thorough investigation into the existence of a mammalian UBD with the capability to differentiate between wildtype and phosphoubiquitin was based around OPTN (Richter *et al.* 2016). The autophagic receptor protein OPTN contains an endogenous tandem UBD; a C-terminal UBAN and ZnF, separated by approximately 60 residues (Laplantine *et al.*, 2009). As an autophagy receptor it recruits ubiquitinated cargo proteins to the autophagy machinery for degradation; such an important process requires tight control and specificity which is in part controlled by phosphorylation (Liu *et al.*, 2014). The study of crosstalk in this context has been especially interesting as both optineurin and ubiquitin can be phosphorylated during ubiquitin recognition.

Richter *et al.* (2016) investigated the effect of phosphoubiquitin on its recognition by the optineurin UBAN domain. Co-expression of phosphomimetic S65E- or S65D-ubiquitin with GFP-OPTN in HEK293 cells resulted in higher coprecipitation compared to wildtype ubiquitin. This indicated that optineurin showed a preference for phosphorylated ubiquitin. However, when combined with data from other experiments the effects of ubiquitin phosphorylation became less clear. Phosphorylation of ubiquitin at Ser65 by PINK1 was seen to increase mitochondrial recruitment of autophagy receptors in cell-based systems (Richter *et al.* 2016; Lazarou *et al.*, 2015; Heo *et al.*, 2015) but could not be seen *in vitro* (Richter *et al.* 2016; Heo *et al.*, 2015; Ordureau *et al.*, 2015). It may be that the *in vitro* experiment was missing a crucial interaction partner and that phosphorylation of ubiquitin alone could not bring about this effect, or that ubiquitin was phosphorylated in a different manner *in vitro* compared to endogenously.

Whilst phosphorylation of either ubiquitin or optineurin caused some increase in binding, the highest levels of interaction were seen when both binding partners were phosphorylated (Richter *et al.* 2016), suggesting that phosphorylation and ubiquitination crosstalk was possible at the ubiquitin and UBD level simultaneously.

Identification of OPTN phosphorylation sites was conducted by stable isotope labelling of amino acids in cell culture (SILAC)-based quantitative mass spectrometry (Richter *et al.* 2016). Of all the sites identified, only Ser473 within the UBAN domain was found to increase polyubiquitin binding in co-precipitation studies. This was Tank-binding kinase 1 (TBK1)-dependent and showed preference for longer and Lys63-linked chains, which only weakly bound unphosphorylated optineurin.

More recent work by Li *et al.* (2018) has begun to determine the structural basis for this effect. The OPTN-UBAN domain acts as a dimer that preferentially binds Met1- and Lys63-polyubiquitin in a 2:1 complex (Rahighi *et al.*, 2009; Lo *et al.*, 2009; Li *et al.*, 2018). The mouse optineurin Ser476 is equivalent to human Ser473 and when in complex with Met1-diubiquitin the Ser476 of one mOPTN was found to be in close proximity to the distal ubiquitin (Li *et al.*, 2018). The side chain of the Ser476 formed a weak hydrogen bond with Arg72 of the positively charged ubiquitin surface. As phosphorylation consists of addition of a negative phosphate group, pSer476 would strengthen the interaction with Arg72. This would stabilise binding between the optineurin dimer and Met1-diubiquitin to promote recruitment of the ubiquitinated proteins for autophagy. This was corroborated by Herhaus *et al.*, (2019) who used pS473-UBAN from OPTN and the corresponding A303E phosphomimetic UBAN domain of NEMO to increase binding affinity to Met1-polyubiquitin as observed by SPR.

Li *et al.* (2018) also suggested that phosphorylation of the OPTN UBAN could alter the binding specificity of the UBD, acting as a molecular switch to change optineurin's function. Whereas unphosphorylated OPTN-UBAN bound Met1- and Lys62-polyubiquitin, the phosphomimetic S476E mutation converted the mouse UBAN to a nonselective UBD, binding monoubiquitin and all polyubiquitins tested (monoubiquitin, Met1-, Lys48- and Lys63-diubiquitin) with equivalent dissociation constants as per ITC. This suggested a model whereby cargo proteins were Met1- and Lys63-polyubiquitinated, and PINK1 phosphorylated pSer65 to stabilise recruited OPTN. After initial OPTN recruitment, TBK1 phosphorylated the critical serine within the OPTN UBAN to switch specificity and recruit further ubiquitinated cargos without linkage restriction for amplified autophagic clearance.

An interesting paper has been recently published that may help explain how ubiquitin phosphorylation can affect its recognition by UBDs (Dong *et al.*, 2017). Using NMR the conformational state of phosphoubiquitin was determined and observed under different pH conditions. Phosphoubiquitin was found to exist in two conformations; relaxed and retracted. The former was similar to wildtype ubiquitin, and, due to the altered pK_a of the phosphoryl group, was protonated and enriched at acidic pH. The retracted state, however, was deprotonated and more likely to be enriched at basic pH. The retracted conformation buried more of the C-terminal tail, extended the Ser65 loop, potentially promoted ubiquitin phosphorylation, and altered the Ile44 patch (Dong *et al.*, 2017; Ye *et al.*, 2019; Kazansky *et al.*, 2018). This effect was enhanced by double phosphorylation, for example at both Ser65 and Thr66 by PTEN-induced novel kinase protein 1 (PINK1) (Ye *et al.*, 2019). As pH varies throughout the cell and under certain conditions, phosphoubiquitin could play a role in the detection of this fluctuation, fine tuning ubiquitin recognition and the processes that it regulates. One key example of this could be mitophagy, during which PINK1 phosphorylates Ser65-ubiquitin to promote interaction with the E3 ligase Parkin (Koyano *et al.*, 2014).

Healthy mitochondria are weakly alkaline (Llopis *et al.*, 1998), but during mitophagy they acidify (Wang *et al.*, 2012), presumably favouring the protonated/relaxed conformation of phosphoubiquitin which is most distinct from that of wildtype ubiquitin, affecting its recognition by UBDs and altering the binding or recruitment of effector proteins.

The relationship of Parkin with phosphoubiquitin is intricate. As an E3 ligase it decorates target proteins, such as Miro1, with ubiquitin and is involved in mitochondrial regulation, protein degradation and aggresome formation (Dawson and Dawson, 2010; Kahle and Haass, 2004; Lim *et al.*, 2006). Mutations of Parkin lead to early-onset Parkinson's disease and the function of this ligase is much studied within the literature (Kitada *et al.*, 1998). As mentioned above, PINK1 phosphorylates Ser65 of ubiquitin, which promotes pS65-ubiquitin interaction with Parkin, phosphorylation of Parkin Ser65 by PINK1 and ultimately leads to increased Parkin E3 ligase activity. Structurally, pS65-ubiquitin causes a straightening of the Parkin helix 3-In Between Ring (IBR) region, which loosens the interface between the IBR and the Parkin Ubl. This removes Parkin from autoinhibition (Chaugule *et al*., 2011) and shifts the Ser65 of Parkin to a more accessible position for phosphorylation by PINK1. From this it was suggested that a major function of pS65-ubiquitin could be for the promotion of Parkin phosphorylation and E3 ligase activity. The principal amino acids involved in ubiquitin-binding by Parkin have been identified through interrogation of crystal structures and ITC but the full UBD remains undefined (Kazlauskaite *et al.*, 2015; Kumar *et al.*, 2017).

Phosphorylation of ubiquitin also affects the activity of DUBs during mitophagy. The deubiquitinases ataxin-3, USP8, USP15 and USP30 all regulate Parkin-mediated ubiquitination but were unable to hydrolyse phosphoubiquitin (Wauer *et al.*, 2015). This demonstrates a physiologically-relevant function for phosphoubiquitin, where phosphorylation of ubiquitin at the mitochondrial surface promotes Parkin activity and

further ubiquitination whilst simultaneously protecting the phosphoubiquitin from degradation by DUBs prolonging the ubiquitin signal and ensuring clearance of the damaged mitochondria.

Despite the phosphoubiquitin having vast potential as a layered method of specific ubiquitin-recognition, little conclusive evidence has been found to suggest its use in endogenous cellular events. The concurrent phosphorylation of both ubiquitin and the OPTN-UBAN is likely the most understood example to date however this is contested (Richter *et al.* 2016; Li *et al.*, 2018). The yeast Rad23-UBA has been suggested to show preference for phosphoubiquitin but this requires further work (Swaney *et al.*, 2015). Given the variety within phosphoubiquitination it seems likely that further phosphorylation-specific or -sensitive UBDs exist but are yet undiscovered. The discovery of such phosphoubiquitin-selective UBDs would not only be of great interest to fundamental biology but would also allow the study of the physiological and pathological processes related to phosphoubiquitin, which currently remain little understood.

1.7 THE POTENTIAL FOR FUTURE SELECTIVE UBDS

Ubiquitination is a highly conserved and diverse mechanism of targeting proteins to a wide variety of fates and biological processes. For this reason, specific encoding and decoding of the ubiquitin signal is crucial for precise, timely and localised ubiquitin-mediated processes. Selectivity is induced at both the encoding and decoding levels by E3 ligases and UBDs within the ubiquitin-binding proteins. A variety of methods exist to engender specificity within ubiquitin recognition by UBDs, and it is likely that they combine with each other and with other methods that have yet to be discovered. Another alternative is that there remain novel UBDs that have not yet been identified.

This seems all too possible when one considers the number of ubiquitin-mediated processes compared to the comparatively low number of ubiquitin-binding domains. That there remain undiscovered aspects of specific ubiquitin recognition and decoding seems obvious, and yet the literature has been struggling to significantly advance in this area. The continued identification of novel UBDs, such as the CUBAN (Castagnoli *et al.*, 2019), CoCUN (Nepravishta *et al.*, 2019), SOUBA (Agromayor *et al.*, 2012), ZRF1-UBD (Richly *et al.*, 2010), UBW (Anindya *et al.*, 2010) and MyUb (He *et al.*, 2016) domains, provides convincing evidence and hope that we have not yet exhausted the capability for specific ubiquitin recognition. Of course, it may be that instead of looking for novel UBDs we should turn our attention to cooperative crosstalk between ubiquitination and other cell signalling methods, such as phosphorylation cascades, glutamate signalling, neurotransmitters, calcium signalling and membrane depolarisations, amongst many others.

To tackle such a challenge, a variety of approaches will be required, ranging from small-scale physiologically-relevant proteomics to high-throughput unbiased bioinformatics and Next-Generation Phage Display. These methods have all previously been used to good effect in the examples given above and should provide a comprehensive starting point for the investigation of novel aspects of specific ubiquitin recognition.

1.8 BACTERIOPHAGE AS AN EXPERIMENTAL TOOL

1.8.1 Basic bacteriophage biology

The use of bacteriophage and Next-Generation Phage Display in discovering novel ubiquitin interactions is intricate and arguably deserves its own dedicated thesis. The cross-disciplinary approach needed by this thesis prevents that, but an overview of bacteriophage, their uses in biotechnology and the combination of Next-Generation Sequencing (NGS) with Phage Display is introduced below.

Bacteriophages, or phages, are viruses that infect every known type of bacteria. They are the most numerous biological entities, being found in the most mundane to the most extreme environments. Phages are considered to be the great controllers of bacterial population and evolution, contributing to the movement of bacterial genes through the release of newly replicated virions from one infected bacterium to another (Bergh *et al.*, 1989; Brussow and Hendrix, 2002; Chibani-Chennoufi *et al.*, 2004; Brussow and Kitter, 2005; McCallin *et al.*, 2013; Santos *et al.*, 2014; Abedon, 2015; Pirnay *et al.*, 2015). Infection to bacteria is highly specific, with estimates of ten species of phage able to infect one bacterium. Specificity is ensured by the interactions between bacterial cell surface receptors and phage tail fibres (Trojet *et al*., 2011; Harada *et al.*, 2018; Kabanova *et al*., 2019).

Phage gross morphology is well defined and maintained across species, although there exist some exceptions to the following general rules. The genetic material can be DNA or RNA, double or single stranded and is enclosed within a protein capsid decorated with surface proteins which enable virion infectivity and release. Beneath this is a core channel surrounded by a spiral contractile sheath that work together to inject RNA or DNA into the bacterial host during infection. The channel ends in a baseplate hub which coordinates the movements of the surrounding tail fibres that detect and anchor to a prospective host (Terzaghi *et al.*, 1979; Rossmann *et al.*, 2005) (Fig.1.8.1.1). The commonly used M13 phage is filamentous in shape, formed of helically arranged pVIII major coat proteins containing the twisted loop of single stranded DNA (ssDNA) (Morag *et al*., 2015). There are five copies each of the minor coat proteins; at one end of the channel are the capping pIII and pVI proteins, and the at the other the replication initiating pVII and pIX proteins (Rakonjac *et al*., 1999; Smeal *et al*., 2017).

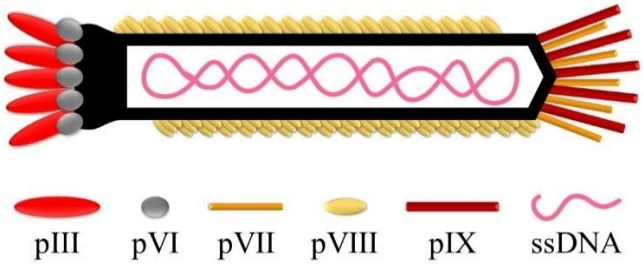


Fig.1.8.1.1. The generalised morphology of a bacteriophage with key proteins marked. There are three major compartments. First, the head of pIII and pVI proteins, then the channel (black) formed of the pVIII coat proteins and containing the ssDNA, and finally the tail of proteins pII and pIX.

Bacteriophage can be grouped by their life cycle; lytic, lysogenic or both (known as temperate). Lysogenic phages incorporate their genome into that of the bacterium, and so live along with the host, producing high numbers of phage particles throughout their lifespan. Whilst this does not cause lysis of the bacterium, it does have a detrimental effect on host growth. Lytic phages hijack the host replication machinery to produce their progeny, then lyse the host to release the new phages into the surrounding environment (Huber and Waldor, 2002).

1.8.2 Bacteriophage proteins and their uses

One phage commonly used in biotechnology is the non-lytic M13 phage. As a filamentous phage it infects bacteria with F pili, preferably gram-negative, including species frequently used in research, such as *Salmonella*, *Pseudomonas* and *Escherichia* (Deng *et al.*, 1999; Russel and Model, 2006; Haq *et al.*, 2012).

M13 have 11 phage proteins which are functionally divided into four groups; phage assembly and secretion (pI, pIV and pX1), DNA replication (pII, pV, pVI and pX), minor coat proteins for capsid assembly (pIII, pVII and pIX) and major coat protein for capsid assembly (pVIII). Both pIII and pVIII are the most commonly used to display randomised peptide libraries in Phage Display experiments.

The bacterial F pilus is the primary phage receptor, binding to phage pIII to begin transfer of the 6.4kb viral ssDNA into the host (Riechmann and Hollinger, 1997; Nilsson *et al.*, 2000),

This is converted into the double-stranded supercoil replicative form by the host topoisomerase, polymerases, RNA and DNA (Ray, 1969). The replicative form and pII are essential in replication and ligation of the phage DNA to form progeny phages (Russel *et al.*, 2004; Russel and Model, 2006). Both pV and pX regulate this process and preserve the quality of DNA replication. The progeny particles are then assembled by pI, pIV and pXI, which combine to create a channel in the host membrane through which the progeny are secreted and released. Provided that growth or replication are not impaired, approximately 1,000 phages are initially released, with 100-200 phages being released in every subsequent generation after infection (Webster, 1996). This allows for production of high titre phage cultures for use in experiments.

A diverse range of biotechnology applications have been developed from phage, including; non-chemical antibiotics, biosensors, food biopreservation and diagnostics, vaccinations, biofilm and plant pathogen control, DNA delivery, anti-corrosives (Cairns *et al*., 2009; Baldwin and Summer, 2012; Bartolacci *et al*., 2018; Namdee *et al*., 2018; Matamp and Bhat, 2019; Pinheiro *et al.*, 2019; Singh *et al*., 2019; Yang *et al.*, 2019b; Sedki *et al*., 2020). Of relevance to this thesis, one common application is the use of short peptide libraries to screen for new protein-binding peptides, either for therapeutic development or for fundamental biology.

1.9 NEXT-GENERATION PHAGE DISPLAY

1.9.1 Development of Next-Generation Phage Display

Phage Display was first described as an experimental technique by Smith in 1985, since then it has become a common method for epitope analysis, selection of synthetic peptides mimicking endogenous ligands (Scott and Smith, 1990; Cortese *et al.*, 1994), identifying pathogen-related proteins from blood sera, and providing lead peptides for further analysis or study (Folgori *et al.*, 1994; Cortese *et al.*, 1996; Germaschewski and Murray, 1996; Scala *et*

al., 1999; Enshell-Seijffers *et al.*, 2001; Mintz *et al.*, 2003). Indeed, one review named it the "most popular and successful surface display application" (Georgieva and Konthur, 2011) with over 9,000 entries in PubMed at the time of writing this thesis. This method allows the isolation of peptide sequences that bind a bait protein and directly links them to their encoding DNA. The bacteriophage model is resilient to genetic change, is stable in a variety of experimental conditions, produces a high titre of progeny phage particles and is fast growing (Dueñas and Borrebaeck, 1995; Ackermann *et al.*, 2004; Jepson and March, 2004; Lin *et al.*, 2010), with little specialised or expensive equipment required – although a significant amount of dedicated laboratory space is required to prevent cross-contamination. In recent years the combination of two powerful strategies, Phage Display and NGS has attracted great interest and scientific merit and is now commonly referred to as Next-Generation Phage Display (NGPD) (Rahim *et al.*, 2003; Dias-Neto *et al.*, 2009).

The basic NGPD experiment starts with creation of a phage culture displaying a randomised peptide library, with one unique peptide per phage. Each peptide is of defined length and can be either completely randomised or can be variations of a base protein. The peptide is fused to one of the phage coat proteins which span the capsid to extend into the surrounding microenvironment. Randomised peptide libraries, such as that used in the present study, are created by inserting random oligonucleotides between the endogenous coat protein and the leader sequence that guides the terminus of the protein out through the capsid.

Each phage will display only one type of peptide although the number of peptide copies per phage will vary depending on which surface coat protein is chosen. The commonly used pIII consists of 406 amino acids with N-terminal domains that bind the F pilus of the host bacteria to initiate infection and a C-terminal domain involved in phage assembly (Nilsson *et al.*, 2000; Rahbarnia *et al.*, 2017). Correct display of pIII is necessary for both phage particle assembly and infectivity, which can cause functional problems when a peptide library is fused to pIII

(Greenwood *et al.*, 1991). To overcome this, co-infection with an additional helper phage is often used, providing additional copies of wildtype pIII to support normal phage function whilst the endogenous pIII is used to display the library.

pIII-based libraries display fewer copies of the peptide per individual phage compared to the common alternative pVIII: there are only five copies of pIII compared to 2,700 copies of pVIII on the phage surface. Libraries based on pVIII are therefore more likely to support avidity or bipartite interactions with the target protein. On the other hand, pVIII is just 50 amino acids long, and so the length of the fused display peptide must be limited to avoid disrupted packaging and secretion (Greenwood *et al.*, 1991; Iannolo *et al.*, 1995; Kristensen and Winter, 1998; Sieber *et al.*, 1998; Cabilly, 1999; Barbas *et al.*, 2001; Baek *et al.*, 2002).

1.9.2 Conducting Next-Generation Phage Display

The phage library is panned against an immobilised target, which in the current study is ubiquitin. Phages that display ubiquitin-interacting peptides are recovered and amplified for further rounds of panning. Selection pressure can be increased between rounds to gradually select for the strongest binding sequences by increasing the number or detergent strength of washes applied to the bound phages during panning. Once the desired number of panning rounds has been completed, the phages are harvested, the peptide-encoding DNA amplified by PCR and used as the input for NGS (Fig.1.9.2.1). Sequencing data can be translated to amino acid sequences and the results interrogated for potential UBD-candidates.

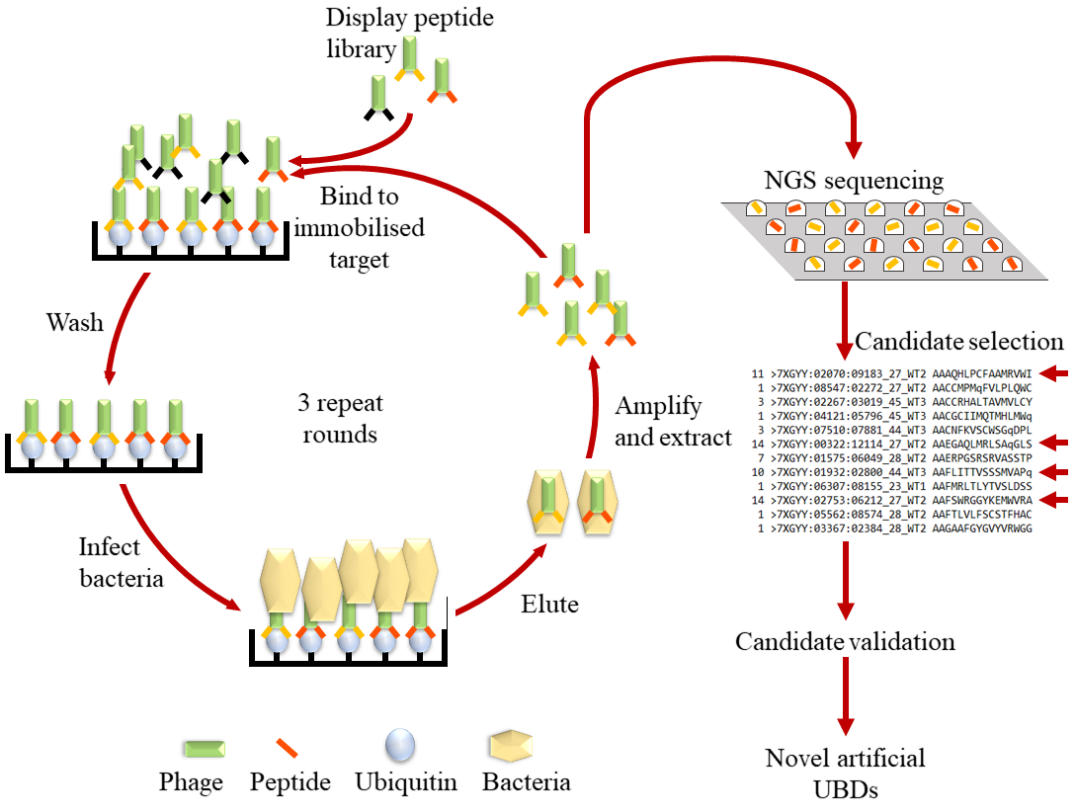


Fig.1.9.2.1. The basic Phage Display experiment leading to NGS of binding sequences. A naïve phage library was panned against immobilised ubiquitin, unbound phages removed through washing and bound phages harvested by elution into TG1 bacteria. These were amplified and used as the input for the next round. After three rounds, bacteria are harvested and the phage DNA sent for NGS and candidate selection.

The combination of traditional Phage Display with Next-Generation Sequencing maximises the high throughput capability of the method whilst maintaining comprehensive analysis of the panned library. Provided that the amount of DNA submitted for sequencing is within the readable limits every sequence in the sample is reliably detected. Although there remains the possibility that the DNA sample was not representative of the entire population and that some sequences had been missed.

The greatest advantage of Phage Display is that no knowledge of the randomised peptide sequence or structure is required – the peptides can be completely randomised removing any bias or restriction of scope imparted by the researcher.

A high binding affinity in the single to double-digit micromolar range is typical for interactions resulting from Phage Display. For example, sequences targeting the DUB USP11 achieved dissociation constants of 6.92 ± 2.25μM, Japanese encephalitis virus 6.06μM, and insulin-degrading enzyme 2.5μM (Zu *et al.*, 2014; Spiliotopoulous *et al.*, 2018; Suire *et al.*, 2018)

Overall, the NGPD system is well suited for discovery of protein-binding peptides. The basic bacteriophage biology permits high-throughput peptide discovery and provide a physical link between the binding peptide and the encoding genetic material. It's a relatively flexible system that can be tailored to suit a variety of targets, including ubiquitin.

1.10 PHAGE DISPLAY AND THE STUDY OF UBIQUITIN

To the author's knowledge there have been no published examples of using NGPD to identify novel UBDs. Indeed, until very recently there had been no attempt to use basic Phage Display to identify novel UBDs. Just this year Castagnoli *et al.* (2019) published the use of a lambda phage displayed human cDNA library to identify the CUBAN UBD. The scientific outcomes of this paper were discussed previously (section 1.5.1. The Unique CUBAN UBD) whereas the evaluation of the methodology is discussed here.

The library was designed from a human brain cDNA library, and as such consisted of peptide fragments of endogenous proteins. Whilst using a source library based on existing proteins meant that UBDs were highly likely to be identified by Phage Display, it did restrict the results to known proteins. This was unable to generate truly novel UBDs or identify non-endogenous binding peptides that could be developed into highly specific UBD-based tools. What is more, the approach taken by Castagnoli *et al.* did not exploit NGS and relied upon random selection of binding phages to identify candidate binding domains. Rather than sequence the entire binding library for identification of the best candidates, they grew bacteria infected with the panned phage on agar plates and selected random clones for individual PCR sequencing. Whilst

this approach yielded results it was inherently high risk, as random colony picking could have yielded only false positive clones or missed the strongest binding peptide. Combining Phage Display with NGS exploits the full high throughput power of the phage-based system and to restrict it by selecting candidates through clone picking loses much of the power of NGPD.

There are several instances of Phage Display being used elsewhere in ubiquitin biology, for example, to identify motifs representing deubiquitinase ligands, E3 ligase substrates, and sensors for ATG8 proteins (Guo *et al.*, 2013; Stolz *et al.*, 2017; Spiliotopoulous *et al.*, 2018).

Phage Display can not only be used to identify novel binding peptides but also to optimise the interaction of an existing binder by creating a library based on the interacting peptide. This can be used to optimise either an existing protein-protein interaction or the output from a preliminary round of Phage Display.

The UIM-ubiquitin interaction was artificially optimised through this method by Manczyk *et al.* (2017). A phage library was created displaying 9 $x10^9$ variations of ubiquitin with a diversified 2,000$Å^2$ surface area around the Ile44 patch targeting UIMs from the yeast Vps27 protein. This is a comparatively large surface area, considering that structural studies have indicated that the UIM-ubiquitin interaction only buries an area of around 490 ±90$Å^2$ on the surface of ubiquitin, and 470 ±70$Å^2$ on the UIM α-helix surface; although it should be noted that this figure is a comparison based on several studies with different UIM-containing proteins (Swanson *et al.*, 2003 and PDB 1Q0W; Hirano *et al.*, 2006 and PDB 2D3G; Raiola *et al.*, 2013 and PDB 2MBH). The ubiquitin-variant library was created using a soft randomisation method which is biased towards display of the wildtype ubiquitin but allows for small variations across the surface (Sidhu *et al.*, 2000; Fellhouse and Pal, 2005). 27 residues were targeted for randomisation, resulting in a library where each position was represented with approximately 50% mutation and 50% wildtype residues. Five panning rounds were

conducted against the first UIM of yeast Vps27 and the resulting candidates underwent DNA sequence analysis to give seven unique ubiquitin-variants with a consistent mutation of Gly10Val on the ubiquitin surface. These variants were screened with competitive phage ELISA and the variant which gave the highest affinity for the target UIM compared to controls (as indicated by IC_{50}) was taken forward for characterisation.

This high affinity variant, referred to as UbV.v27.1, contained 18 substitutions compared to wildtype, including the G10V mentioned above. Of these, only six were in close proximity to the UIM peptide as determined by a 2.35Å crystal structure of the UbV.v27.1-UIM complex: L8I, K48M, T66N, H68Y, L71S and R72S. 'Close proximity' was defined by the authors as being within 6Å. The interaction, unsurprisingly given that the authors focused on varying the ubiquitin surface around the hydrophobic Ile44 patch, was mostly through hydrophobic interactions, although there were some hydrophilic contacts observed. These modifications increased the *Kd* from 114μM to ~0.27μM, as determined by fluorescence polarization assays and competitive phage ELISA. Back mutations of the six close-proximity residues concluded that only H68Y contributed to this increased affinity. This worked cooperatively with the distant G10V mutation, as simultaneous back mutation of both resulted in a UbV.v27.1 with nearly no binding affinity for the UIM. Using a wildtype ubiquitin, the variant mutations G10V and H68T were independently introduced leading to 13- and 8-fold increases in affinity, respectively, as measured by fluorescence polarisation binding. This not only reinforced the importance of those two mutations for optimising ubiquitin and UIM binding, but also highlighted how a residue far from the binding interface can affect affinity.

Initially no clear structural reason could be found to explain the effects of the distant G10V mutation. The new amino acid had a higher molecular weight than the original (117kDa compared to 75kDa) (Lide, 1991) and was considerably more hydrophobic (hydrophobicity

index of 76 compared to 0 at pH7) (Monera *et al.*, 1995), however as this residue was far from the canonical hydrophobic binding site it was unlikely to be a straightforward interaction.

It wasn't until the second publication on the topic from the same laboratory (Manczyk *et al.*, 2019) that this was rationalised. The UbV.v27.1 variant was crystallised in the presence of the first UIM of yeast Vps27 (yUIM), and the Val10 was found to both support the formation of the UIM binding site and hydrophobically interact with the non-canonical binding surface of yUIM1. The G10V mutation also appeared to stabilise dimerization of UbV.v27.1 and promote a 2:2 binding ratio with yUIM1, allowing for avidity effects.

This piece of work, whilst highly interesting in its own right, was of particular relevance to the work proposed here. It not only starkly demonstrated the power of Phage Display to discriminate between single mutants within the context of ubiquitin-based NGPD but it also provided a hint as to how phosphorylation could affect the binding ability of ubiquitin. A concept similar to that of the G10V substitution could explain how the addition of the structurally small phosphate group to Ser57 or Ser65 could affect ubiquitin binding, as discussed earlier in the Introduction. The work discussed above also served to highlight the sensitivity and strengths of Phage Display, as such a modification was unlikely to have occurred to a researcher using rational design and structure-based processes to optimise ubiquitin and UIM interactions.

Despite never being reported to be used to target ubiquitin, NGPD has been used to identify novel peptides that bound the ubiquitin like domain of the USP11 deubiquitinase (Spiliotopoulos *et al.*, 2019). The ubiquitin-like domain (UBL) shares structural similarity with ubiquitin and some UBDs can also bind UBLs, for example, the UBA domain can also bind the ubiquitin-like FAT10 and the CUBAN domain of N4BP1 can bind both ubiquitin

and the ubiquitin-like NEDD8, as noted previously (Schmidtke *et al.*, 2006; Castagnoli *et al.*, 2019). The pVIII-displayed library was panned against immobilised peptide representing the USP11 DUSP and UBL domains in tandem for three rounds and the binding phage sent for NGS (Spiliotopoulos *et al.*, 2019). This resulted in two motifs; `[YNHC]-c-L-c-ε-R` and `L-x-L-ε-x-x-S-[RP]`, where the square brackets denoted conservation of any of the contained amino acids, `c` stood for a charged residue, `ε` for non-polar and `x` for any amino acid. Synthetic peptides containing this motif were tested for DUSP-UBL binding by ITC, with an average dissociation constant of 10μM. One of the peptides was found to be highly specific for the USP11 DUSP-UBL *in vitro*, as no binding was seen to the DUSP-UBL of USP4 or USP15, which share domain structure with USP11 (Harper *et al.*, 2011; Clerici *et al.*, 2014; Harper *et al.*, 2014). The interaction was localised to the UBL domain, in particular by the major binding pocket of the USP11 UBL, comprising His161, Asn162, Leu164, Gly206, Pro241, Gln244 and Glu246. This application of NGPD to UBL targeting demonstrated how a common motif could arise from the NGS data and be assessed for specificity and affinity to the target. This supported the possibility of using NGPD to identify ubiquitin-binding peptides.

In summary, the existing literature using Phage Display to identify and explore such a small modification suggested that NGPD could be a viable method to develop phosphoubiquitin-targeting UBDs.

1.11 THE DARK MATTER OF UBIQUITIN RECOGNITION

There are over 600 human E3 ligases (reviewed, Zheng and Shabek, 2017), ubiquitinating several thousand proteins (Chen *et al.*, 2014; Huang *et al.*, 2016) with chains of currently unquantifiable possible polyubiquitin variations (when taking into account the combinations of eight linkage types, the existence of branched and multi-linkage chains, the variable

length of a polyubiquitin chain, the modification of ubiquitin by other PTMs and the existence of multiple ubiquitin chains on a single target) which specifically direct their target protein to one of over 600 biological processes (Huntley *et al.*, 2014). All of this grand scope and variety is apparently mediated by 29 types of ubiquitin-binding domain. The maths of ubiquitination does not add up: logic, the need for specificity and the continued discovery of new UBD families all suggest that there remains much more to ubiquitin recognition that we do not yet understand (Fig.1.11.1).

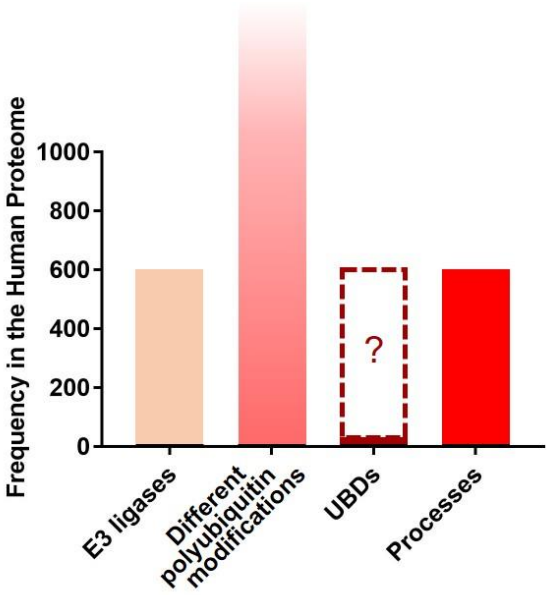


Fig.1.11.1. The undiscovered potential of ubiquitin-binding domains. The number of specific ubiquitination events and ubiquitin-mediated processes does not correlate to the number of UBDs able to decode this specific signal. This suggests the existence of either undiscovered UBDs or alternative mechanisms of ensuring UBD specificity. Frequency values are indicative rather than absolute and are collated from the data presented in Chapter 3.

This raises the obvious possibility of either undiscovered UBDs or alternative methods of fine-tuning specificity. This is the 'dark matter' of ubiquitination; we have described the gap that these new UBDs or alternative methods can fill yet we have not been able to elucidate them.

To take on such a challenge will likely require a diverse set of experiments with a multi-disciplinary approach. The discovery of the unique UBDs and the investigations into UBD

specificity described above (sections 1.5 and 1.6) provide some possible methods to tackle this problem. Phage Display was highly successful in identifying a novel UBD, the CUBAN domain, from a human cDNA library (Castagnoli *et al.*, 2019), suggesting that it would also be suitable for the high-throughput discovery of other novel UBDs. NGPD is analogous to peptide array experiments investigating binding of E3 ligases, which print peptide variants onto an immobilised surface such as a membrane and pan target protein (for example, E3 ligases or ubiquitin) over the surface to identify binding peptides. NGPD is considerably higher throughput, however, as standard arrays accommodate up to 600 variants per membrane (Klecker and Dissmeyer, 2016). An alternative method to the high throughput NGPD would be the classical use of ubiquitin to affinity pull-down ubiquitin-binding proteins from cell lysates, which was used in great success during the characterisation of Rad23 as a phosphoubiquitin-selective UBD (Swaney *et al.*, 2015). This has the advantage of being close to physiological conditions. Any protein suspected of containing a UBD could be aligned to known UBD sequences, similar to the discovery of the CoCUN (Nepravishta *et al.*, 2019) and UBW (Anindya *et al.*, 2010) ubiquitin-binding domains, although this does require a well-defined set of confirmed UBD sequences to conduct the alignment against. Finally, NMR has been used in multiple examples discussed previously to great effect and proved invaluable in the characterisation of the Rap80 UIM and its high affinity binding to ubiquitin (Anamika *et al.*, 2014).

And so this brings forth the major hypothesis of this thesis; that as of yet undiscovered families of UBDs and aspects of specific ubiquitin-recognition exist, such as crosstalk with other major cell signalling mechanisms like phosphorylation and calcium signalling. A combination of approaches will be required to work towards proving this hypothesis. Whilst this endeavour will doubtlessly continue for many years past the conclusion of this thesis,

as a starting point, bioinformatics, NGPD and classical proteomics will be combined to help elucidate this dark matter of ubiquitination.

Before any active investigation can take place a thorough examination of the current UBD population is required, and then different methods can be employed to identify novel synthetic and endogenous ubiquitin-binding proteins and UBDs. To achieve this, the following broad aims were set:

1. To assess the current landscape of known ubiquitin-binding proteins and domains and generate an updated resource for the discovery of novel instances of known ubiquitin-binding domains
2. To use high-throughput Next-Generation Phage Display to identify novel synthetic ubiquitin-binding peptides
3. To use brain lysate as a more physiologically relevant model in the identification of novel endogenous ubiquitin-binding proteins and potentially novel UBDs.

It is hoped that this will not only advance the current knowledge of ubiquitin-binding domains but will also be of use, and of interest, to other enthusiastic researchers looking to detangle the fascinating but annoyingly convoluted ubiquitin system.

2 MATERIALS AND METHODS

2.1 MATERIALS AND REAGENTS

For a list of all buffers, stocks and purchased reagents, please see Appendices.

2.2 ANTIBODIES

Primary antibodies

For Western blots the anti-ubiquitin and anti-peflin antibodies were prepared to 1:5,000 (v/v) dilutions in 5% Marvel 0.05% TBS-Tween 20, and the anti-GST was prepared to 1:10,000 in the same buffer.

For enzyme-linked immunosorbent assays (ELISA), anti-ubiquitin was prepared to 1:200 in 3% Marvel PBS.

Antibody	Manufacturer	Product code
Mouse anti-ubiquitin	LifeSensors	VU-101
Rabbit anti-peflin	Novus	NBP1-31613
Goat anti-GST HRP	GE Healthcare	RPN1236
Rabbit anti-USP5	Proteintech	10473-1-AP
Rabbit anti-p62	ENZO	PW9860

Secondary antibodies

All secondary antibodies were used at a dilution of 1:5,000 in 5% Marvel 0.05% TBS-Tween 20, except the anti-mouse for ELISA, which was used at 1:2000 in 3% Marvel PBS.

Antibody	Manufacturer	Product code
Swine anti-rabbit HRP	Dako	P0217
Rabbit anti-mouse HRP	Dako	P0260
Anti-mouse for ELISA	Invitrogen	A16163

2.3 PURCHASED PEPTIDES AND PROTEINS

Phage-derived peptides, Rabex-5-MIU peptide and pS65-ubiquitin were commercially sourced. All production, preparation and quality control was performed by the manufacturer. Peflin, wildtype ubiquitin and all other ubiquitin mutants were grown and purified as described below.

Lyophilised phage-derived and Rabex-5-MIU peptides were purchased as custom peptides from by Genosphere Biotechnologies (https://www.genosphere-biotech.com/) (Table 2.3.2.1). Each phage peptide was produced with the N-terminal 'Ala-Ser' bacteriophage Pel-B leader adduct and C-terminal amidation to more closely replicate display conditions on the phage. The Rabex-5-MIU was also produced with C-terminal amidation to allow it to act as a control for the phage-derived peptides.

Table.2.3.2.1. Specification of purchased phage-derived and Rabex-5-MIU peptides.

Peptide	Purity	N-terminal	C-terminal	Sequence 'N --- C'	g/mol
07	>95%	NH2	[nh2]	ASCYFRYHRVCIRSWEVQ	2,303
10	>95%	NH2	[nh2]	ASEACSVQWRKTRVHRDV	2,127
11	>95%	NH2	[nh2]	ASERRKPHGWTVIRCVRQ	2,179
36	>95%	NH2	[nh2]	ASRVSSRTAILTRNSAHV	1,925
39	>95%	NH2	[nh2]	ASSQWTRKQVQTSVMQSP	2,048
45	>95%	NH2	[nh2]	ASVHDAQWWQEMRLRANE	2,226
50	>95%	NH2	[nh2]	ASYLASPEALYRRLWTAP	2,064
MIU	>95%	NH2	[nh2]	QKQIQEDWELAERLQREE EEAFASSQS	3,265

Human recombinant phospho-Ser65-ubiquitin (#U-102) was purchased from Boston Biochem (www.BostonBiochem.com). It was phosphorylated enzymatically by PINK1 of the Red Flour Beetle (*Tribolium castaneum*). Purity was confirmed at >95%.

2.4 PRODUCTION OF PURIFIED PROTEINS

2.4.1 Ubiquitin production in bacterial culture

The yeast ubiquitin-encoding sequence was inserted into the pKK223-3 plasmid between EcoR1 and HindIII sites with ampicillin resistance in BL21 *E. coli* and back mutated to the human sequence; S19P, D22E and S28A (Sharp and Li, 1987). The phosphomimetic S57E and S65E mutations were introduced by site directed mutagenesis. Glycerol stocks of wildtype and phosphomimetic ubiquitin were kindly donated by Dr Jed Long, University of Nottingham.

15mL Luria Bertani Broth (LB) (Sigma, L3022) with 100µg/mL ampicillin was inoculated from ubiquitin glycerol stocks and cultures were grown for approximately 16 hours at 37°C shaking 180rpm to an OD_{600nm} of 0.6. The culture was centrifuged at 1,000xg for 10 minutes in a Thermo Scientific Heraeus Primo R centrifuge and the pellet resuspended in 25mL minimal media containing 100µg/mL ampicillin, trace elements, vitamins and 1mL/L 100x MEM Vitamins (Sigma #M6895) (Appendix A2.1). Minimal media was used to reduce contaminants. This was grown for a further 30 minutes then used to inoculate a 1L flask of minimal media with the same additives. After growth at 30°C to an OD_{600nm} of 0.7, the culture was induced with 200µM isopropyl β-d-1-thiogalactopyranoside (IPTG) and incubated for 16 hours at 30°C 180rpm. The pellet was harvested by centrifuging at 4,000xg in a Beckman Avanti J-25 I centrifuge with JLA 10.500 rotor for 30 minutes and frozen at -20°C overnight to aid lysis.

Uniformly labelled ^{15}N-ubiquitin was grown to the same protocol, except 1g/L ^{15}N ammonium chloride was in the trace elements instead of unlabelled ammonium chloride.

2.4.2 Ubiquitin purification strategies

The pellet was resuspended in 10mL 5x AE buffer (Appendix A2.1) per 1L original growth and lysed on ice by 30:30 seconds ON/OFF sonication for a total of 10 minutes at 10 microns

using a Soniprep 150 sonicator kept in a 6°C cold room. The lysate was clarified by centrifugation at 30,000xg for 30 minutes at 4°C in a Beckman Avanti J-26S XP centrifuge with JA 25.50 rotor. Additional 5x AE buffer was added to the supernatant until no precipitate was seen and centrifuged again to the same specifications.

The supernatant was diluted 1:4 with degassed deionised water and loaded onto a 5mL HiTrap SP HP sepharose cation exchange column (GE Healthcare, 17115201) pre-equilibrated in 1x AE buffer using an ÄKTA Start chromatography system. Proteins were eluted from the column by increasing the salt concentration and therefore increasing buffer conductivity. Conductivity was increased by increasing the proportion of 1x AE High Salt buffer (Appendix A2.1) from 0 to 30% over a 300mL elution. 10mL fractions were collected and presence of protein indicated by UV absorbance at 280nm. Ubiquitin elutes at approximately 140mL as standard, and this was confirmed by separating fraction samples on SDS-PAGE 5-20% gels and Coomassie staining (Sections 2.5.2 and 2.5.4, respectively). Fractions containing protein of interest, i.e. of the correct molecular weight, were frozen at -80°C and freeze dried for two days.

Lyophilised fractions were resuspended in 1-2mL deionised water, centrifuged at 3,000xg for 10 minutes in a Thermo Scientific Heraeus Primo R centrifuge to remove precipitates, and gel filtrated using a HiLoad 26/600 Superdex 75 pg column (GE Healthcare, GE28-9893-34) and ÄKTA Prime Plus chromatography system. Proteins were loaded onto the column in Gel Filtration buffer (Appendix A2.1) and eluted over 400mL in 10mL fractions at 1mL/minute. Ubiquitin eluted at approximately 240mL as standard. Presence of protein was determined by UV absorbance at 280nm and was confirmed by running samples on SDS-PAGE 5-20% gels and Coomassie staining. Fractions of interest were pooled, frozen to -80°C and freeze dried for two days.

The protein was resuspended in 2mL deionised water and centrifuged at 3,000xg for 10 minutes in a Thermo Scientific Heraeus Primo R centrifuge to remove precipitates. This was loaded onto a HiTrap 5mL Desalting column (GE Healthcare, GE29-0486-84) using an ÄKTA Prime Plus chromatography system and desalted into degassed deionised water in 10mL fractions. Presence of protein was determined by UV absorbance and salt by conductivity. Fractions suspected of containing ubiquitin were confirmed by SDS-PAGE and Coomassie staining. Confirmed fractions were pooled and freeze dried for two days and the amount of protein weighed and stored at -20°C until use.

Approximately 5-10mg of wildtype ubiquitin could be expected from each litre of growth. This was slightly lower for mutants and labelled proteins.

2.4.3 Peflin growth in bacterial culture

2.4.3.1 Peflin sub-cloning

The plasmid containing the PEF1 gene encoding full length human peflin was purchased from Harvard PlasmID (https://plasmid.med.harvard.edu/PLASMID/) (clone ID HsCD00325639). It was created by Dr Gerry Rubic, University of California, Berkeley CA, USA, and the original vector by Dr Michael Brownstein, National Institute of Mental Health, NIH, Bethesda, MD, USA.

The host strain was DH5-alph T1 phage resistant bacteria with chloramphenicol resistance, with PEF1 contained within the pOTB7 vector using the Xho1 site (Fig.2.4.3.1.1). The peflin insert was subcloned into the pGEX-4T2 vector appended to the GST tag. All reactions were conducted in a thermocycler with heated lid.

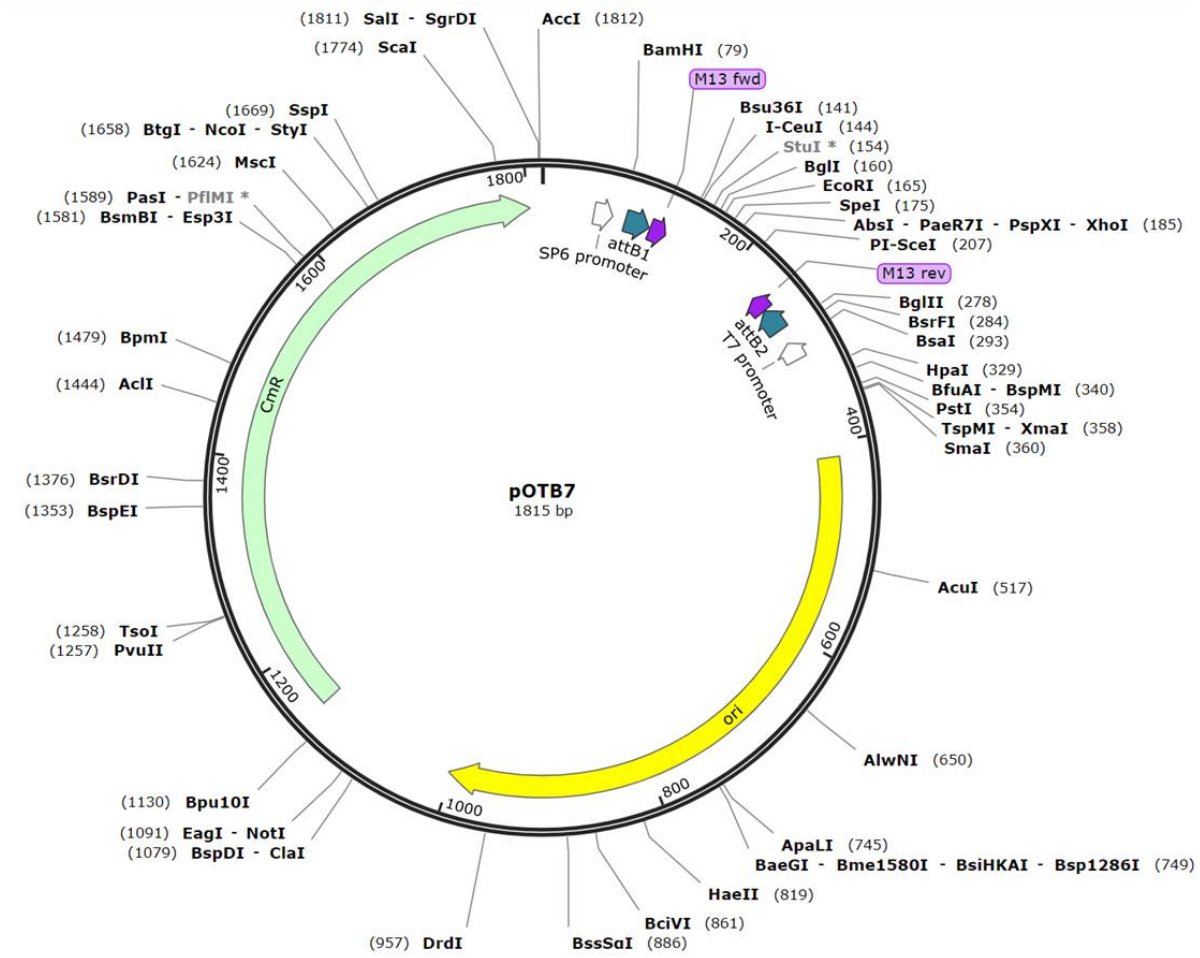


Fig.2.4.3.1.1. The commercially available pOTB7 vector. The peflin plasmid was purchased from Harvard PlasmID Repository within the pOTB7 vector, ligated at the Xho1 site with chloramphenicol resistance. The map is reproduced from the Snapgene Repository (www.spapgene.com).

The glycerol stock provided by Harvard PlasmID was grown overnight under the recommended conditions, 37ºC in LB supplemented with 30µg/mL chloramphenicol. The DNA was extracted using a standard QIAprep Spin Miniprep kit (QIAGEN, #27104) as per the manufacturer's instructions and eluted into 30µL deionised water. Yield was determined by Nanodrop spectrophotometry.

A 20µL double digest was run for 2hrs at 37ºC with Xho1 and EcoR1 restriction enzymes using the CutSmart buffer (all New England Biolabs, #R0146S, #R3101S, and #B7204S, respectively). A double digest of the receiving pGEX-4T2 vector (kindly donated by Dr Simon Dawson, University of Nottingham) was run alongside using the same protocol.

	Volume (µL)
Plasmid	5
dH20	12
Xho1	0.5
EcoR1	0.5
CutSmart buffer	2

The plasmid was then CIP treated (New England Biolabs, #M0290S) to prevent vector re-annealing and the digested product gel purified by DNA electrophoresis (Section 2.5.1) alongside an uncut plasmid control sample on a 1% gel. The insert was recovered using a standard QIAquick Gel Extraction kit (QIAGEN, #28704) according to the manufacturer's instructions and quantified by Nanodrop spectrophotometry.

The insert and receiving pGEX-4T2 vector with ampicillin resistance were ligated using a 3:1 molar ratio with 50ng of vector and T4 system (T4 DNA ligase #M0202S, and T4 10x buffer #B0202S, both New England Biolabs). Ligation into pGEX-4T2 appended a GST tag prior to the PEF1 insert.

	PEF1	Control
Water	12.96µL	14.12µL
T4 10x buffer	2µL	2µL
T4 DNA ligase	0.5µL	0.5µL
pGEX-4T2	3.38µL	3.38µL
Insert	1.17µL	-
Total	20µL	20µL

Ligation was confirmed by size confirmation in DNA electrophoresis on a 1% gel (Section 2.5.1) and sequencing of the ligated plasmid (University of Nottingham DeepSeq) with M13 forward and reverse primers;

M13 forward	GTA AAA CGA CGG CCA GT
M13 reverse	GGA AAC AGC TAT GAC CAT G

The GST-peflin pGEX-4T2 plasmid was heat shock transformed into Rosetta *E. coli* for improved rare codon usage. Transformed cultures were selected by growth on LB agar plates containing 100µg/mL ampicillin and colonies picked to inoculate 10mL LB agar with

100µg/mL ampicillin. This was grown to an OD_{600nm} of 0.6 at 37°C in a shaking incubator at 180rpm. Glycerol stocks were taken and stored at -20°C until use.

Rosetta GST-peflin was grown, expressed and purified through glutathione-Sepharose affinity chromatography and thrombin cleavage as described below. Liquid Chromatography with tandem mass spectrometry (LC-MS/MS) identification of the expression product excised from a Coomassie stained SDS-PAGE gel, and detection by Western blot with anti-Peflin (Novus Biologicals, #NBP1-31613) both confirmed presence of in-frame and expressed peflin protein.

2.4.3.2 Peflin growth

Starter cultures were grown from peflin glycerol stocks in 10mL LB supplemented with 30µg/mL chloramphenicol or 100µg/mL ampicillin, as appropriate, in a shaking incubator at 37°C 180rpm to an OD_{600nm} of 0.6. This was used to inoculate the final scaled up growth medium of the same composition. The final culture was grown under the same conditions as the starter culture until it reached an OD_{600nm} of 0.6. It was then induced by addition of 200µM IPTG for 3hrs at 37°C shaking at 180rpm.

For small scale 25mL culture, the culture was centrifuged at 720xg for five minutes. The pellet was resuspended in 1mL 0.1% TBS Triton X-100 pH 7.5 with 1:1,000 phosphatase inhibitor cocktail 2 and protease inhibitor cocktail (Sigma-Aldrich, #P5726 and #P8340, respectively) and lysed on ice. The lysate was centrifuged at 15,700xg for 10 minutes at 4°C and the pellet discarded.

Large scale 600mL cultures were grown in 2L flasks with IPTG induction overnight at 20°C 180rpm to improve aeration and protein expression. The culture was centrifuged at 4,000xg in a Beckman Avanti J-25 I centrifuge with JLA 10.500 for 20 minutes, and the supernatant discarded. The pellets were combined by resuspension in a total of 50mL 0.1% TBS Triton X-

100 pH 7.5. This was centrifuged at 720xg for 5 minutes and the supernatant discarded. The single combined pellet was frozen at -20°C overnight to aid lysis.

The next day, the frozen pellet was defrosted on ice and resuspended in 30mL 0.1% TBS Triton X-100 pH 7.5 with 1:1,000 phosphatase inhibitor cocktail 2 and protease inhibitor cocktail and lysed on ice by 30:30 seconds ON/OFF sonication for a total of 10 minutes at 10 microns using a Soniprep 150 sonicator kept in a 6°C cold room. The lysate was clarified by centrifugation at 30,000xg for 30 minutes at 4°C in a Beckman Avanti J-26S XP centrifuge with JA 25.50 rotor.

2.4.3.3 Peflin purification strategies

For purification of GST-peflin, Glutathione-Sepharose (GE Healthcare, #17-0756-01) was pre-equilibrated to 0.1% TBS Triton X-100 pH 7.5 according to the manufacturer's instructions at a 1:1,000 ratio of Sepharose to original growth culture volume. Protein was separated from the beads by either reduced glutathione elution with 0.5 volumes of reduced glutathione elution buffer (Appendix A2.1) or thrombin cleavage. The amount of thrombin was proportional to the volume of growth culture, with 5 units per 1L growth culture. Presence of protein was confirmed by anti-peflin Western blot.

2.4.4 GST-p62 growth in bacterial culture

Glycerol stocks of XL10-gold pGEX-T41 GST-p62 were kindly provided by Dr Alice Goode (University of Nottingham). Starter cultures were grown from glycerol stocks, amplified to 25mL cultures and purified using the same method as described for peflin. Total protein was determined by BCA and presence of GST-p62 confirmed by SDS-PAGE and Western blotting.

2.5 GENERAL BIOCHEMISTRY METHODS

2.5.1 DNA electrophoresis

1% or 3% (w/v) agarose TAE gels were cast with 5µL SYBR Safe (Thermo Fisher, #S33102) per 100mL. DNA samples and high or low molecular weight marker (both New England Biolabs, #B7025 and #N0557S, respectively) were diluted 6:1 with DNA gel loading dye (New England Biolabs, #B7024S) and run at 70 V constant until sufficiently separated and imaged using a Gene Genius bioimaging system (Syngene).

2.5.2 SDS-PAGE

SDS-PAGE was used to separate proteins before Western blot or Coomassie staining. All gels were cast in house using resolving Gel buffers A and B (Appendix A2.1) and Gel stacking buffer (Appendix A2.1) in a 1mm plate system. A continuous 5-20% acrylamide gradient for improved protein separation was created across the gel using a gradient mixer. The volumes shown below are for one gel.

	5% gel	20% gel	Stacking gel
Gel buffer A	-	3.33mL	-
Gel buffer B	3.33mL	-	-
Gel stacking buffer	-	-	5mL
Acrylamide	1.67mL	6.67mL	1mL
Water	5mL	-	-
SDS	100µL	100µL	-
APS	100µL	100µL	100µL
TEMED	10µL	10µL	10µL
Volume	10.21mL	10.21mL	6.11mL

Protein samples were diluted 1:1 with gel loading buffer (GLB; Appendix A2.1) and boiled for five minutes at 95ºC in a heat block. Once cooled they were loaded alongside BLUeye Prestained protein molecular weight marker (Geneflow, #S6-0024) and the gel electrophoresed in Electrode buffer (Appendix A2.1) for ~2hrs or until the dye front met the bottom of the gel at 40mV (EC13590 powerpack, Thermo).

2.5.3 Western blot

Proteins were transferred from the SDS gel to 0.45μm nitrocellulose membrane (Amersham Protran, GE Healthcare) pre-soaked in Transfer buffer (Appendix A2.1). The membrane and gel were protected with filter paper (Whatman) and proteins transferred overnight at 40mA at room temperature in Transfer buffer.

Successful transfer was confirmed by Ponceau-S (Appendix A2.1). Then the membrane was blocked in 5% Marvel 0.05% TBS-Tween 20 for 1hr at room temperature and probed with the appropriate primary antibody for 1hr at room temperature. The membrane was then incubated with the appropriate horseradish peroxidase (HRP) secondary antibody under the same conditions. The only exception was anti-GST which was already HRP conjugated and did not require a secondary. All antibodies were prepared in 5% Marvel 0.05% TBS-Tween 20 to the dilutions shown below.

Primary antibodies	
Mouse anti-ubiquitin	1:5,000
Rabbit anti-peflin	1:5,000
Rabbit anti-USP5	1:5,000
Rabbit anti-p62	1:5,000
Goat anti-GST HRP	1:10,000

Secondary antibodies	
Swine anti-rabbit HRP	1:5,000
Rabbit anti-mouse HRP	1:5,000

Antibodies were detected using the Western lighting Plus Enhanced Chemiluminescence (ECL) system (Perkin Elmer) with imaging film (Kodak) according to the manufacturer's instructions. Exposure time varied according to detection levels but did not exceed 25 minutes. The film was then manually developed with PQ Universal developing solution and Hypam fixer solution (#1757314 and #1758285, Ilford).

2.5.4 Coomassie staining

Non-specific protein was detected in SDS-PAGE gels by soaking the gel in Coomassie Blue stain (Appendix A2.1) for 1hr at room temperature on a plate rocker. The used stain was discarded, and the gel soaked in destain (Appendix A2.1) under the same conditions.

2.5.5 Bicinchoninic acid (BCA) assay

BCA assays were conducted using the microplate procedure according to the manufacturer's instructions with the Pierce^{TM} BCA Protein Assay Kit (Thermo Scientific, #23225).

2.5.6 Obtaining porcine brain tissue lysates

Porcine tissue was used in affinity chromatography and LC-MS/MS experiments. It had high homology to both the human genome and proteome and provided a readily available model for the mammalian proteome. Porcine brain halves were obtained from the local abattoir and frozen at -20°C until use.

Myelin-rich tissue was removed by dissection on ice. The unmyelinated white tissue was homogenised in 10mL capture buffer per 10mg of tissue with 1:1,000 protease and phosphatase inhibitors using an Ultra Turrax T25 homogeniser (Janke and Kunkel). The lysate was centrifuged at 20,000xg for 20 minutes at 4°C to yield approximately 10mL of supernatant.

The total protein was determined by BCA and aliquots stored at -20°C until use.

2.5.7 Cyanogen Bromide-activated (CNBr)-Sepharose affinity chromatography

2.5.7.1 Standard CNBr-Sepharose protocol

Sepharose was always prepared fresh as required. One bead volume of dehydrated Cyanogen Bromide-activated (CNBr)-Sepharose 4B (GE Healthcare, #17043001) was hydrated and coupled to bait protein according to the manufacturer's instructions. The CNBr coupling, ethanolamine, tris and acetate buffers are listed in Appendix A2.1.

To capture the prey protein, bait-coupled beads were equilibrated in the appropriate assay buffer before addition of prey protein in assay buffer. Capture was conducted by incubation at room temperature for 2hrs with end-over-end rotation. Samples of flow through buffer containing unbound protein were collected. Bound protein was eluted from the beads by boiling samples in an equal volume of GLB (Appendix A2.1) for five minutes.

Samples were taken at every stage to confirm coupling and capture. These samples as well as the flow through samples received the GLB treatment.

All samples were run on SDS-PAGE and presence of protein determined by Western blot or Coomassie staining. As standard, 70µL of sample was loaded per well, representing 100% of the bead sample and 50% of the flow through.

2.5.7.2 CNBr-Sepharose capture of pig brain lysates

CNBr-Sepharose was coupled to bait ubiquitin at 2mg/mL and prepared for prey protein capture as described above.

The total protein of pig brain lysed in capture buffer was determined by BCA. 6.7mg of total protein in capture buffer was added to 140µL ubiquitin-Sepharose. Where appropriate calcium chloride or EDTA was added to a final concentration of 250µM or 1mM, respectively. The samples were incubated for 2hrs at room temperature, collected and processed as described above. SDS-PAGE gels were loaded with 70µL of beads and 35µL of flow through.

This general protocol aligned well with previous experiments within the laboratory, although the buffer composition was altered to remove previous optimisation for known ubiquitin-binding proteins. This included a high level of glycerol and Triton X-100 for stabilisation and to solubilise poorly soluble proteins, and low salt to help preserve electrostatic protein complexes.

2.5.7.3 CNBr-Sepharose capture of recombinant protein lysates

CNBr-Sepharose was coupled to bait ubiquitin at 2mg/mL and prepared for prey protein capture as described above.

For experiments comparing ubiquitin-Sepharose capture of GST-peflin or GST-p62 from cell lysates the amount of total cell protein in over expression culture lysate was not determined and was not equalised between conditions. This was due to the differences in recombinant protein expression between cell lines. Instead, protein input was equalised qualitatively so that there were equivalent amounts of GST-peflin and GST-p62 loaded onto the ubiquitin-Sepharose, rather than equal amounts of total protein. Samples from each over expression lysate were separated by SDS-PAGE and Coomassie stained. The volume of XL10gold GST-p62 lysate was then proportionally adjusted so that the amount of GST-p62 present was equivalent to that of GST-peflin found in 50μg/25mL of Rosetta GST-peflin over expression lysate.

50μg of total Rosetta protein and the adjusted amount of XL10-gold total protein in capture buffer was added to 50μL ubiquitin-Sepharose. Where appropriate calcium chloride, magnesium chloride or EDTA was added to a final concentration of 250μM, 250μM or 1mM, respectively. The samples were incubated for 2hrs at room temperature, collected and processed as described above. SDS-PAGE gels were loaded with 50μL of beads and 35μL of flow through.

2.5.7.4 CNBr-Sepharose capture of purified peflin

CNBr-Sepharose was coupled to bait ubiquitin at 2mg/mL and prepared for prey peflin capture as described above.

Peflin was purified from 25mL of Rosetta culture as described above through either reduced glutathione elution to yield GST-peflin or thrombin cleavage to yield cleaved peflin. The

amount of peflin purified was determined by culture volume rather than by concentration to remain comparable to previously conducted lysate experiments where 25mL of culture was used as prey sample. This 25mL had previously been shown by BCA to equate to 50µg of total protein. The purified peflin in capture buffer was added to 35µL ubiquitin-Sepharose. Where appropriate calcium chloride was added to a final concentration of 250µM. The samples were incubated for 2hrs at room temperature, collected and processed as described above. SDS-PAGE gels were loaded with 35µL of beads and 35µL of flow through.

2.5.8 LC-MS/MS

LC-MS/MS was used to identify all peptides within a sample, either for confirmation of protein expression or for identification of interaction partners after affinity chromatography. The samples were run on an SDS-PAGE gel until they had passed through the stacking gel and had just formed a defined band within the resolving gel. This was Coomassie stained, excised and sent to the Cambridge Centre for Proteomics (CCP) Core Facility, University of Cambridge, for analysis. The samples were digested with trypsin by the CCP and the run time was extended to 2hrs to reduce sample complexity and increase the number of positive protein identifications.

Data received from CCP were processed with Scaffold Viewer v4. The total spectrum data were analysed for the number of times a peptide had been identified within the sample. The peptide and protein thresholds for identification of proteins was set to 95%, allowing high confidence that the detected peptide correlated to the called protein. The minimum threshold for determining the presence of a protein within the sample was set to a minimum of one. This was lower than the standard minimum threshold (two) to allow detection of infrequently or previously missed proteins.

2.6 NEXT-GENERATION PHAGE DISPLAY

Next-Generation Phage Display (NGPD) of a randomised peptide library kindly donated by Dr Alison Gray (founder, AFABILITY, France and visiting researcher, University of Nottingham, UK) was conducted to identify short ubiquitin-binding peptides as putative UBDs. The library targeted wildtype (WT)-, phosphomimetic S57E-, S65E- and pS65-ubiquitin immobilised on 96 well plates. At every stage correct immobilisation of ubiquitin to the plate surface was confirmed by ELISA. All steps were carried out in sterile conditions unless stated otherwise.

2.6.1 ELISA

ELISA was conducted to confirm even immobilisation of ubiquitin on the plate surface. Autoclaved sterile PBS was used for all stages and all incubations were conducted with a plate lid to prevent contamination. 100µL of 1µg/mL WT-ubiquitin in PBS was immobilised per well of the Nunc Immobilizer Amino 96 well plate (Thermo Fisher, #436006) by incubation for 1hr at room temperature. Wells were blocked with 400µL 3% Marvel in PBS for 1hr at room temperature then incubated with anti-ubiquitin primary antibody at a dilution of 1:200 in 3% Marvel in PBS for 1hr at room temperature. The secondary anti-mouse Alkaline Phosphate antibody (Invitrogen, #A16163) was added at 1:2,000 in 3% Marvel in PBS for 1hr at room temperature. The fluorescent substrate was *p*-nitrophenyl phosphate (pNPP) (Thermo Fisher Scientific, #37620) prepared according to the manufacturer's instructions and was exposed to the captured antibodies for 10 minutes at room temperature to allow for fluorescence to develop. Plates were read using a VARIOSKAN FLASH spectrophotometer (Thermo Scientific) at 405nm. Fluorescence was measured in arbitrary absorbance units and normalised to a blank PBS control.

2.6.2 Preparation of TG1 stocks

A glycerol stock of TG1 *E. coli* with ampicillin resistance was kindly donated by Dr Alison Gray (as above). This was streaked onto a M9 minimal media agar plate (Appendix A2.1) containing 150µg/mL ampicillin and grown overnight at 37°C; minimal media was used to select for the F' pilus that allows bacteria to be infected by phage. Single colonies were used to inoculate 20mL of 2YT (Appendix A2.1) then grown at 37°C 180rpm to OD_{600nm} 0.6. Glycerol stocks were prepared at a final concentration of 25% glycerol, aliquoted and frozen at -80°C until use. New TG1 selection plates and cultures were prepared every week.

2.6.3 Preparation of helper Exphage stocks

Helper Exphage were co-cultured with the phage library and TG1 to help infectivity as display of the library impairs bacteriophage growth.

TG1s were prepared on minimal agar as described above. A single colony was used to inoculate 10mL of 2YT which was grown at 30°C 220rpm overnight. This starter culture was used to inoculate 1 litre 2YT and grown under the same conditions to OD_{600nm} 0.4-0.6. 10µl of the helper Exphage with kanamycin resistance glycerol stock (Gough lab, University of Nottingham School of Veterinary Medicine and Science) was added per 1L of TG1 culture. This was incubated at 37°C static for 30 minutes to allow regrowth of pili, then 30 minutes at 37°C 220rpm. A further 1 litre of 2YT was added and the culture split between two 2-litre flasks. Kanamycin was added to 50µl/mL and the cultures grown overnight at 30°C, 180rpm.

The next day, a new colony of TG1 was grown to OD_{600nm} 0.5-1.0. The Exphage culture was centrifuged at 3,000xg 4°C for 20 minutes and the pellet discarded. The Exphage was precipitated from the supernatant by lowering to pH 4.2 with hydrochloric acid diluted 1:1 with dH_2O. This was incubated on ice for 60 minutes. Precipitated Exphage was collected by centrifuging at 8,000xg 4°C for 30 minutes. The pellet was re-suspended in 80mL of the

supernatant, transferred to two Falcon tubes and centrifuged again at 3,000xg 4°C for 10 minutes. The supernatants were discarded and all traces of 2YT removed by two further centrifugations at 3,000xg 4°C for 5 minutes, discarding the supernatants between spins. The pellets were softened by standing for 10 minutes in 25mL sterile PBS each, and then gently re-suspended. 2M sodium hydroxide was used to readjust the suspension to pH 7.0-7.4, taking care to swirl the liquid to prevent denaturation of the Exphage. Any bacterial proteins that were carried through with the Exphage were removed by centrifugation at 8,000xg 4°C for 1 minute and discarding the pellet. Glycerol stocks were prepared to a final concentration of 25% glycerol and stored at -80°C until use.

To check for absence of contamination and to determine the concentration of the Exphage, stocks were titred according to the dilutions below (Fig.2.6.3.1) and grown on kanamycin 2YT agar plates overnight at 37°C, with appropriate control plates.

Plate	Titration		Dilution
1	1µL neat phage + 100µL PBS	+ 900µL TG1	10^{-2}
	Whilst in PBS		
2	Dilute 1µL in 100µL PBS	+ 900µL TG1	10^{-4}
	Whilst in PBS		
3	Dilute 1µL in 100µL PBS	+ 900µL TG1	10^{-6}
	Whilst in PBS		
4	Dilute 1µL in 100µL PBS	+ 900µL TG1	10^{-8}
	Whilst in PBS		
5	Dilute 1µL in 100µL PBS	+ 900µL TG1	10^{-10}
	Whilst in PBS		
6	Dilute 1µL in 100µL PBS	+ 900µL TG1	10^{-12}

Fig.2.6.3.1. Titration method for determining concentration of input phage. Serial dilutions were conducted of the prepared phage solution into PBS. After dilutions were complete, 900µL of TG1 culture was added to each sample. These were grown on agar plates and the number of colonies counted and the representative pfu/mL calculated.

If contamination was seen on any of the control plates then the entire batch was discarded. If a plate grew less than 30 colonies it was discarded. The plate with the most distinct colonies was counted and each colony was equated to one Exphage. The concentration of the Exphage preparation was then calculated as 6 $x10^{12}$ pfu/mL.

2.6.4 Preparation of the M13 bacteriophage library

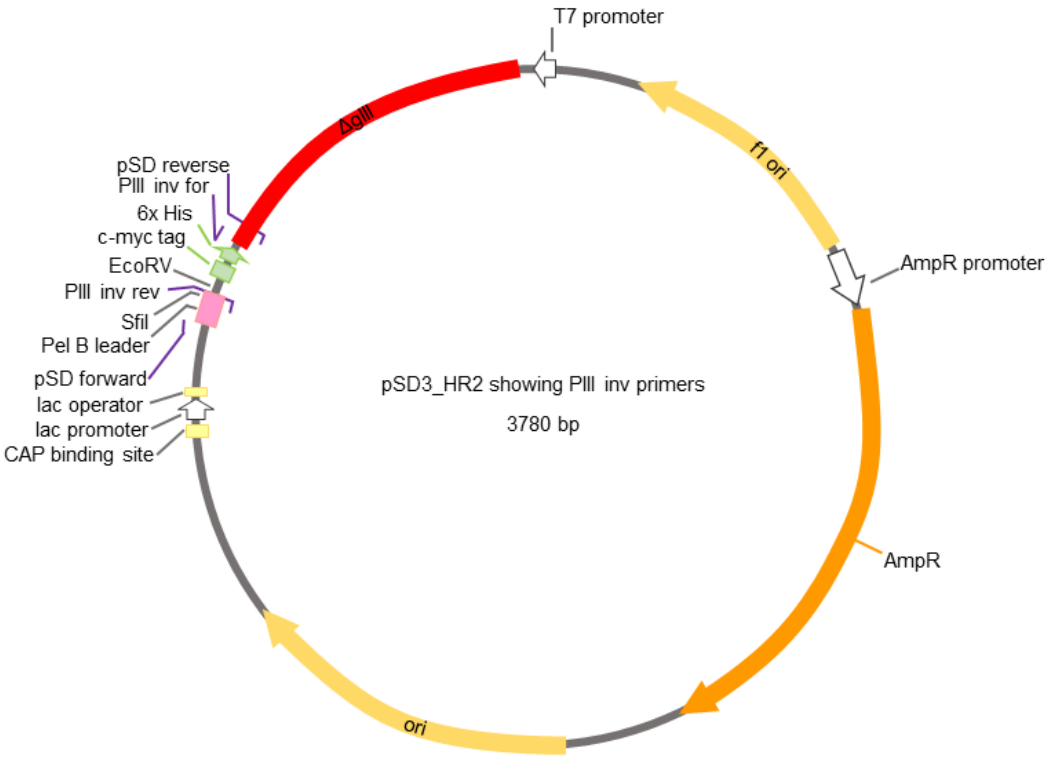
A TG1 *E. coli* stock infected with the naïve M13 phage library displaying a pIII 16mer peptide was kindly provided by Dr Alison Gray. Whilst the creation of the library did not form part of this thesis an overview of the process is included below for completeness and as it informs the results presented in Chapter 4.

The library was created by saturation mutagenesis (also known as hard randomisation) of the pSD3_HR2 vector (created in house) using inverse primers with (NNK)16 ligated into the Sfi1 site, between the C-terminus of pIII and the N-terminus of the PelB leader sequence (Fig.2.6.4.2). The NNK system creates a DNA sequence where N = A/C/G/T and K = G/T, resulting in 32 possible codons that produce a randomised peptide with a length of (NNK)*n*, where *n* = the number of amino acids encoded. Each of the positions can be occupied by any one of 21 possible amino acids, with a total of 21^{16} possible combinations for a 16mer peptide. Post-production analysis of the library determined the actual number of variants at 10^7, which is a considerable, albeit expected, loss of variation.

Using the NNK system as opposed to NNN reduced the occurrence of stop codons: of the 32 codons permitted by NNK only one represents a stop equivalent to a premature stop probability of 3.125%. The NNK method also reduced the skew towards amino acids which have more common codons, with a ratio of common to rare codons of 0.09:0.03 compared to 0.10:0.02 for the alternative NNB system, where B = C/G/T (Nov, 2012).

During the construction of this library the Amber Stop system was used. This includes an amber stop TAG codon between the pIII and library peptide encoding regions. When present in a non-supressing bacterial strain, such as HB21551, the TAG codon is read as a stop codon which 'turns off' expression of the pIII fusion. This results in expression of pIII alone which can be of use during the construction of the library. When present in a supressing strain however, such as TG1, the TAG is read through as a Gln permitting expression of the pIII fusion peptide for use in panning. During sequencing this Gln resulting from the amber stop was denoted by a 'q', whereas as the natural Gln codons were denoted by 'Q' (Carmen and Jermutus, 2002).

The final pSD3_HR2_(NNK)16 vector was transformed into TG1 *E. coli* to result in a protein product formed of the pIII-(NNK)16-PelB in a M13 library of 10^7 16mer variants (Fig.2.6.4.1).



GATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTT
TTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCT
GATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAA
CGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCC
TCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAA
GCGGGCAGTGAGCGAGTACCCGATAAAAGCGGCCTCCTGACAGGAGGCCGCTTTGNTNGCA
GCCCACCTCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACT
TTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAA
CAGCTATGACCATGATTACGCCAAGCTCGAAATTAACCCTCACTAAAGGGAACAAAAGCTG
GCCACCGCGGTGGCGGCCGCAAATTCTATTTCAAGGAGACAGTCATA**ATGAAATACTTACT**
GCCGACTGCAGCTGCAGGCTTACTGTTACTGGCGGCCCAACCGGCCATGGCTATGGATATC
CCAGTTGGTGGGTGCTGAACAGAAACTGATCTCTGAAGGAGACCTGAACGGTGCTCACCAC
CATCACCACCATACTAGTTCTAGTGGTGGCGGTGGCTCTCCATTCGTTTGTGAATATCAAG
GCCAATCGTCTGACCTGCCTCAACCTCCTGTCAATGCTGGCGGCGGCTCTGGTGGTGGTTC
TGGTGGCGGCTCTGAGGGTGGTGGCTCTGAGGGTGGCGGTTCTGAGGGTGGCGGCTCTGAG
GGAGGCGGTTCCGGTGGTGGCTCTGGTTCCGGTGATTTTGATTATGAAAAGATGGCAAACG
CTAATAAGGGGGCTATGACCGAAAATGCCGATGAAAACGCGCTACAGTCTGACGCTAAAGG
CAAACTTGATTCTGTCGCTACTGATTACGGTGCTGCTATCGATGGTTTCATTGGTGACGTT
TCCGGCCTTGCTAATGGTAATGGTGCTACTGGTGATTTTGCTGGCTCTAATTCCCAAATGG
CTCAAGTCGGTGACGGTGATAATTCACCTTTAATGAATAATTTCCGTCAATATTTACCTTC
CCTCCCTCAATCGGTTGAATGTCGCCCTTTTGTCTTTGGCGCTGGTAAACCATATGAATTT
TCTATTGATTGTGACAAAATAAACTTATTCCGTGGTGTCTTTGCGTTTCTTTTATATGTTG
CCACCTTTATGTATGTATTTTCTACGTTTGCTAACATACTGCGTAATAAGGAGTCTTAATC
ATGCCAGTTCTTTTGGCTA

Fig.2.6.4.1. Circular vector map and DNA sequence of the original plasmid, pSD3_HR2. The pIII_InvPCR_for and pIII_InvPCR_rev primer sites are marked for inverse PCR, and the pIII region is identified as ΔgIII. In the DNA sequence, the Sfi1 restriction site is underlined.

The pIII primers used for inverse PCR are shown below, where all sequences are displayed left to right as 5' to 3', introducing a Sfi1 restriction site and (NNK)16. Lower case text represents the annealing region.

pIII_InvPCR_for:

Forward	CAGGCCCAACCGGCCATGGCTgcatct-(NNK)16-cggagttctagtggtggcggtggctctcc
Reverse	cctctcggtggcggtggtgatcttgaggc-tctacgTCGGTACCGGCCAACCCGGAC
Complement	GTCCGGGTTGGCCGGTACCGAcgtaga-gcctcaagatcaccaccgccaccgagagg
Reverse complement	ggagagccaccgccaccactagaactccga-gatgcAGCCATGGCCGGTTGGGCCTG

pIII_InvPCR_rev:

Forward	CAGGCCGGTTGGGCCGCCAGTAACAGTAAGCC
Reverse	CCGAATGACAATGACCGCCGGGTTGGCCGGAC
Complement	GTCCGGCCAACCCGGCGGTCATTGTCATTCGG
Reverse complement	GGCTTACTGTTACTGGCGGCCCAACCGGCCTG

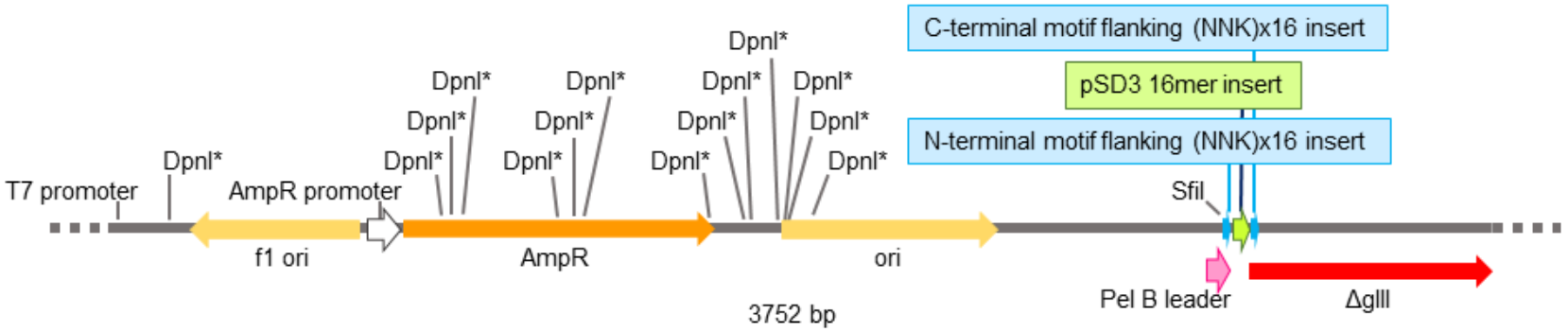


Fig.2.6.4.2. Linear vector map for the ligated pSD3_HR2_(NNK)16 vector. The pelB leader is marked in pink, the pIII region is identified as ΔgIII and the 16mer insert is highlighted in green with blue flanking regions. This shows the orientation of the final fusion protein product; PelB-16mer-pIII.

The library in TG1 was grown at 37ºC in a shaking incubator at 180rpm in 2YT media containing final concentrations of 150µg/mL ampicillin and 1% glucose in a baffled 2L flask. The culture for 3 hours after reaching an OD_{600nm} of 0.4 then spun down by centrifuging for 10 minutes at 3,000xg. The supernatant was discarded and the pellet suspended in 10mL 2YT with 25% glycerol, aliquoted and stored at -80ºC until use. Growth was not measured by titration as this library was never used neat but was instead grown on from glycerol stocks as required.

2.6.5 Phage display and bio-panning method

Three rounds of panning were performed, with the output from Round 1 becoming the input for Round 2 and so on. Selection pressure for the most selective binders was increased by including an additional 10 washes each round; 5 of buffer with detergent and 5 of plain buffer.

2.6.5.1 Preparation of Round 1 phage

The method of library and helper phage growth and preparation for panning was broadly identical to that of preparation for creating stocks (Sections 2.6.2 and 2.6.3). To prevent repetition, a brief summary will be given below with differences highlighted.

The naïve library was grown from a previously prepared glycerol stock to an OD_{600nm} of 0.4-0.6 as described above in 500mL 2YT media containing 150µg/mL ampicillin and 1% glucose in a baffled 2L flask in a shaking incubator at 180rpm and 37ºC. 500µL helper Exphage was added to achieve a multiplicity of infection (MOI) of 10. This was based on the generalised assumption that OD_{600nm} 1.0 equates to $8x10^8$ cells and that OD_{600nm} 0.5 equates to $4x10^8$ cells. Therefore, to achieve a MOI of 10, $4x10^9$ phage particles were required, equivalent to a 1:1,000 dilution from the $6x10^{12}$ pfu/mL Exphage stock. The culture was incubated for 30 minutes at 37ºC static and 30 minutes shaking at 180rpm. Cultures were centrifuged at 3,000xg for 10 minutes, the supernatant discarded and the pellet rinsed with 2YT before another 1 minute spin at 3,000xg and removal of the second supernatant. This ensured that no glucose remained. The

pellet was then re-suspended in double the original volume of 2YT containing 150µg/mL ampicillin and 50µg/mL kanamycin for selection of the library and helper phages, respectively, and grown overnight at 30ºC at 200rpm.

The phages were collected and prepared for panning. From this point on the cultures were no longer kept under sterile conditions and experiments were carried out on the lab bench. The culture was centrifuged for 20 minutes at 4,000xg at 4ºC. As previously described, the supernatant was collected, the phages acid precipitated, pH adjusted to 7.0-7.4 and any bacterial proteins that had also precipitated were removed by centrifugation. The phage concentration was determined by titration and remaining phages were stored at 4ºC until use.

2.6.5.2 Determining phage concentration by titration

Titration dilutions were prepared according to the diagram previously described (Fig.2.6.3.1) under sterile conditions with *E. coli* TG1 cultures used at an OD_{600nm} of 0.5-1.0. The dilutions were incubated for 45 minutes at 37ºC static to allow for infection then spread on 2YT agar plates containing 150µg/mL ampicillin and 1% glucose, allowed to dry then grown at 37ºC overnight. The plate with the most distinct colonies was counted and each colony was equated to one phage, as infection by one M13 bacteriophage prevents subsequent multiple infections. The concentration of the phage preparation was then calculated at pfu/mL.

2.6.5.3 Round 1 bio-panning

All subsequent steps were carried out on the benchtop in non-sterile conditions unless stated, although all reagents and buffers were kept sterile until the time of use. In brief, phage preparations were panned against immobilised ubiquitin, the binding phage recovered for amplification and preparation for another round of panning (Fig.2.6.5.3.1). Three rounds were completed in total and the eluates from each sent for NGS to identify binding peptides.

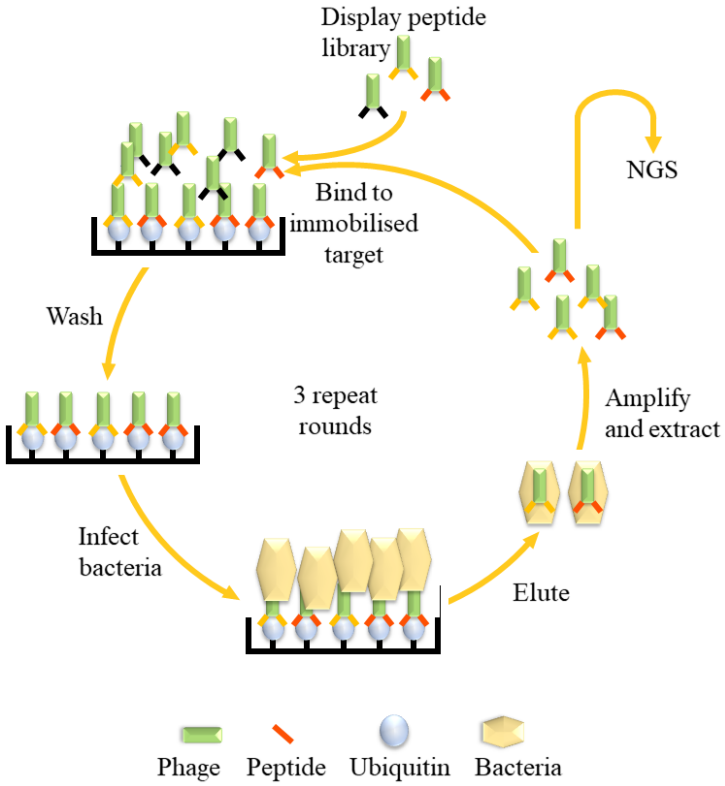


Fig.2.6.5.3.1. An overview of phage panning against immobilised ubiquitin. The targets were WT-, S57E-, S65E- and pS65-ubiquitin. Three sequential rounds were conducted. Samples from each round were sent for NGS.

Nunc Immobilzer Amino 96 well plates (Thermo Fisher Scientific) were coated for 1hr with 100µl of 10µg/mL of the target ubiquitin protein in PBS; final available protein per well was 1µg. The targets were WT-, pS65-, S57E- or S65E-mono ubiquitin as specified. A negative control was included by coating the wells with PBS only. This will be referred to as 'no protein' or 'Marvel' control, as any phage that result from this condition will have bound to the Marvel plate block.

Typically, ten wells were coated for panning per condition and an eleventh for ELISA. After the hour incubation a further 100µl PBS was added for 30 minutes to increase the surface area coated and to ensure that all the potential reactive coupling had taken place. As per the plate manufacturer's suggestions a pre-wash step of was included after coating the wells with peptide; wells received three washes of TBS Triton X-100 0.05% (v/v). All wells were blocked with 400µl 3% Marvel in PBS and the phage blocked with 200µl 18% Marvel in PBS per mL

of phage supernatant. The blocked phage was kept on a roller mixer to prevent settling. The ten experimental wells were incubated with 300µl blocked phage for 1hr, and the remaining eleventh wells reserved for ELISA. Excess or weakly bound phage were aspirated, the wells washed five times with TBS Triton X-100 0.1% (v/v), and then a further five with TBS.

All the following steps were carried out under sterile conditions. Remaining phage were eluted by adding 300µl of TG1s at an OD_{600nm} of 0.5-1.0 and incubating at 37ºC static for 30-45 minutes. The infected TG1s were collected and all ten wells of each condition were pooled, giving 3mL total for each condition.

Of this 1µL was titred to determine the eluted phage concentration according to the figure below and spread on 20mL 2YT agar petri dishes containing 150µg/mL ampicillin and 1% glucose (Fig.2.6.5.3.2).

Plate	Titration	Dilution
1	10µL neat eluted in TG1	neat
2	Dilute 10µL in 90µL PBS	10^{-1}
3	Dilute 10µL in 90µL PBS	10^{-2}

Fig.2.6.5.3.2. Titration method for determining concentration of eluted output phage.

The remaining eluate was spread on a Nunc bioassay plate (Thermo Fisher) of 150mL 2YT agar containing 150µg/mL ampicillin and 1% glucose and incubated overnight at 37ºC. Colonies from the Round 1 bioassay plates were harvested in 5mL 2YT with 1mL 50% glycerol for analysis by NGS and an aliquot was taken for use as Round 2 inputs.

2.6.5.4 Subsequent bio-panning rounds

Round 2 phage library preparation, titration and bio-panning were carried out to the same method as Round 1. The only differences were that the aliquoted eluate from Round 1 was used instead of naïve phage as the input library, that plates coated with blocked immobilised target

ubiquitin were washed 10 times with TBS Triton X-100, then a further 10 with TBS, and that subtractive blocking was included for the S57E-, S65E- and pS65-ubiquitin conditions. This involved addition of 10μg/mL WT-ubiquitin to the phage preparation blocking step to pre-bind any phage displayed peptides that bound sites that were conserved between wildtype and phospho/phosphomimetic ubiquitin. This left only phage that bound to the area surrounding Ser57/Ser65 and therefore were likely to represent phospho-selective binding peptides. The addition of subtractive panning was denoted by '-sub' in the experiment name.

Again, Round 3 was carried out to the same protocol as Round 1 and 2, except for the eluates from Round 2 were used as the input library for Round 3, and the plates received 15 TBS Triton X-100and 15 TBS washes. Subtractive blocking was included as described above. As there was no Round 4 the whole glycerol eluate from Round 3 was stored for DNA preparation and NGS.

Using the eluate from the previous round as the input for the next allowed for selection of the ubiquitin-binding phage. This selection pressure was increased between rounds by the rising number of washes.

2.6.5.5 Phage Display experimental design

Two sets of experiments were carried out. The first consisted of the phage library panned against WT-, S57E- and S65E-ubiquitin with Marvel controls for three rounds. Each Round was panned once (Fig.2.6.5.5.1).

Whilst Set 1 successfully yielded phage-infected TG1 colonies after three round the levels of growth were inconsistent between conditions, days and rounds. This was supported by anecdotal evidence from others in the lab who had seen the same variability day to day, despite every effort to replicate conditions exactly. To protect against this, Set 2 made use of repeated pans, where the same phage preparation was panned twice and the elutates combined. This protected the experiment from stalling due to failure of one condition.

The second set consisted of the same starting library panned against WT-, pS65- and S65E-ubiquitin with Marvel controls for three rounds. Round 1 was panned once, and both Rounds 2 and 3 were panned twice. In this case the Round 2 input phage (eluate from Round 1) was grown and prepared once, and this was panned once as per the protocol above and then again the next day under identical conditions. The eluates from the two repeats were kept separate. Round 3 was grown from an equal mixture of both repeated eluates, maintaining the same total volume of input phage. Again, Round 3 phage was grown and prepared once, then panned twice and the eluates kept separate (Fig. 2.6.5.5.2).

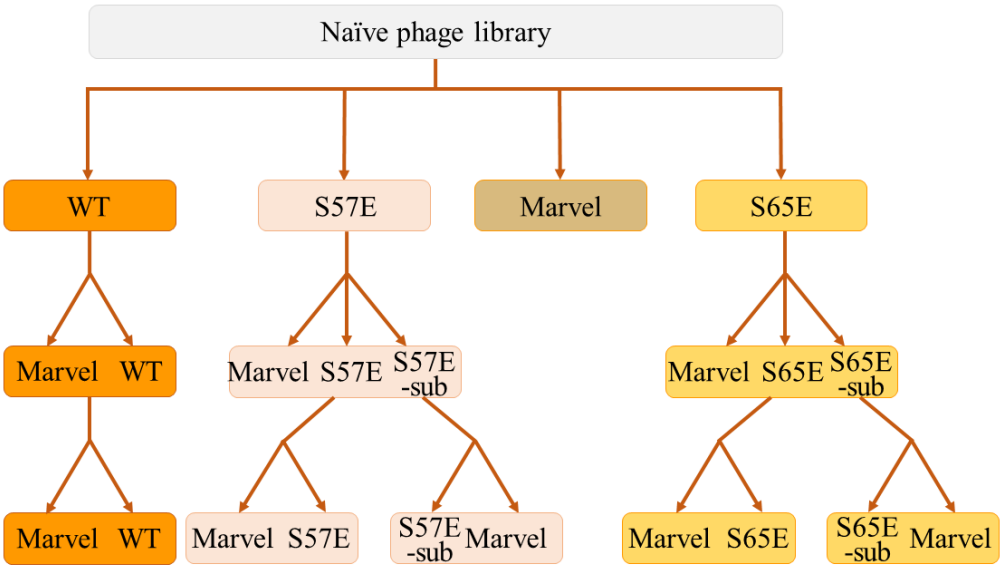


Fig.2.6.5.5.1. Diagram representing the growth plan for Set 1 phage bio-panning experiments. Targets were Marvel, WT-, S65E- and S57E-ubiquitin. Where applicable the presence of subtractive blocking against WT-ubiquitin is shown as "+ mono". The naïve phage library was grown once, then this preparation was panned through three rounds, shown here as rows.

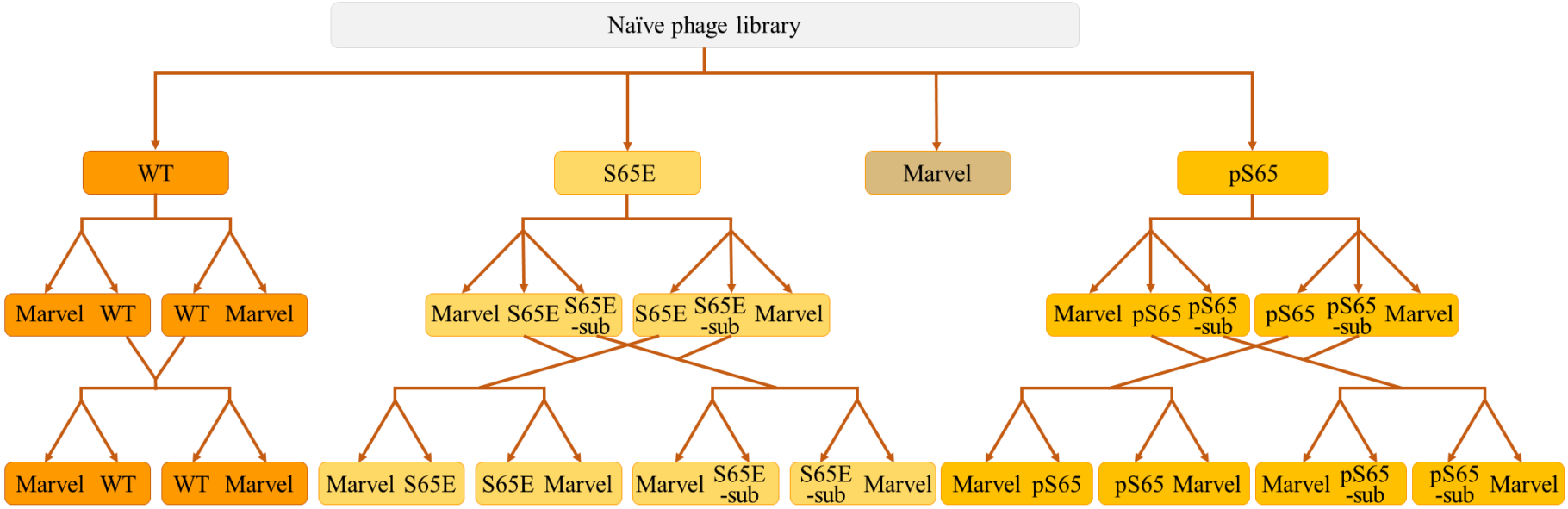


Fig.2.6.5.5.2. *Diagram representing the growth plan for Set 2 phage bio-panning experiments. Targets were Marvel, WT-, S65E- and pS65-ubiquitin.* Where applicable the presence of subtractive blocking against WT-ubiquitin is shown as "-sub". The naïve phage library was grown once, then this preparation was panned through three rounds, shown here as rows. Pans were repeated twice for Round 2. Round 3 was grown from a combination of Round 2 eluates and was also panned twice.

As previously mentioned, each colour block represents one experiment formed of multiple samples. Each sample is created by exposing the phage preparation to the target ubiquitin in ten wells in a 96 well plate. The ten wells were considered to be experimental duplicates as they were conducted at the same time from the same preparations and solutions and were all labelled with the same identifying barcode (Fig.2.6.6.1.1). For the purposes of this thesis, that means the *n* number of each condition was the number of samples sent for sequencing. Overall this equated to eight experiments across the two sets, with *n* numbers for each condition as laid out below (Table 2.6.5.5.1).

	***n* number**		
Condition	**Set 1**	**Set 2**	**Total**
Marvel	1	1	2
WT	3	5	8
WT marvel	2	4	6
S57E	3	0	3
S57E-sub	2	0	2
S57E marvel	2	0	2
S57E-sub marvel	1	0	1
S65E	3	5	8
S65E-sub	2	4	6
S65E marvel	2	4	6
S65E-sub marvel	1	2	3
pS65	0	5	5
pS65-sub	0	4	4
pS65 marvel	0	4	4
pS65-sub marvel	0	2	2
Total	22	40	62

Table 2.6.5.5.1. The distribution of experiments and n numbers for NGPD. The 62 samples sent for NGS were divided across sets and conditions as shown.

2.6.6 Next-Generation Sequencing of Phage Display samples

Eluates from each pan, round and experimental condition were prepared for NGS, including no-protein Marvel controls. This equated to 62 samples in total. A unique DNA barcode was attached to the DNA yielded from each of these samples, allowing the DNA to be easily traced back to its original growth conditions during analysis of the NGS data. Sequencing was conducted by the University of Pennsylvania Perelman School of Medicine's DNA Sequencing Facility.

2.6.6.1 Preparation of samples for NGS

The sample preparation for NGS was extensive, and can be summarised in 13 steps;

1. DNA Extraction
2. Nanodrop quantification
3. First PCR round
4. DNA electrophoresis of first PCR
5. Clean first PCR
6. Second PCR round
7. DNA electrophoresis of second PCR
8. Qubit quantification
9. DNA electrophoresis
10. Clean final product
11. Qubit quantification
12. DNA electrophoresis
13. Prepare sample for transport

As standard 100µL of glycerol stock from each sample was purified using a QIAprep Spin Miniprep kit (QIAGEN, #27104) as per the manufacturer's instructions. This volume was altered to provide equivalent amounts of material if the sample was from a culture with particularly high or low growth. The DNA was eluted into 50µL of the NE buffer supplied with the Nucleospin Gel and PCR clean up kit (Macherey-Nagel, #740609). The DNA yield was determined by Nanodrop spectrophotometry with 2µL of sample and the remainder diluted to 10ng/µL with NE buffer.

A 50µL PCR reaction was conducted on each sample with high-fidelity Q5 DNA Polymerase (New England Biolabs, #M0491S) with a water control as described below. The Q5 buffer and GC enhancer were obtained from the Q5 Reaction Buffer pack (New England Biolabs, #B9027).

	Amount per 50µL
5x Q5 buffer	10µL
GC enhancer	10µL
dNTP mix 10mM	1µL
pSD3_16_F_lnk2 10µM	1µL
pSD3_16_R_lnk1 10µM	1µL
Nuclease free distilled water	25.5µL
Q5 polymerase	0.5µL
DNA 10ng/µL	1µL

The primers were specific to the 16mer library and were used to incorporate the linker sequence. This linker would later be used to attach the sample-specific barcode. The primers are listed below, where plain text matches the barcode primer and bold text matches the plasmid.

Primer	Sequence
Reverse pSD3_16_R_lnk1	GTAATCCTTGTGGTATCG**CGAATGGAGAGCCACCGC**
Forward pSD3_16_F_lnk2	CTAGAACATTTCACTTA**CATGATTACGCCAAGCTCG**

The PCR reaction was conducted for 25 cycles to amplify the DNA in a thermocycler with a pre-heated lid. The amplification steps (#2-4) ran for 25 cycles and were each set to 30 seconds as the PCR products were typically 200-250bp in length.

Step		Temperature	Duration	
1		95°C	3 minutes	
2	Denaturing	95°C	30 seconds	
3	Annealing	60°C	30 seconds	x25
4	Extension	68°C	30 seconds	
5		72°C	5 minutes	
6		4°C	Hold	

PCR products were refrigerated but never frozen in between re-amplifications. 10µL of the PCR product and the low DNA molecular weight ladder (New England Biolabs, #N0557S) were electrophoresed on a 3% agarose gel in TAE buffer to confirm amplification. The gel was run at 100V constant voltage for approximately 1hr.

Using a Macherey-Nagel Nucleospin Gel and PCR clean up kit, the remaining PCR product was cleaned. The NTI buffer was diluted 1:5 with sterile distilled water. The DNA was eluted from the column with 20µL NE elution buffer as supplied with the kit.

The second round of PCR amplification was conducted using 1µL of the eluted DNA as the template. The primers incorporated the barcode into each plasmid (Fig.2.6.6.1.1).

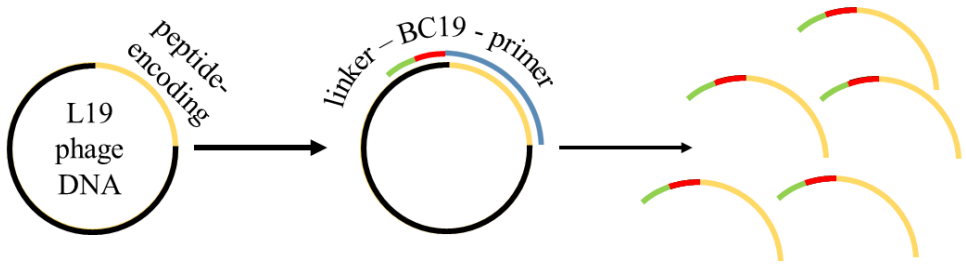


Fig.2.6.6.1.1. The incorporation of barcodes into the phage plasmid for NGS. There were 62 phage panning samples. First the generic linker was included in the DNA of all samples. Then a sample-specific barcode (e.g. BC19 for sample L19) was annealed.

As each primer was barcode-specific, a different primer was used with each DNA sample to ensure that the DNA was tagged with a unique barcode that could be readily linked back to sample and panning information. The list of barcode primers can be found in the Appendices (A2.3). Primer and Barcode 42 were not used as they caused bleed through from other barcodes. The P1 primer sequence is included below;

Primer	Sequence
P1	CCTCTCTATGGGCAGTCGGTGATCTAGAACATTTCACTTAC
Barcode primer	As applicable

A master mix was prepared and template DNA, barcode primer and Q5 polymerase were added separately. The amplification was conducted in a thermocycler with pre-heated lid for 25 cycles.

	Amount per 50µL
5x Q5 buffer	10µL
GC enhancer	10µL
dNTP mix 10mM	1µL
P1 primer 10µM	1µL
Nuclease free distilled water	25.5µL
Q5 polymerase	0.5µL
Barcode primer 10µM	1µL
DNA 10ng/µL	1µL

Step		Temperature	Duration	
1		95°C	3 minutes	
2	Denaturing	95°C	30 seconds	x25
3	Annealing	63°C	30 seconds	
4	Extension	68°C	30 seconds	
5		72°C	5 minutes	
6		4°C	Hold	

10µL of the PCR product was analysed by gel electrophoresis as before.

The concentration of the PCR product was determined using a Qubit Fluorometer (Thermo Fisher Scientific) using 2µL of product and the dsDNA Broad Range Assay kit (Thermo Fisher Scientific, #Q32850) as per the manufacturer's instructions.

All 62 samples were then pooled to give a single combined sample. The sequencing facility required a sample of 1,000ng dsDNA (pre-cleaning weight) encompassing 96 barcodes. This equated to an average 10.53ng of dsDNA per barcode. As only 62 barcodes were required for this thesis, the remaining 33 were used for another project within the laboratory. Therefore, the 62 barcodes in this thesis were allocated a total of 652.63ng dsDNA.

However, this was not evenly divided between the samples. To fully exploit the power of NGS over the 62 barcodes, the amount of DNA per barcode was adjusted to weight the number of

reads towards those samples with a higher pfu/mL, and therefore a higher potential sequence diversity. Samples were sorted into 7 bins according to their pfu/mL, with bin 1 being the lowest, bin 4 the median and bin 7 the highest. Bin 4 was assigned a proportional value of 1; every barcode in that bin was allocated 10.53ng dsDNA. All dsDNA allocations were weighted relative to this, with more dsDNA being provided for higher pfu/mL bins to increase the number of reads, and less dsDNA and fewer reads provided for lower pfu/mL bins. This was to exploit the high diversity possible in the higher pfu/mL bins, and to avoid wasting repeated reads on low diversity of low pfu/mL bins. The weight pfu/mL and dsDNA submitted to NGS and the reads and sequences per sample were included in the Appendices (Table A4.1).

The pooled DNA was cleaned by DNA electrophoresis on a 3% agarose gel, excising the band, solubilising in neat NTI buffer and using the Macherey-Nagel Nucleospin Gel and PCR clean up kit.

The sample was further cleaned using an AMPure XP system (Agencourt, #A63880) as per the manufacturer's using an EasySep Magnet (StemCell Technologies, #18000). 25µL of NE buffer from the Macherey-Nagel clean up kit was used in the final elution step and the supernatant collected into a clean Eppendorf.

The concentration of the cleaned sample was determined by Qubit as previously described. Typically, cleaning led to considerable loss of material and the final yield was expected to be 300-500ng.

50ng of the final sample was run on a 3% agarose gel. The image from this gel was sent to the NGS facility along with the sample to demonstrate purity. The remaining sample was diluted with NE buffer from the Macherey-Nagel clean up kit to within the parameters requested by the NGS facility; 30-50µL of 100-300ng DNA in a 0.5mL LoBind tube (SLS, #E0030108035).

2.6.6.2 Next-Generation Sequencing

The pooled sample containing barcoded dsDNA from 62 phage display experiments was sent for NGS at the University of Pennsylvania Perelman School of Medicine's DNA Sequencing Facility. NGS was conducted on an Ion Torrent S5 with a 540 chip, giving 1,242,184 expected reads, each with a read length of 200bp.

Basic data processing was completed by the facility. As it formed part of the candidate selection process, further processing of the NGS data is included within the results in Chapter 4.

2.7 NUCLEAR MAGNETIC RESONANCE

All NMR was conducted with the advice and technical support of the University of Nottingham High Field NMR Facility.

HSQC NMR relies upon the detection of two types of nuclei; the ^{1}H (practically all naturally occurring hydrogen atoms are ^{1}H) and ^{15}N, which is naturally abundant at low levels (0.366% (De Biévre and Taylor, 1993)) but can be enriched by protein production in minimal media supplemented with a labelled nitrogen source. This combination of two nuclei gives it the name 2D-NMR. In a 2D HSQC NMR experiment the 800 MHz field applied by the magnet results in a slightly different electronic microenvironment and magnetic field surrounding each of the ^{1}H or ^{15}N nuclei that make up the protein's amino acids. This can be detected as a signal peak and assigned to each of the residues within the molecule, except proline and the most N-terminal residue, which are undetectable due to the lack of amide protons and the presence of the -NH2 moiety. A spectrum of uniformly labelled ^{15}N-ubiquitin, generated by growing wildtype-ubiquitin in the presence of labelled nitrogen, on its own acts as the baseline for interaction studies. If the spectrum changes after addition of an un-labelled ligand then binding has taken place: interaction between amino acids changes the microenvironment surrounding

each nucleus and changes the reported signal. This is detected as a chemical shift around the nucleus and is measured in parts per million (ppm).

The extent of the shift can also indicate the affinity of the interaction. As the reported shift is an average of all the spectra captured during the sampling time, a weak interaction will frequently bind and disassociate, resulting in a small average shift. A stronger interaction will be maintained throughout the duration of the sampling time and be reported as two discrete peaks; one before and one after addition of ligand. Weak interactions are referred to as fast exchange between binding states, and strong interactions as slow exchange. A large shift does not necessarily indicate direct involvement of that residue, however. If interaction results in wide-spread structural changes of the labelled ubiquitin, the microenvironment around each residue nuclei will change regardless of its involvement in the binding event.

2.7.1 NMR experimental design

All samples were purified, lyophilised and suspended in deionised distilled water with 10% D_2O before use. The only except was for peflin experiments in Chapter 5: 3.5L of GST-peflin culture was purified through glutathione elution as described above to a volume of 50mL. This was condensed to ~0.1mL and diluted with minimal capture buffer (Appendix A2.1) and 10% D_2O. The same buffer was used for the matched ^{15}N-ubiquitin control samples.

The molar ratio of ligand compared to ^{15}N-ubiquitin and the concentration of both partners varied between experiments but was set to 4:1 or 8:1 with 50μM ^{15}N-ubiquitin as standard. All experiments were conducted at room temperature (298K) or 8ºC (281K) and samples equilibrated to magnet temperature before acquisition began.

As all experiments were designed to initially determine if binding occurred or not, each interaction was run once (n=1) with no titration steps.

Data were acquired using a Bruker 800 MHz Avance III (Chapter 4) or NEO spectrometer (Chapter 5), equipped with a 5mm QCI ($^{1}H/^{13}C/^{15}N/P/^{2}H$) cryoprobe. Prior to acquisition basic instrument calibration and optimisation of ^{1}H 90 degree pulse and water frequency were conducted.

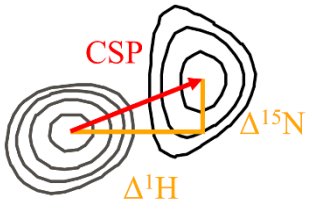
Proton 1D NMR were acquired with solvent suppression employing the excitation sculpting method. Data were acquired with 32k complex points with acquisition time of 2.5 s over 15 ppm. Data were averaged over 64 transients, with an overall recycling delay of 6 seconds. For ^{15}N enriched samples 1D data was additionally acquired using ^{15}N GARP4 composite pulse decoupling, with 16k points and 132 transients.

^{15}N, ^{1}H correlation experiments were recorded using the Heteronuclear Single Quantum Coherence (HSQC) sequence, with 2k points in t2 and 256 points in the indirect (t1) dimension with 16 or 384 transients depending on desired signal to noise. Spectral width in the indirect ^{15}N dimension was 45 ppm, and quadrature detection was achieved using the States-TPPI method.

^{1}D data were processed following standard methods of phasing and baseline correction, and zero-filled by a factor of 4 prior to apodisation using an exponential window function with a 2 Hz line broadening. 2D data were subjected to standard processing, including zero-filling and were apodised using a sine squared window function with a shifted sine bell value optimised to give good resolution with no artefacts. Additionally, linear prediction was used in F1.

In all instances, 1D ^{1}H spectra were acquired first for both ^{15}N-ubiquitin and the potential ligand, be that peflin or NGPD-derived peptide, to check that each individual protein was detectable with reasonable folding. Then the 2D HSQC spectra for ^{15}N-ubiquitin in the absence of binding partner were acquired. The potential ligand was then added to the ^{15}N-ubiquitin sample at the same molar concentration as was used for ligand 1D spectra. The ^{15}N-ubiquitin

2D HSQC was repeated in the presence of binding partner, and the chemical shift perturbations (CSP) were calculated as change from ^{15}N-ubiquitin to ^{15}N-ubiquitin in the presence of binding partner (Fig.2.7.1.1), where ΔH is the change in ^{1}H (ppm, γ axis) and ΔN is the change in ^{15}N (ppm, *x* axis) from the control condition (Williamson, 2013). Assignments were made and CSP calculated using CcpNMR analysis. CSP were considered significant if they were greater than the mean plus one standard deviation (+1SD). The assigned 2D HSQC spectra for ^{15}N-ubiquitin, as assigned by the Layfield and Searle laboratories, can be found in the appendices (Fig.A4.2).



$$\mathrm{CSP} = \sqrt{0.5\left(\Delta H^2 + \frac{\Delta N^2}{5}\right)}$$

Fig.2.7.1.1. Chemical Shift Perturbation calculation from example NMR data. CSP (red) of an example 2D HSQC peak in the absence of ligand (left) compared to the presence of ligand (right) is calculated in ppm from the change in ^{1}H and change in ^{15}N (orange) according to Williamson, 2013.

3 DEFINING CONSENSUS SEQUENCES FOR HUMAN UBDS – A BIOINFORMATICS APPROACH

3.1 INTRODUCTION

A common approach for identifying novel instances of binding domains is often bioinformatics. Alignment of a protein sequence with a known domain or motif is rapid, simple, well established and trusted. So much so that there are over 96,000 publications associated with the search "protein sequence alignment" (as per PubMed, September 2019).

A downside to this ease of use is that protein alignments are conducted frequently by different research groups with different parameters, creating redundant and conflicting data. It is uncommon that old data is comprehensively reviewed in light of new information, leading to out of date consensus sequences that are still being used to interpret novel datasets. Further, domain boundaries are often set by the first structural publication and not updated to account for differences between protein isoforms or new instances of the domain. Consensus sequences are often defined with little consideration for the natural variation that occurs between different proteins or for variations across species. For example, a 'consensus' might be defined across many species or across multiple proteins from several species. The former actually defines sequence conservation through evolution, and the latter includes two sources of variation: both protein and species.

In short, whilst bioinformatic domain annotation and identification are easy and highly useful, they are far from perfect and require a repeatable, systematic approach that can be periodically updated to reflect the current knowledge as new occurrences of a particular domain are identified.

For UBDs in particular, there is no collated systematically produced list of agreed consensus sequences for the field. To find a UBD consensus the user must conduct a lengthy literature

search making sure to note the species, isoform and methods used as it may not apply to the model or species in question or may rely on outdated information. In simple terms, consensus sequences evolve and are not static. This turns what should be a simple five-minute task into a day's work combining disparate information sources. With one of the key aims of this thesis being the identification of the UBD 'dark matter', i.e. overlooked UBDs, a reliable body of consensus sequences was required, both for annotation of novel instances of UBDs and for reference whilst identifying novel families of UBDs. Whilst current domain annotation was not fit for purpose, the development of a systematic method for the identification and annotation of consensus sequences should provide an easily updated resource. This led to the hypothesis that a standardised method for the identification and bioinformatic validation of UBD consensus sequences could be used to identify novel instances of UBDs and highlight unknown aspects of the UBD-related proteome.

To examine this hypothesis, the following aims were set:

- To develop a standardised, cross-resource validated approach to creating consensus sequences
- To create consensus sequences for all known families of human UBDs
- To validate this resource by searching the human proteome for known UBDs
- To use this resource to search existing datasets for novel proteins with putative UBDs.

Here a systematic review of current human UBD annotations between the popular databases Pfam, iUUCD, SMART, UniProt and Prosite (El-Gebali *et al.*, 2019; Zhou *et al.*, 2017; Letunic and Bork, 2018; UniProt Consortium, 2018; Sigrist *et al.*, 2012), is presented. Domains that were present in more than one database were aligned using Jalview and PROMALS3D (Waterhouse, 2009; Pie *et al.*, 2008) and the consensus sequences compared to those currently in the literature, where present. For the purposes of this thesis, a 'motif' refers to a short stretch

of amino acids, typically less than ten, a ‘domain’ refers to a longer, more diverse sequence, and a ‘full consensus sequence’ to the most frequent amino acid sequence across the full length of a given domain.

3.2 DEFINING HUMAN UBD-PROTEINS

Before creation of the UBD resource, a comprehensive list of human proteins containing ‘known’ UBDs had to be compiled. Given the heterogeneity between databases and lack of literature agreement on the precise number of UBD families, the iUUCD classification of UBDs was taken as the standard.

To create a reproducible pipeline, a standardised approach was taken to determine UBD classes, protein lists and domain boundaries. In brief, the UBD groups and families were defined by iUUCD as the only database dedicated to ubiquitin systems; the protein list by Prosite; and the domain annotations by UniProt (Fig.3.2.1).

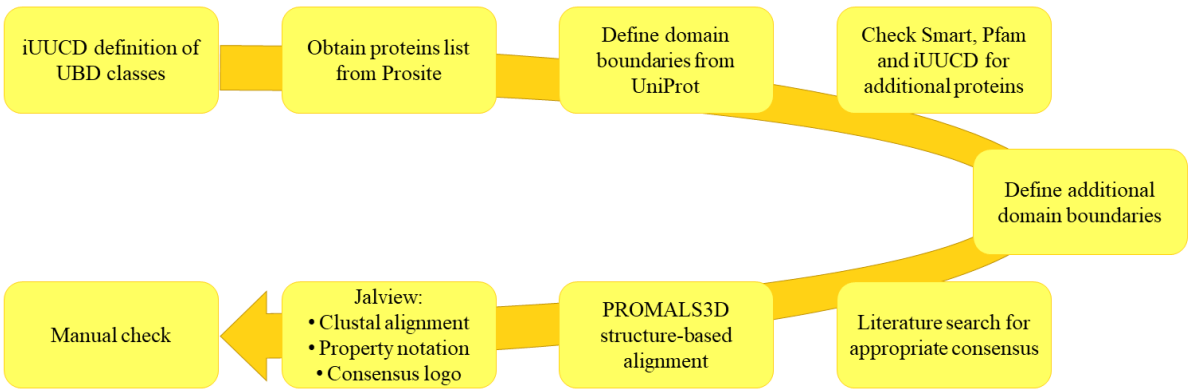


Fig.3.2.1. Pipeline for the standardised creation of UBD consensus sequences.

The iUUCD lists 74 groups of protein related to ubiquitin or ubiquitin-like systems based on domain structure and protein function. Within this, the iUUCD defined 27 UBD families present in humans. Additional searches were conducted through the literature and the other commonly used bioinformatic databases; Prosite, SMART and Pfam.

Each family of UBDs was first searched in Prosite to find the list of human proteins containing that UBD. Prosite was chosen as the first choice as iUUCD, whilst being ubiquitin-centric, was found to be incomplete. iUUCD did not include all UBD containing proteins, missed several multiple UBDs within a single protein and did not differentiate between functionally-verified and predicted UBDs. Prosite, whilst having a higher number of duplicates, provided a good balance between encompassing all UBDs and not containing excessive numbers of false positives.

Supplementary searches of SMART, Pfam, iUUCD and the general literature were conducted to find any proteins missed by Prosite. In all cases, these additional proteins and domains required confirmation by multiple sources before being trusted.

Throughout the process only the canonical protein sequence was used; any variants, homologous proteins, uncharacterised proteins, predicted proteins or isoforms were ignored.

This resulted in identification of 296 domains in 212 proteins across 29 families. Compared to this cross-referenced and curated list the other, commonly used and publicly available, databases contained considerably higher numbers of domains and proteins (Fig.3.2.2).

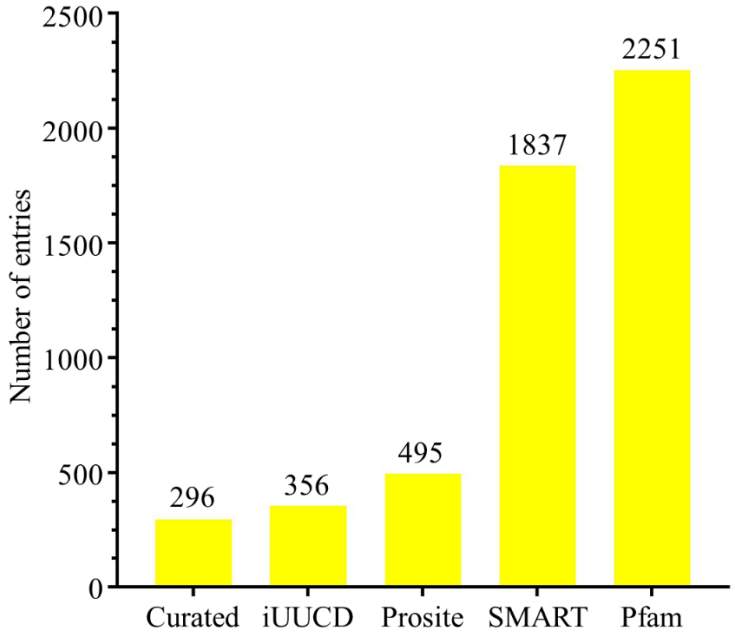


Fig.3.2.2. A comparison of human ubiquitin-binding domains and proteins according to select databases. The number of entries for all human UBD families was extracted from four commonly used databases (iUUCD, Prosite, SMART and Pfam). If a family was misrecognised then it was not included in the total. This was compared to the numbers generated in this thesis (Curated).

Despite the same search parameters being used for each database, and each database being well-established, there was great variety in the number of UBD entries in each. Both SMART and Pfam reported markedly more UBDs compared to the other databases; interrogation of the database outputs confirmed that they contained multiple uncharacterised proteins, isoforms, fragments and predicted proteins. Both databases were found to contain duplicated or out of date data, with some proteins having been removed from UniProt as many as five years ago. Prosite, whilst reporting lower number of UBDs, still contained 199 more entries than could be verified through cross-referencing. The iUUCD output was the most similar to that curated by this thesis, which is to be expected given that it asserts to be both curated and ubiquitin-specific, i.e., to not contain proteins or domains that have not been functionally verified. However, iUUCD had 60 more entries than expected, likely due to the presence of predicted domains. Despite the higher numbers, newly identified, multiple or tandem UBDs were found to be infrequently reported in the database.

The pipeline developed to create this list was standardised and cross-checked at many levels, providing quality assurance and although false positives from the source experimental data are always possible. It was designed to be easily reproducible, allowing the data to be regularly updated and reviewed to account for new developments. Database and literature cross-validation resulted in 212 proteins containing 296 UBDs in 29 families across four groups forming the input for the consensus resource.

The disparity between databases has undoubtably affected the literature, either in misrepresenting the profile of a given UBD or by necessitating lengthy manual verification before use. This database comparison neatly demonstrated that existing bioinformatic resources were not fit for purpose and the importance of regular and thorough data review.

To create up to date, human-specific consensus sequences for all UBD families the 296 domains identified above were processed using the standardised pipeline (Fig.3.2.1). UBD domain boundaries and sequences were preferentially taken from UniProt as the most curated database but in the absence of UniProt annotation were included from SMART, Pfam and iUUCD.

As it was impossible to generate consensus sequences from families that contained only one human protein, all orphans were discarded. In this thesis, orphan UBDs were defined as those unrelated to all other families; whilst it is likely that other examples of the UBD in question exist, they have yet to be discovered. The SOUBA domain was classed as an orphan, separate from the rest of the UBA domains, as it is structurally distinct (Agromayor *et al.*, 2012). Recently discovered unique UBDs which have not been affiliated with any family, such as the CUBAN, CoCUN and cryptic Parkin domains, although this last one is thought to bind ubiquitin in a manner similar to that of HOIP (Lechtenberg *et al*., 2016; Kumar *et al.*, 2017; Castagnoli *et al.*, 2019; Nepravishta *et al.*, 2019), were not included.

Several of the UBD families were not solely ubiquitin-binding domains, i.e. only a subset had ubiquitin-binding activity. For example, of the 248 human SH3 domain-containing proteins listed in SMART and Pfam, only 13 were annotated with ubiquitin-binding activity in iUUCD. A fourteenth protein, BCR/ABL p210 fusion protein, was included in iUUCD but no domain was provided and the corresponding UniProt protein entry was unreviewed. Where possible, only those instances with known ubiquitin-binding activity were considered.

These constraints led to the loss of the SOUBA, Jab1/MPN, GLUE, PFU, PRU, ZRF1-UBD, UBW, CUBAN, MyUb and CoCUN domains which are not included in the analysis. The remaining 19 families all contained at least two examples within the human proteome. These families consisted of 286 domains within 202 proteins, and could be collated within four groups; α-Helix, Zinc Finger (ZnF), Ubiquitin Conjugating enzyme (UBC)-like and Others, It should be noted that the table below indicates a total number of proteins higher than 202 as several proteins contain more than one type of UBD and so appear several times between different families (Table 3.2.1).

Table 3.2.1. Classification of human UBD-containing proteins.

Group	Family		Proteins	Domains
α-Helix	UBA	Ubiquitin associated	52	60
	UIM	Ubiquitin-interacting motif	22	43
	CUE	Coupling of ubiquitin conjugation to endoplasmic reticulum degradation	14	15
	VHS	VPS-27, Hrs and STAM	9	9
	GAT	GGA and TOM	6	6
	MIU	Motif interacting with ubiquitin	5	6
	UBAN	Ubiquitin binding in ABIN and NEMO	5	5
	DUIM	Double-sided UIM	3	3
	UMI	Motif interacting with ubiquitin	2	2
	UBM	Ubiquitin binding motif	2	2
ZnF	NZF	Npl4 zinc finger	16	35
	UBP	Ubiquitin-specific processing protease	14	14
	UBZ	Ubiquitin-binding zinc finger	10	12
	ZnF_A20	A20 type zinc finger	7	11
UBC-like	Ubc	Ubiquitin conjugating enzyme	5	5
	UEV	UBC E2 variant	2	2
Other	beta-Prp	Beta-Propeller	36	36
	SH3	Src homology 3	13	13
	CARD	Caspase recruitment domain	4	7
Total	**19**		**227**	**286**

The proteins listed in Table 3.2.1 had been validated by comparison between databases supplemented with manual scrutiny of the literature and other resources. It should be noted that one of the UBZ-containing proteins, NEMO, was not listed in any database despite evidence of involvement in ubiquitin binding for over a decade (for examples, see Cordier *et al.*, 2008; Rizzo *et al.*, 2014; Toma *et al.*, 2015). As various publications have referred to the NEMO-ZnF as a UBZ domain, it is included in the UBZ family here.

As mentioned above, any domains that could not be cross-referenced were discarded. Also, for large multi-function families such as SH3, where possible only domains with ubiquitin-binding activity were considered. This was to refine the consensus sequence to those domains with functional relevance.

For these reasons, the list was not meant to be exhaustive of all UBDs, but rather a representative list to generate consensus motifs of the frequently occurring domains.

After a comprehensive analysis of both databases and literature, the list of 286 UBDs, their boundaries and their sequences was generated as described above. Now that suitable input data had been curated, human-specific and up to date consensus sequences for each UBD family could be generated to aid identification of new family members.

3.3 APPROACHES TO GENERATING CONSENSUS SEQUENCES

Whilst the iUUCD is a useful reference for ubiquitin-related proteins with a wealth of protein database information, it does not supply reliable information on domain consensuses or boundaries. An automated consensus sequence is provided but given the incomplete list of proteins underpinning it and the disputed domain boundaries (which are often contradictory to those found elsewhere), the consensus sequences provided can only be considered consensuses for the iUUCD and not suitable for use with novel proteins. As the consensus spans the whole domain, which can be hundreds of amino acids long, it is of limited use when trying to scan novel proteins; a short identifying motif that can indicate presence of a UBD for later alignment to a full consensus would be of much more practical use.

To help fill this knowledge gap, the cross-checked list created above was used to create an updated resource of ubiquitin-binding Short Linear Motifs (SLiMs). SLiMs differ from true UBDs in that they do not cover an entire domain, as defined as a protein region responsible for a given interaction and that can independently fold. There are several easily usable and publicly

accessible programmes for generating consensus sequences and motifs; the most appropriate of these were used to create consensuses based on amino acid, physicochemical property, occupancy and structural similarities. This information was then manually curated for the best pattern identification and a motif created for each UBD family. Where possible the canonical sequence of the parent protein from UniProt was used to remove skew by the presence of isoforms or variants.

The resources available for consensus generation included MEME Suite which, whilst providing a wide range of specialised tools, did not report the best motifs, just the first motifs found (Bailey *et al.*, 2009). CLUSTAL and the related MUSCLE servers were both highly useful at generating sequence alignments with indications of amino acid residue at each position, as well as the ability to cope with inserts or deletions. However, amino acid physicochemical properties and sequence structure were not considered, and the long consensuses were of little practical use (Edgar, 2004; Sievers *et al.*, 2011). PROMALS3D had the advantage of automatically performing homologue searches to gather pertinent information about the whole protein and considering secondary structure when performing alignments. The output could be difficult to use and, in the writer's experience, often missed useful patterns with a tendency to over-fragment the sequences. Whilst this nominally improved the physiological relevance, it did reduce the quality and occupancy of the consensus. The open source EMBOSS set of tools provided specialised sequence analysis for a wide variety of input formats and could manage several different types of sequence or protein. It was less user-friendly than the other services as it was command-line based and required a familiarity with code manipulation (Rice *et al.*, 2000). Logo-based tools, such as WebLogo and SeqLogos produced attractive, easily interpretable infographics which contain more information that a flat alignment or population analysis (Reche 2018). However, they suffer when sequences are of different lengths or when a position is defined by the residue properties rather than the exact

amino acid. The Jalview desktop programme is a highly useful tool that combines information from multiple sources for structure predictions, functional site analysis, consensus conservation, clustering, pairwise and multiple alignments and protein disorder analysis. It is a highly visual and intuitive programme but still can only create predicted alignments and sequences: a manual check to identify patterns alignments was still required (Table 3.3.1).

Table 3.3.1. A comparison of publicly available tools for generating amino acid sequence alignments.

Database or programme	Advantages	Disadvantages
MEME Suite	Diverse, specialised, identifies a small usable motif	Does not identify the best motifs
CLUSTAL/MUSCLE	Trusted sequence alignment tools, easily copes with inserts or deletions	No physicochemical or structural consideration, often generated overly long consensuses
PROMALS3D	Homolog searches, accounts for secondary structure, increased physiological relevance	Often over-fragments sequences, output can be difficult to use
EMBOSS	Open source, specialised, can be tailored to fit purpose	Not user-friendly, requires knowledge of code manipulation
Logo tools, e.g. WebLogo and SeqLogo	Highly visual, easily interpretable, display all residues at a position ranked by consensus	Do not easily identify physicochemical conservation, input sequences must be of equal length
Jalview	Desktop-based, combines multiple sources of information, intuitive, visual	Occasionally misaligns sequences, requires a manual check

To create a reliable motif, a combination of tools was used. First, a literature search was conducted to identify any published consensus sequences for the list of UBDs identified above. Then, Clustal W and Clustal X v2.0 (Larkin *et al.*, 2007) sequence alignments were conducted within Jalview. Whilst some alignments appeared to be mis-called they were not manually

adjusted. This was to preserve the reproducibility created by the computer-based alignment. The Zappo-defined physicochemical properties of the aligned sequences were indicated. Site occupancy and data quality were also indicated in Jalview with the automatically determined consensus logo. This was compared to PROMALS3D alignment for structural consensus analysis. Together, this represented amino acid alignment, structural, physicochemical, consensus and occupancy information, which were all combined and condensed to create UBD-identifying SLiMs.

3.4 PARAMETERS IN DETERMINING UBD CONSENSUSES

As discussed previously a major part of identifying new instances of existing protein domains is the similarity to a full consensus sequence or motif. The latter is generally more useful in the first instance, for example, when faced with an unannotated functional ubiquitin-binding protein it is much easier to quickly scan the sequence by eye for a conserved motif than it is to align the whole protein to all know UBD families. In this scenario, the SLiM acts as a quick marker to suggest which UBD could be present, which can then be checked later. It also acts to identify domains which are more divergent from the consensus as homology is more likely to be conserved across a short common motif than a large variable domain sequence. Whilst well-accepted motifs have been previously identified for the more common UBD families, not all have been determined, making it harder to identify novel instances. The primary aim of this chapter was to create practical and updated consensus motifs for all UBD families for use in interrogating specialised datasets or the wider proteome. To this end, each family within the UBD list created above was analysed with a standardised motif-discovery pipeline.

It should be noted that the sequences from PROMALS3D may not directly align with those from Jalview due to the different methods of alignment. In all instances, the PROMALS3D

structural alignment is shown at the top of the figure in the yellow box, with the Jalview alignment and analysis below.

PROMALS3D amino acid conservation was calculated on a scale from 1-9, with 9 indicating conservation through all proteins submitted. Only scores of 5 and above were shown. The consensus indicated the automatically determined amino acid consensus sequence, with capital letters denoting the consensus amino acid. Lower case letters presented physicochemical conservation in the absence of direct amino acid conservation as per the table below (Table 3.4.1). The conserved structural features were predicted, where e = beta-strand and h = alpha helix.

Table 3.4.1 Physicochemical properties of amino acids aligned by PROMALS3D.

Symbol	Property	Amino Acids
l	Aliphatic	I, V, L
@	Aromatic	Y, H, W, F
h	Hydrophobic	W, F, Y, M, L, I, V, A, C, T, H
o	Alcohol	S, T
p	Polar	D, E, H, K, N, Q, R, S, T
t	Tiny	A, G, C, S
s	Small	A, G, C, S, V, N, D, T, P
b	Bulky	E, F, I, K, L, M, Q, R, W, Y
+	Positively charged	K, R, H
-	Negatively charged	D, E
c	Charged	D, E, K, R, H
.	No conservation	

Within Jalview, amino acids were highlighted according to Zappo physicochemical properties (Table 3.4.2) and the conservation, quality, consensus and occupancy calculated. Jalview conservation analysis measured the conservation of physicochemical properties for each position and was displayed on a scale of 0-11, with 0 indicating no conservation, 10 indicating conservation of properties but not amino acid, and 11 conservation of amino acid (and by default, also properties). The * symbol was used in place of the number 11, and + in place of 10 to ease display. Any position with an occupancy too low for reliable conservation scoring

was presented as a -. The quality analysis was conducted as the probability of mutation at a given position. The consensus logo indicated what percentage of the aligned column was occupied by the displayed amino acid. Finally, occupancy indicated the number of sequences with an aligned amino acid at the given position.

Table 3.4.2. Zappo classification of physicochemical properties.

Symbol	Colour	Property	Amino acids
w	Green	Hydrophilic	S, T, N, Q
*	Pink	Unique	P, G
C	Yellow	Cysteine	C
−	Red	Negative	D, E
+	Blue	Positive	K, R, H
@	Orange	Aromatic	F, W, Y
#	Peach	Aliphatic/Hydrophobic	I, L, V, A, M

The colours and classifications were determined according to Zappo convention, and the symbol was included for reference within this thesis.

These symbols applied to the outputs from PROMALS3D and Jalview analyses, but given the overlap between the two sets of symbols and as colour could not be reliably integrated into consensuses (due to variation in reproduction through printing and potential for misidentification through colour perception), the different property symbols were combined and a comprehensive annotation system defined (Table 3.4.3). Whilst the original symbols were used for the PROMALS3D and Jalview data, these new combined symbols were used when identifying new consensuses. Any conserved amino acids were represented as per the conventional single letter amino acid symbols. Repeats of a property were annotated in round brackets, for example `(x7)` indicated seven `x` amino acids in a row. Square brackets denoted conservation of any of the amino acids within the brackets for example `[DS]` indicated a requirement for either Asp or Ser at that position.

Table 3.4.3. Annotation symbols for consensuses identified in the text.

Symbol	Property	Amino acids
l	Aliphatic	I, V, L
@	Aromatic	Y, H, W, F
h	Hydrophobic	W, F, Y, M, L, I, V, A, C, T, H
o	Alcohol	S, T
p	Polar	D, E, H, K, N, Q, R, S, T
t	Tiny	A, G, C, S
s	Small	A, G, C, S, V, N, D, T, P
b	Bulky	E, F, I, K, L, M, Q, R, W, Y
+	Positive	K, R, H
-	Negative	D, E
c	Charged	D, E, K, R, H
w	Hydrophilic	S, T, N, Q
*	Unique	G, P
#	Aliphatic/Hydrophobic	A, I, L, M, V
x	Any	
.	Variable occupancy	

Due to the input data containing multiple annotation systems, a combined system was created to report consensuses within the text. Any consensuses reported from the literature were converted to match this system and any deviations noted. Variable occupancy was included to allow for high levels of amino acid deletions or inserts between sequences.

In some UBD families due to the number or length of the domains only areas of interest are shown. In these cases, a diagram of the full alignment is also shown with displayed areas highlighted.

3.5 CONSENSUS MOTIFS FOR THE A-HELICAL UBD FAMILIES

The α-Helical UBD group is the largest, both in terms of number of families and number of proteins and contains some of the most studied domains; the UBA and UIM-related families. As a consequence of this high level of study several of the families already have published consensus motifs or full sequences which were often accepted by the field, e.g. the 'LALAL' motif of the UIM family. As not all of these consensus sequences were derived from up to date or human-specific lists of proteins, all of the families underwent analysis with the standardised domain pipeline outlined above. The literature search identified three families with no current

consensus, four families with consensuses that required review and three families with appropriate, human-relevant and comprehensive consensuses.

3.5.1 A consensus motif for the UBA family

The motif for the largest of the α-Helical families, the UBA family, is well established as 'MGF' within the wider consensus of LxQLxx.MG.FxxxEx.xxALxAxxx...xxAxxxLL (Hofmann, 1996). Although the accepted motif was derived from UBA domains of multiple species, the Jalview and PROMALS3D analyses were unable to produce a substantially improved human-specific consensus or motif (Appendix A3.1.1). That the analysis conducted here resulted in a motif that matched the published motif demonstrated the accuracy of the pipeline.

3.5.2 A consensus motif for the UIM family

Sometimes considered as the 'minimal' α-Helical UBD, UIMs are formed of just one α-helix around 20 amino acids in length (Hicke *et al.*, 2005) with a well-established consensus defined as −xxhxxAhx[h−]Szx− where z = an aliphatic bulky or hydrophobic residue (Swanson *et al.*, 2003) but more commonly referred to as the 'LALAL' motif. The published consensus was comparable to that generated here, but a note of caution should be included for the LALAL motif. Whilst well accepted, the motif was not comparable or identical to either the full consensus or the PROMALS3D and Jalview analysis of human UIM domains (Fig.3.5.2.1). The same span of amino acids would be more accurately described by hxxA[ILM], which was notably less memorable. However, it would still be useful for a manual scan of a single protein sequence, especially if the researcher were experienced enough to view it as an aliphatic/hydrophobic patch rather than the exact amino acids.

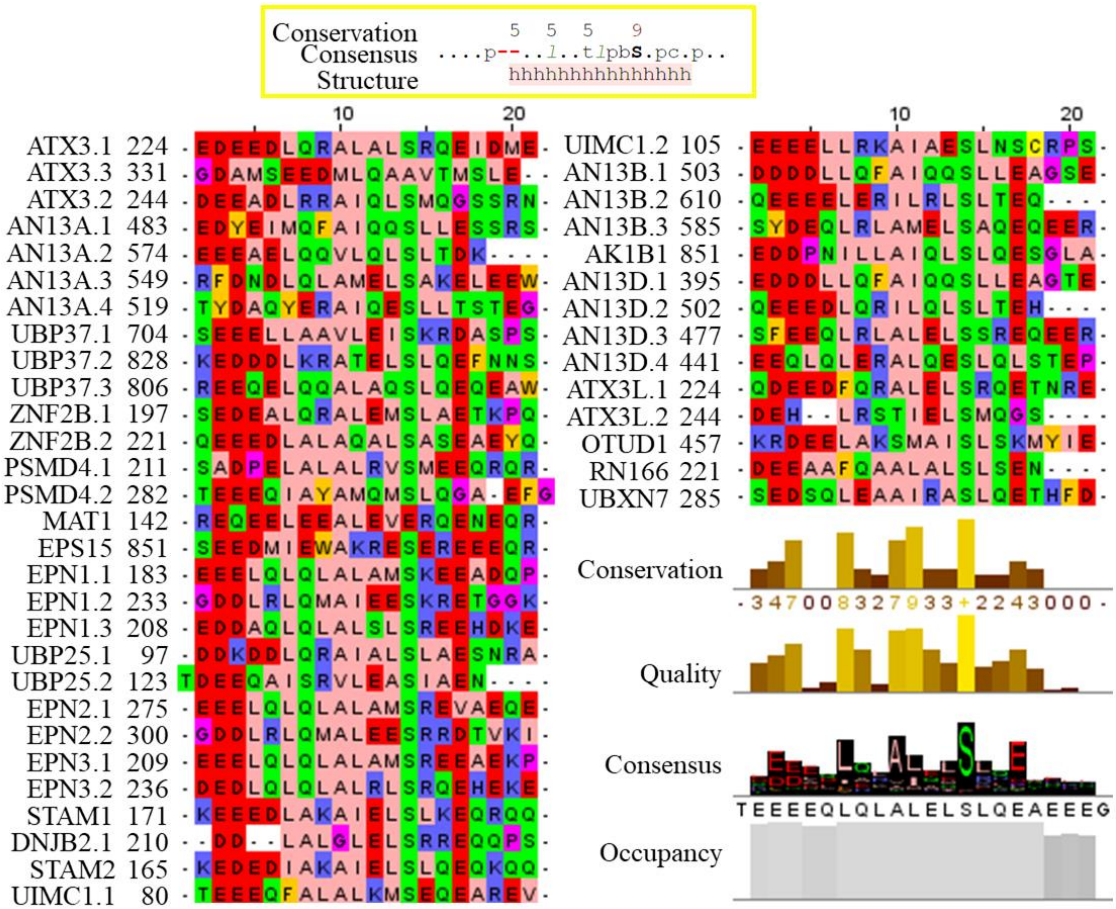


Fig.3.5.2.1. Structural, physicochemical and consensus alignments of human UIMs. 43 human UIM domains were determined by cross database comparison, then aligned with PROMALS3D (upper, yellow panel) and Clustal W and X within Jalview (lower panel). Zappo physicochemical properties were indicated by colour within Jalview and alignment conservation, quality, suggested consensus and occupancy calculated automatically. The position of the domain within each protein was indicated on the left by the number of the first amino acid and the relative position within the alignment at the top.

3.5.3 A consensus motif for the CUE family

Previous studies established a fairly well-conserved motif for the CUE domain: a conserved methionine-phenylalanine-proline trio, then, around 24 non-conserved amino acids later, a dileucine motif, noted as `MFP(x24)LL` (Prag *et al.*, 2003; Kang *et al.*, 2003; Ponting, 2000). With such a broad range of positions the existing motif was not hugely practical: a great number of false positives could be identified on the basis of just two combinations of conserved amino acids. The original characterisation of the CUE motif involved direct amino acid comparison between proteins in *Arabidopsis*, yeast, *Caenorhabditis elegans*, *Candida albincans*, *Drosophila*, *Fubu rubripes*, *Leishmania major*, *Schizosaccharomyces pombe*, rats and humans.

The varied species input indicates that the CUE domain was comparatively well conserved between species, but likely contributed to the broad definition of the motif. To help solidify this motif and make finding putative CUE domains within a dataset easier, the 15 human CUE domains were aligned (Fig.3.5.3.1).

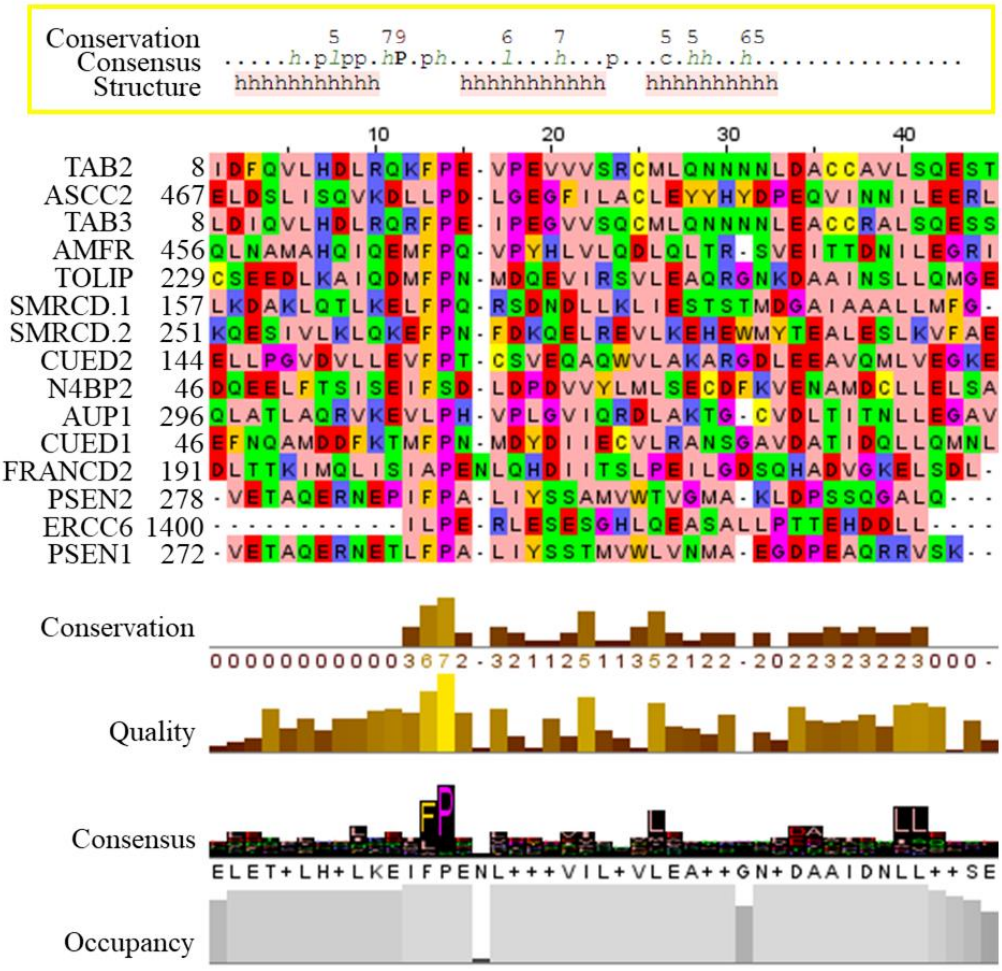


Fig.3.5.3.1. Structural, physicochemical and consensus alignments of human CUE domains. 15 human CUE domains were determined by cross database comparison, then aligned with PROMALS3D (upper, yellow panel) and Clustal W and X within Jalview (lower panel). Zappo physicochemical properties were indicated by colour within Jalview and alignment conservation, quality, suggested consensus and occupancy calculated automatically. The position of the domain within each protein was indicated on the left by the number of the first amino acid and the relative position within the alignment at the top.

Reassessment of the 15 human CUE domains to create a concise, usable characterising motif was conducted through comparison of structural, physicochemical, and consensus alignments. Only human proteins verified through several databases were included in an effort to create a

stricter motif. After removal of the non-human sequences the methionine was considerably less conserved and it is suggested that it should not be included in a human CUE motif. To do so would likely exclude many human CUE domains as they did not fit the non-species-specific motif. Instead, the new motif is suggested as lxphFPp(x25)LLp.

3.5.4 A consensus motif for the VHS family

Despite VHS domains being aligned several times throughout the literature no identifying motifs or consensuses have been published beyond the requirement for a 'DW', which was too short to be of practical use due to the high probability of false positives (Takatsu *et al.*, 2001; Mao *et al.*, 2000; Ren and Hurley, 2010; Lohi *et al.*, 2002; Lohi and Lehto, 1998). This was surprising, given it frequently occurs in proteins with multiple UBDs and that it is unusually long for a UBD (approximately 135 amino acids), both of which make it an interesting target for further study. These alignments were, once again, across several species and were not human-specific. Despite the length of the domain there were several areas of conservation (Fig.3.5.4.1).

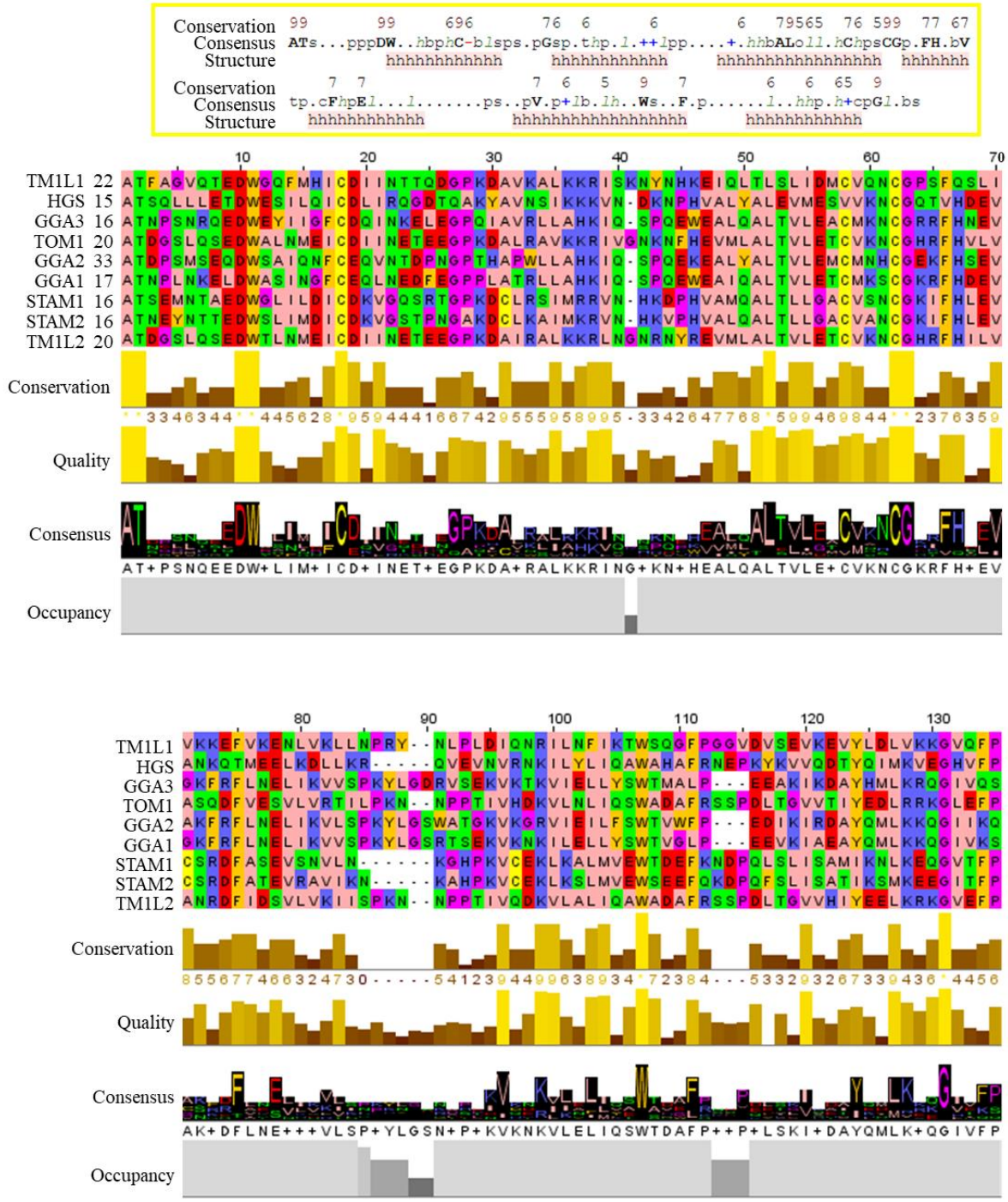


Fig.3.5.4.1. Structural, physicochemical and consensus alignments of human VHS domains. Nine human VHS domains were determined by cross database comparison, then aligned with PROMALS3D (upper, yellow panel) and Clustal W and X within Jalview (lower panel). Zappo physicochemical properties were indicated by colour within Jalview and alignment conservation, quality, suggested consensus and occupancy calculated automatically. The position of the domain within each protein was indicated on the left by the number of the first amino acid and the relative position within the alignment at the top.

As occupancy, quality and physicochemical properties stayed well conserved throughout the majority of the sequence, any areas of conservation were considered reliable and good targets for motif finding (Fig.3.5.4.1). There were ten amino acids with perfect conservation, at relative positions Ala1, Tyr2, Asp10, Trp11, Cys18, Leu52, Cys62, Gly63, Trp107 and Gly131. An identifying motif of `AT(x7)DW(x5)[FI]C[DE]` was found to have perfect conservation across all known human VHS domains. There was also a highly conserved cysteine pair (Cys58 and Cys62) which was traditionally identified in ZnF domains (Miller *et al.*, 1985).

3.5.5 A consensus motif for the GAT family

There have been several attempts at a GAT sequence alignment within the literature although these have never been condensed into a predictive motif (Collins *et al.*, 2003; Shiba *et al.*, 2004; Bilodeau *et al.*, 2004). The `P[ED]DL` motif was identified from the GAT domains within GGA proteins specifically (Collins *et al.*, 2003), however, this is outdated; it was originally based on an alignment of the GAT domains within the three human GGA proteins and yeast Gga1p and Gga2p homologues. There are two issues with this; the yeast homologues weight the consensus towards the GGA, skewing the data away from the human specificity required here, and it does not include the other human GAT-containing proteins; TOM1, TM1L1 and TM1L2. The paper acknowledges the lack of TOM1 in the alignment but dismisses it due to differing domain boundaries. Exclusion of those proteins only permits a motif for the GGA-GAT domains not GAT in general, as the GAT domains of TOM1, TM1L1 and TM1L2 do not cover this same stretch of sequence and have no notable homology to that section of the alignment (relative positions 17-20 in Fig.3.5.5.1).

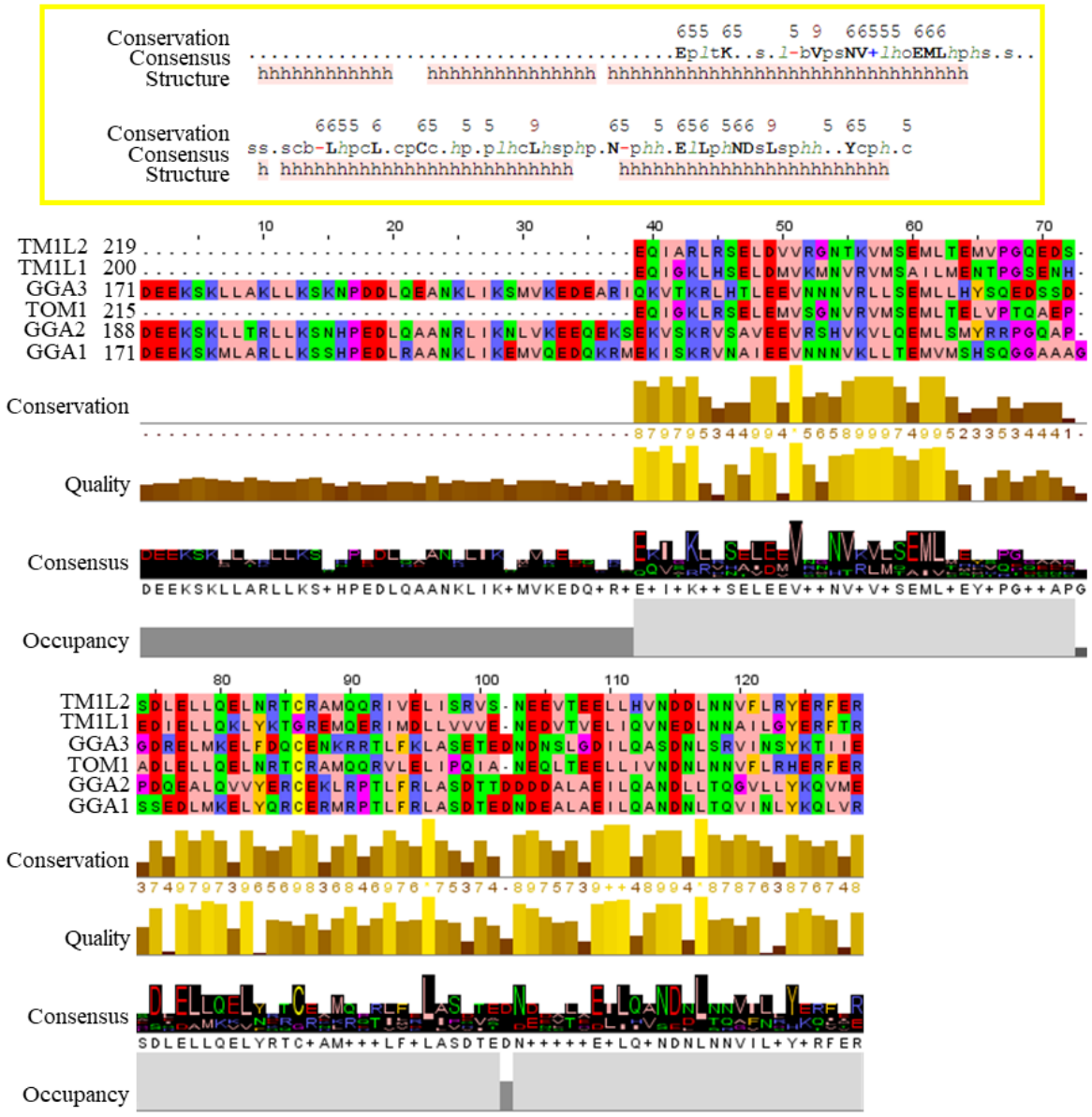


Fig.3.5.5.1. Structural, physicochemical and consensus alignments of human GAT domains. Six human GAT domains were determined by cross database comparison, then aligned with PROMALS3D (upper, yellow panel) and Clustal W and X within Jalview (lower panel). Zappo physicochemical properties were indicated by colour within Jalview and alignment conservation, quality, suggested consensus and occupancy calculated automatically. The position of the domain within each protein was indicated on the left by the number of the first amino acid and the relative position within the alignment at the top.

Multiple sequence alignment of the six human GAT domains highlighted several areas of homology that could be the basis of an identifying motif (Fig.3.5.5.1). There were three perfectly conserved amino acids at the relative positions of Val51, Leu96 and Leu117. After testing several possible areas using the MOTIF search engine of the human KEGG genome database (Kanehisa *et al.*, 2002) for matches, a potential motif was identified in the middle of the alignment. The `[DS]x[ED][LA][LM][KQ]x[LV]xxx[CG]` motif found at relative positions 75-86 covered an area of generally high physicochemical conservation, as noted in both structural and sequence-based alignments. Searching for this motif within the MOTIF/KEGG database resulted in 17 human proteins that contained the motif, compared to the six GAT domains previously identified. Interestingly, one of these 17 hits was the NEDD4 binding protein 2 (N4BP2). It contains a putative CUE domain (UniProt Consortium, 2018) and has been linked to DNA repair (Watanabe *et al.*, 2003), a classically ubiquitin-mediated process. NEDD4 is an ubiquitin-like E3 ligase that shares structural similarity with ubiquitin (Mayer *et al.*, 1998). Each position of the `[DS]x[ED][LA][LM][KQ]x[LV]xxx[CG]` motif was important for its specificity, as replacing any of the constrained positions with an `x` led to 50+ hits. With the high conservation of this motif, its dramatic loss of specificity after modification of even a single position and its links to ubiquitin-related proteins, this was considered a strong motif for the identification of human GAT UBDs.

3.5.6 A consensus motif for the MIU family

A published motif is well established and accepted for the MIU family and the analysis conducted here added little new information or specificity (Appendix A3.1.2). For completeness, the motif is included; `DφxLAxxLxx...x` where φ = large hydrophobic (Penengo *et al.*, 2006).

3.5.7 A consensus motif for the UBAN family

There were only five human members identified within the UBAN family, which show high homology, allowing for a well conserved and established consensus. Within the literature this has already been defined as a short motif for quick identification, `DFxxERxxRE`, and a longer consensus sequence for verification, `ExxxxLxxQxxxYxxDFxxERxxRE(x5)KxxL(x6)L` (Wagner *et al.*, 2008). Whilst extra constraints could be included, such as the conserved polar or hydrophobic residues indicated by the PROMALS3D analysis, the existing motif was already sufficiently specific (Appendix A3.1.3).

3.5.8 A consensus motif for the DUIM family

The DUIM family has been shown to be functionally distinct from the UIM family, despite sequence and structural similarities (Hirano *et al.*, 2006). Indeed, the domains presented here were originally identified as UIM domains until the functional distinction was realised. DUIMs have a particular binding mode, whereby a single DUIM helix can bind the Ile44 patches of two ubiquitin molecules on opposite faces of the DUIM. At the time of submitting this thesis (September, 2019), the DUIM family was only recognised by the iUUCD database and not by UniProt, Pfam, SMART or Prosite. The consensus for the DUIM domain was defined as `-x-xhxhAhAzSz[SA]-`, where `z` represented a bulky hydrophobic or polar residue (Hirano *et al.*, 2006). However, this was originally derived from the experimentally derived DUIMs listed below as well as several predicted DUIMs across five different species. As such, a consensus for the human-specific experimentally validated DUIM domains was created here.

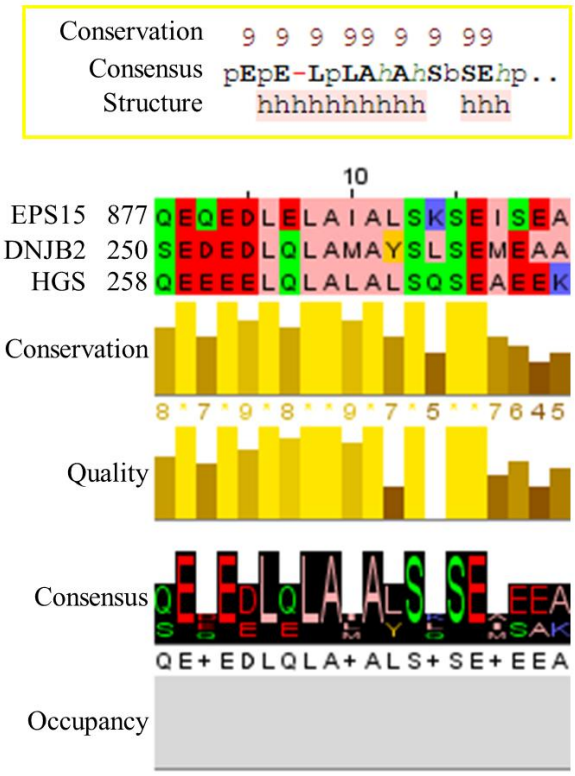


Fig.3.5.8.1. Structural, physicochemical and consensus alignments of human DUIM domains. Three human DUIM domains were determined by cross database comparison, then aligned with Clustal W and X within Jalview. Zappo physicochemical properties were indicated by colour within Jalview and alignment conservation, quality, suggested consensus and occupancy calculated automatically. The position of the domain within each protein was indicated on the left by the number of the first amino acid and the relative position within the alignment at the top.

There was excellent agreement between the structural, sequence and physicochemical alignments, likely in part due to the small number of family members. The Glu in relative position four and both the Ala residues were found to contribute significantly to the double-sided binding of HGS-DUIM towards ubiquitin and so were key in defining a consensus that was distinct from the UIM family. The consensus was defined as `ExExLxLAxAxSxSE`. Whilst there was considerable physicochemical conservation at the x positions, given the high number of amino acid constrained positions, these x residues were given more flexibility to accommodate development of the motif as new DUIM domains are identified.

3.5.9 A consensus motif for the UMI family

The UMI family was little studied and so far only identified in two proteins; the RNF168-UMI was identified by biochemical methods and the ESR1-UMI by homology to RNF168 (Pinato *et al.*, 2011; Pesiri *et al.*, 2013). Both were short domains of only nine amino acids long, which was too short for PROMALS3D structural analysis (Fig.3.5.9.1).

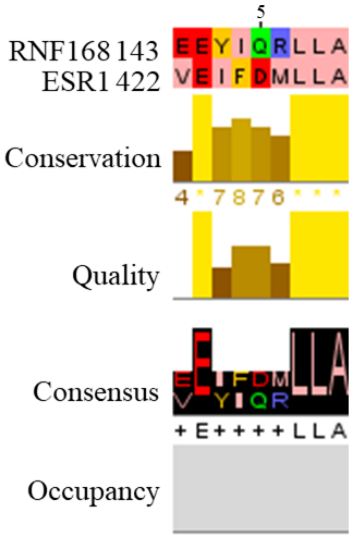


Fig.3.5.9.1. Structural, physicochemical and consensus alignments of human UMI domains. Two human UMI domains were determined by cross database comparison, then aligned with Clustal W and X within Jalview. Zappo physicochemical properties were indicated by colour within Jalview and alignment conservation, quality, suggested consensus and occupancy calculated automatically. The position of the domain within each protein was indicated on the left by the number of the first amino acid and the relative position within the alignment at the top.

As the two domains were too short for structural analysis and as the ESR1-UMI was identified due to homology with the RNF168-UMI, any motif derived from this analysis was likely to be overly focused. For these reasons if more instances of this domain are identified then they may not fit the motif suggested here, and this motif will require updating. However the motif was identified as `ExxxxLLA` and was included here for completeness. Using the MOTIF search for human proteins containing this motif resulted in 585 proteins (data not shown), which demonstrated the further work required to create a usable identifying motif for the human UMI domain.

3.5.10 A consensus motif for the UBM family

The UBM also has only two confirmed human members to date, the Rev1 and Pol iota proteins (Bienko *et al.*, 2005). No motif has been provided for this family, with previous instances identified through a yeast two hybrid experiment. An alignment of human, mouse, chicken and *Saccharomyces cerevisiae* (*S. cerevisiae*) Rev1 and Pol iota was conducted for a cross-species analysis, but this was to prepare for a structural analysis and no consensus conclusions were drawn (Bomar *et al.*, 2010). An invariant Leu-Pro pair was identified at the centre of the domains (Bienko *et al.*,2005). This is near useless as a mechanism for identifying novel UBMs, however, as a multitude of proteins contain an Leu-Pro pair - so many so that the MOTIF online library reached the maximum number of proteins within the display when searched for Leu-Pro within the human proteome (Kanehisa *et al.*, 2002). The two domains are fairly well conserved, however, with ample scope for a suggested motif (Fig.3.5.10.1).

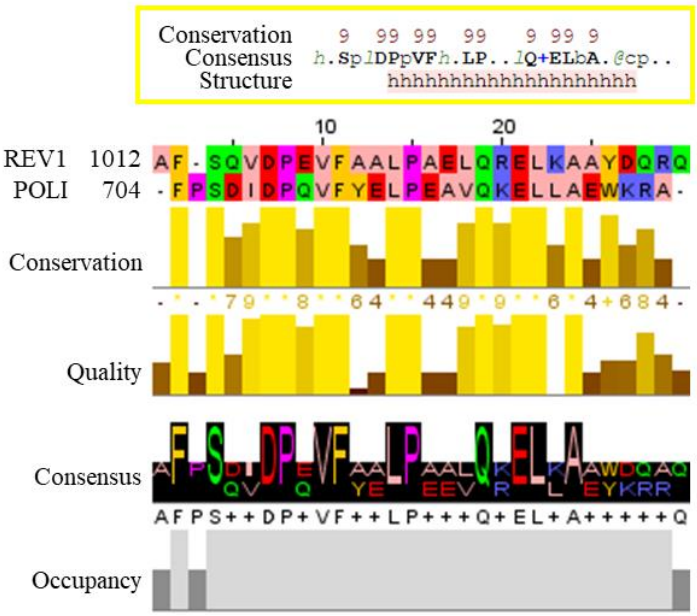


Fig.3.5.10.1. Structural, physicochemical and consensus alignments of human UBM domains. Two human UBM domains were determined by cross database comparison, then aligned with PROMALS3D (upper, yellow panel) and Clustal W and X within Jalview (lower panel). Zappo physicochemical properties were indicated by colour within Jalview and alignment conservation, quality, suggested consensus and occupancy calculated automatically. The position of the domain within each protein was indicated on the left by the number of the first amino acid and the relative position within the alignment at the top.

As can be seen from the PROMALS3D structural alignment (Fig.3.5.10.1, top panel) there was considerable conservation of both structure and amino acid sequence, which was highly likely given that there were only two input sequences. As the quality of the alignment remained high until the Ala in relative position 24 (bottom panel), this was taken as the boundary of the potential motif. The initial motif was defined as `SxxDPxVFxxLPxxxQxELxA`. Through MOTIF searching for aligned human proteins within the human KEGG GENES database, it was determined that the motif could withstand refinement to `DPxVFxxLP` before excessive number of unconfirmed proteins were included in the search results. This motif was both short enough for easy identification within a manual scan of a protein sequence and encompassed the invariant Leu-Pro pair.

3.6 CONSENSUS MOTIFS FOR THE ZNF UBD FAMILIES

All of the ZnF families are well characterised and the majority have published consensus sequences or motifs. Each existing motif or sequence was assessed for current suitability based on the up to date list of human ZnF proteins. Two families were found to be consistent with their published consensuses, one required review and one had no published motif.

3.6.1 A consensus motif for the NZF family

The consensus for the general NZF family was determined by this reanalysis as `xxxxWxCxxCxxxN(x6)CxxC(x5)` which fit well with that determined for human NZF domains (Alam *et al.*, 2004) (Appendix Fig.A3.1.4).

3.6.2 A consensus motif for the UBP family

In keeping with the trend developing throughout this chapter, whilst several papers have published multiple sequence alignments for the UBP domain, they have always been for immediate use to confirm a new UBP rather than generate a consensus for future use (Bonnet

et al., 2008; Yang *et al.*, 2019a; Reyes-Turcu *et al.*, 2006; Pai *et al.*, 2007; Allen and Bycroft, 2007; Zhang *et al.*, 2011). In an effort to create this resource and isolate identifying motifs or consensus sequences, the 14 human UBP domains were analysed with PROMALS3D and Jalview (Fig.3.6.2.1).

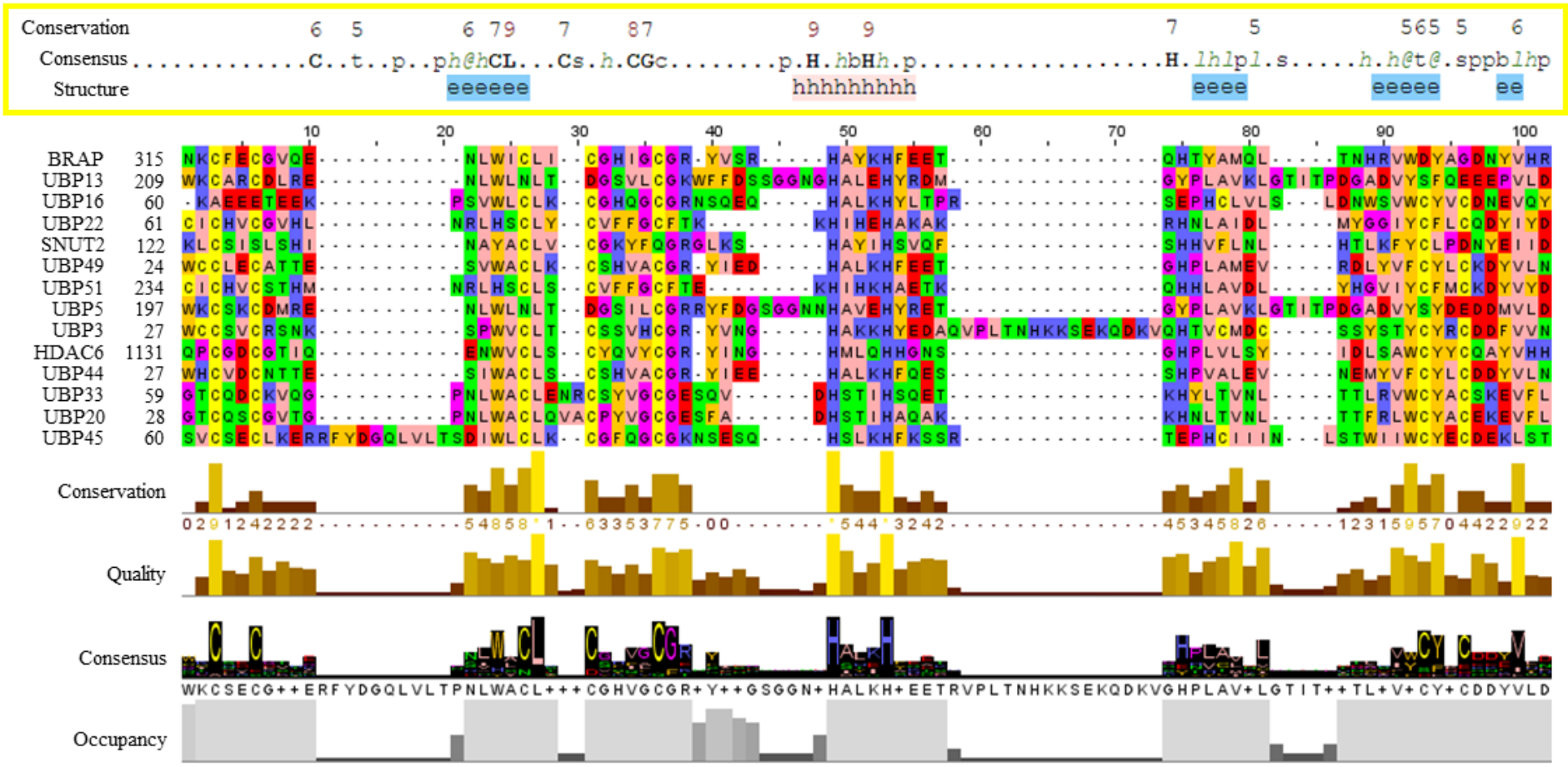


Fig.3.6.2.1. Structural, physicochemical and consensus alignments of human UBP domains. 14 human UBP domains were aligned with PROMALS3D (upper, yellow panel) and Clustal W and X within Jalview (lower panel). Zappo physicochemical properties were indicated by colour within Jalview and alignment conservation, quality, suggested consensus and occupancy calculated automatically. The position of the domain within each protein was indicated on the left by the number of the first amino acid and the relative position within the alignment at the top.

The alignment of UBP domains in Fig.3.6.2.1 showed the varying lengths of the domains, with several inserted sequences that were unique to a single protein, for example the additional ten amino acids towards the N-terminus of UBP45. These inserts made it difficult to identify a single motif of defined length for all proteins. Instead several unconnected smaller motifs were identified. These included two of the so-called Cys-boxes that coordinate metal ion binding within zinc finger domains and the catalytic His-box (Baker *et al.*, 1992; Papa and Hochstrasser, 1993): `[CA]xx[CES]`, `pl[WH]x[CN]Lx..[CD]xxx[CQ][GF]x` and `H#xpH`. To combine these into one single consensus would create the long consensus of `[CA]xx[CES](x4)(.11)pl[WH]x[CN]Lx..[CD]xxx[CQ][GF]x(.10)H#xpH`

Whilst there was potentially a final Cys box at the C-terminus of the domain with a motif of `@[CDS]@x[CGED]xxxl`, this was considerably less conserved and so was ignored.

3.6.3 A consensus motif for the UBZ family

The UBZ family has a published motif based on several studies, `CxxC(`x_n`)[CH]xDx[CH]`, where `(`x_n`)` stands for a variable number of any amino acids (Rizzo *et al.*, 2014). However, when compared to the Jalview and PROMALS3D analysis conducted here the published motif could be improved upon (Fig.3.6.3.1).

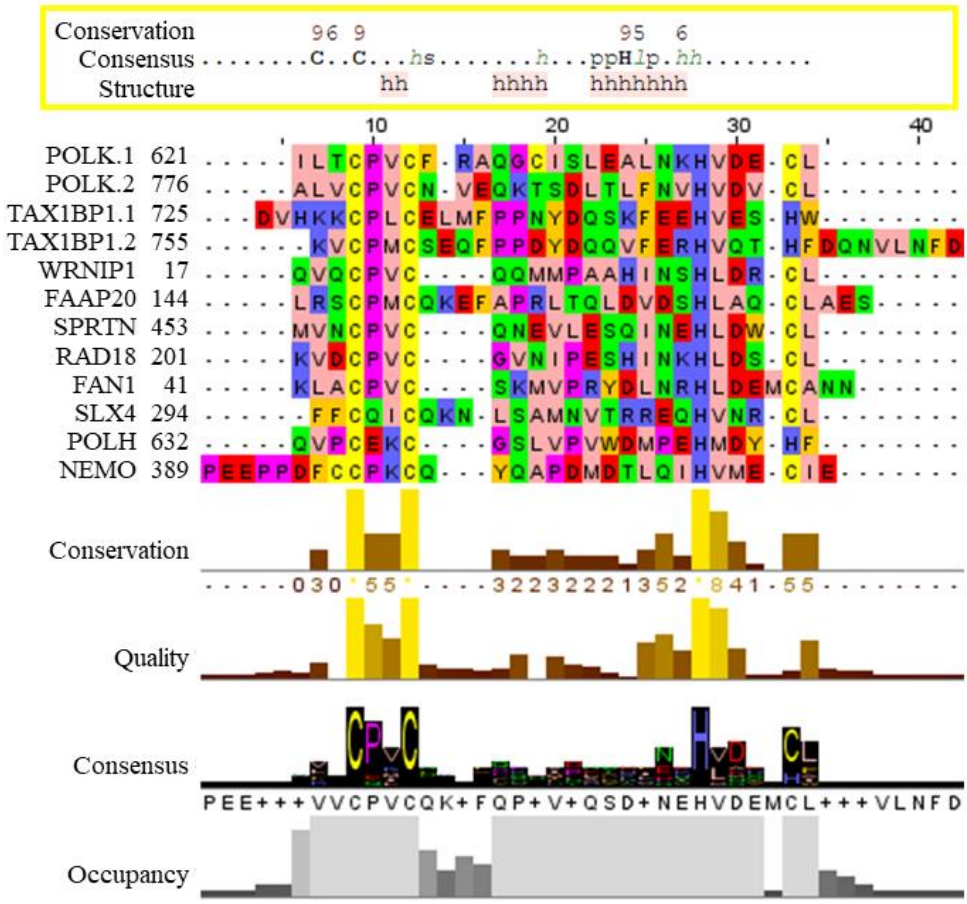


Fig.3.6.3.1. Structural, physicochemical and consensus alignments of human UBZ domains. 12 human UBZ domains were determined by cross database comparison, then aligned with PROMALS3D (upper, yellow panel) and Clustal W and X within Jalview (lower panel). Zappo physicochemical properties were indicated by colour within Jalview and alignment conservation, quality, suggested consensus and occupancy calculated automatically. The position of the domain within each protein was indicated on the left by the number of the first amino acid and the relative position within the alignment at the top.

As can be seen in the alignment of the 12 human UBZ domains, the published motif of `CxxC(x`$_n$`)[CH]xDx[CH]` (Rizzo *et al.*, 2014) was no longer applicable (Fig.3.6.3.1). The `CxxC` aligned to relative positions 9-12. Then, 12-16 amino acids later there was a fully conserved His instead of the quoted `[CH]`. Then the Asp two amino acids further did not align well. Seven of the 12 domains aligned with Asp30, which is not sufficiently conserved to justify the absolute requirement for an Asp residue indicated by the published motif. However, whilst the amino acids at relative position 10 were more conserved, with ten of 12 domains presenting

as Pro10, the motif did not indicate any form of consensus at that position. The new, human-specific motif for UBZ domains was therefore determined as `CP#C(xn)H#xx.[CH].`

3.6.4 A consensus motif for the ZnF_A20 family

The ZnF_A20 domain has been published with a well conserved motif, albeit hidden away in the supplementary data (Wertz *et al.*, 2004). The motif, `CxxxsCsxhxssxxxshCpxC@`, agreed well with the analysis conducted here (Appendix Fig.A3.1.5).

3.7 CONSENSUS MOTIFS FOR THE UBC-LIKE UBD FAMILIES

The UBC-like group of UBDs was small, with only two families, seven proteins and seven domains. Identifying motifs or sequences for such small families were difficult to determine confidently, as conservation of an amino acid was more likely to be coincidental when only a few domains were compared. Whilst there were published consensus sequences for the UBC enzymes in general, no consensus motifs or sequences for the UBDs could be found within the literature.

3.7.1 A consensus motif for the Ubc family

The Ubc domain exists in a small number of UBC enzymes. Whereas most did not display the ability to non-covalently bind ubiquitin, five UBC proteins have been confirmed to contain a UBD (Zhou *et al.*, 2017). These UBDs were often secondary, low affinity sites (Wright *et al.*, 2016), which made it more difficult to study them or generate consensus motifs. Added to this, all of the members of this UBD family were closely related, all being ubiquitin conjugating E2 enzymes, and so any areas of conservation may have been due to factors other than conservation of ubiquitin-binding activity. For these reasons any consensus identified could not definitely be considered an identifying motif for the Ubc UBD family but may have instead

identified the UBC enzymes in question. An analysis of conservation within the Ubc domains and an identifying motif is presented regardless for completeness (Fig.3.7.1.1).

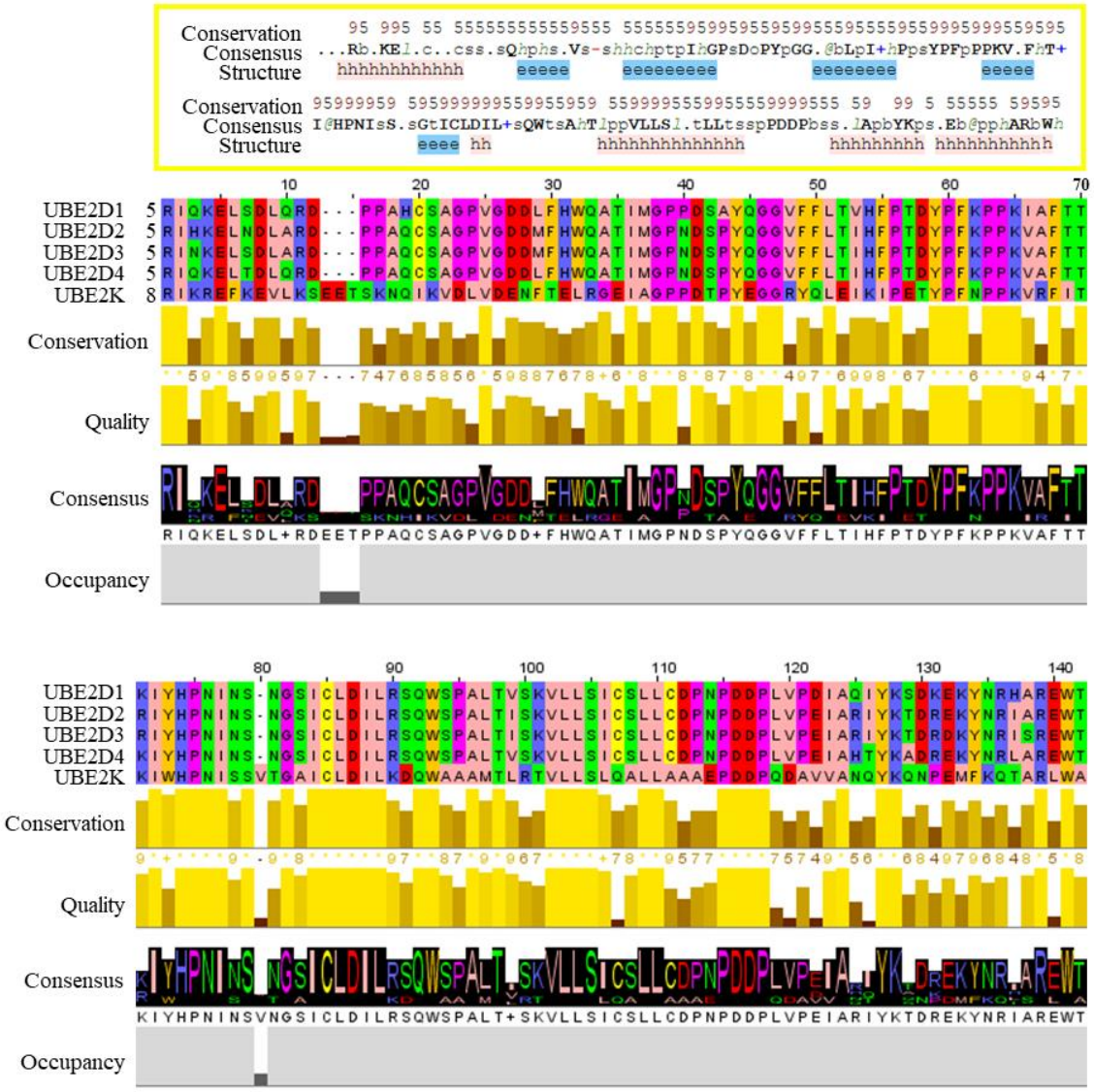


Fig.3.7.1.1. Structural, physicochemical and consensus alignments of human Ubc domains. Five human Ubc domains were determined by cross database comparison, then aligned with PROMALS3D (upper, yellow panel) and Clustal W and X within Jalview (lower panel). Zappo physicochemical properties were indicated by colour within Jalview and alignment conservation, quality, suggested consensus and occupancy calculated automatically. The position of the domain within each protein was indicated on the left by the number of the first amino acid and the relative position within the alignment at the top.

There was considerable amino acid and physicochemical homology between the Ubc domains, as was to be expected from such closely related proteins (Fig.3.7.1.1). A large number of potential of consensus motifs could have been created from the alignments. In an effort to extract a motif that was specific to the Ubc domain rather than the E2 protein family, motifs were checked with the MOTIF database, and a motif selected that did not also occur in other E2 enzymes. If more Ubc domains are identified then this motif will need revision, to ensure that it is specific to the UBD and not the broader family. Regardless, the motif is included here for reference: `LDILxxQW`.

3.7.2 A consensus motif for the UEV family

Through database cross-referencing the UEV domain was identified in two human proteins, TS101 and UEVLD. The literature mentioned other UEV-containing proteins, such as Mms2, however these could not be validated through Uniprot, iUUCD, Pfam, SMART or Prosite and so were discarded (Pornillos *et al.*, 2002). As there was no established motif found within the literature, the two UEV domains were analysed through PROMALS3D and Jalview for motif identification (Fig.3.7.2.1).

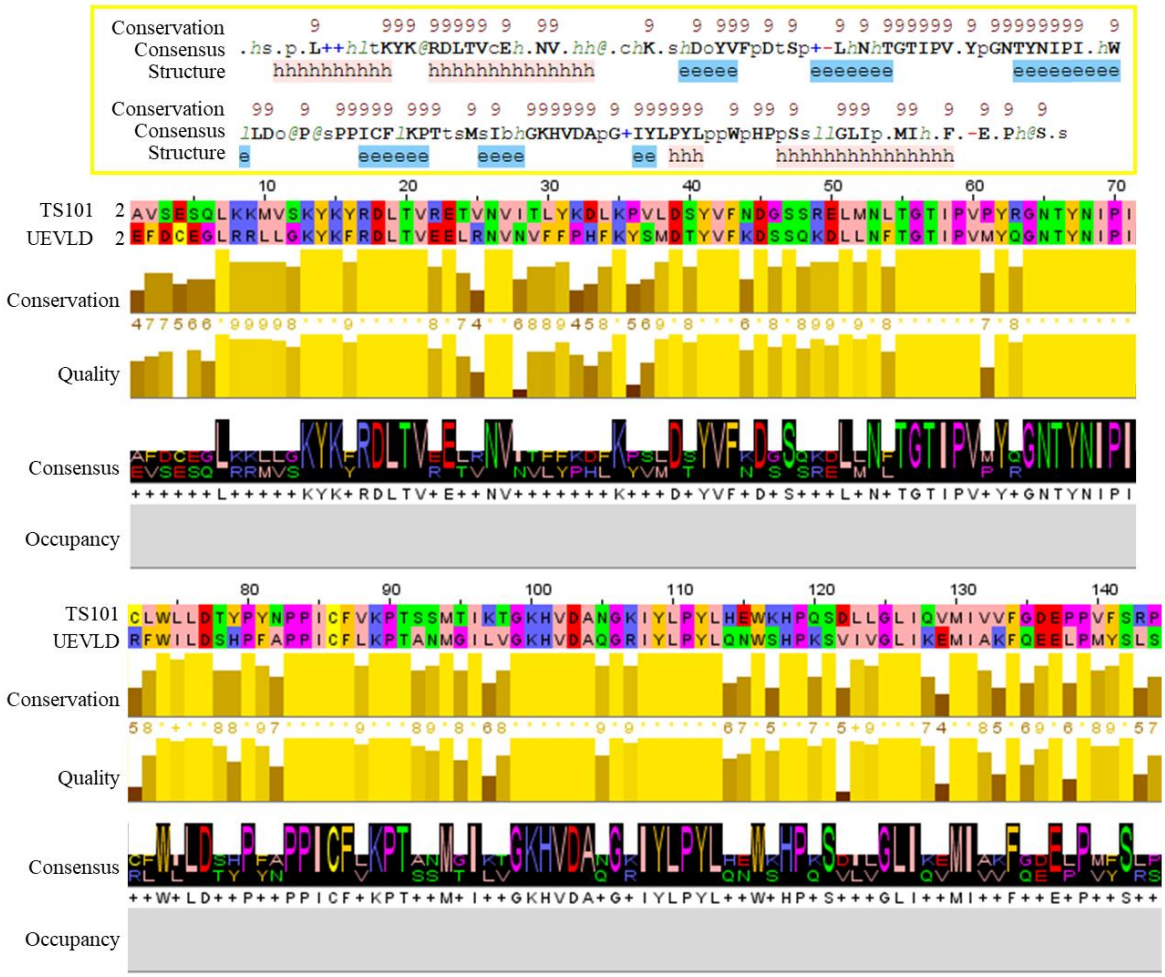


Fig.3.7.2.1. Structural, physicochemical and consensus alignments of human UEV domains. Two human UEV domains were determined by cross database comparison, then aligned with PROMALS3D (upper, yellow panel) and Clustal W and X within Jalview (lower panel). Zappo physicochemical properties were indicated by colour within Jalview and alignment conservation, quality, suggested consensus and occupancy calculated automatically. The position of the domain within each protein was indicated on the left by the number of the first amino acid and the relative position within the alignment at the top.

Similar to the Ubc family, there were considerable areas of homology between the two human UEV domains (Fig.3.7.2.1). In an effort to generate a motif for the UBD rather than generalised homology between the proteins, the amino acids found to be critical for ubiquitin-binding were identified through literature searches. The specific amino acids involved in binding by UEVLD have not yet been experimentally determined so TS101 was used as a representative. For TS101 the ubiquitin-binding site had been identified through NMR of a TS101 isoform different to the canonical form presented here: all amino acids

reported by NMR were shifted one position towards the C-terminus compared to the canonical sequence. Accounting for this, the NMR-identified amino acids were at relative positions Val42, Asn44, Asp45, and Phe87 (Pornillos *et al.*, 2002). All bar Asn44 were conserved between the two domains, with UEVLD presenting with Lys44 instead. Both Asn and Lys are polar residues, however with information from only two domains to draw on it was decided to represent position 44 as non-conserved rather than a required polar residue. Accordingly, the motif was chosen to cover three of the four binding residues that were located close to each other: `KxxxDxYVFxD`.

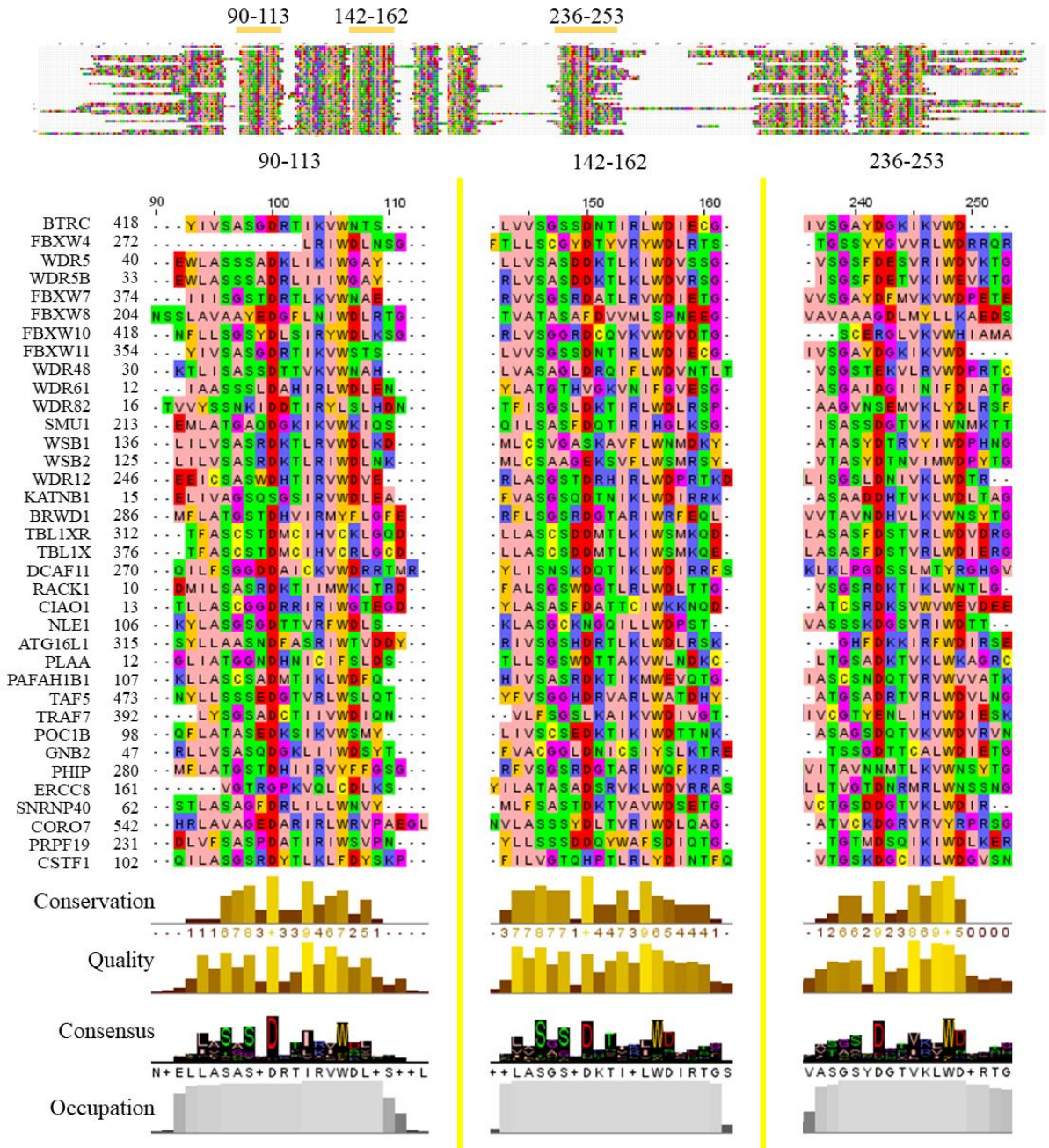
3.8 CONSENSUS MOTIF FOR THE OTHER UBD FAMILIES

The UBD families within the Other group bore little to no similarity to one another and were grouped based solely on lack of homology to other families. These four families included three partial motifs that required updating and one family with no published consensus sequence or motif.

3.8.1 A consensus motif for the beta-Prp family

The beta-Prp domain was named for its folded beta propeller structure of four antiparallel beta strands. The domain, also known as WD40 repeat propeller, had a fairly well established Trp and Asp motif followed by approximately 40-60 amino acids which was repeated multiple times within the protein (Li and Roberts, 2001). Given the variable number of amino acids between the WD repeats database discovery of new beta-Prp domains will always be difficult as most motif searching programmes do not allow for inserts of variable length. Whilst a single WD40 repeat could be identified through database searches, a manual check of the number of repeats would always be required. In order to allow quicker and more confident identification of beta-Prp domains, both manually and through database searches, an updated motif with more localised constraints was identified. Due to the length of the human beta-Prp domains, up to

330 amino acids long, only the regions of interest were presented (Fig.3.8.1.1). The PROMALS3D structural alignment provided little new information and was too long for practical comparison to the Jalview data and so was not shown here (Appendix A3.1.6).



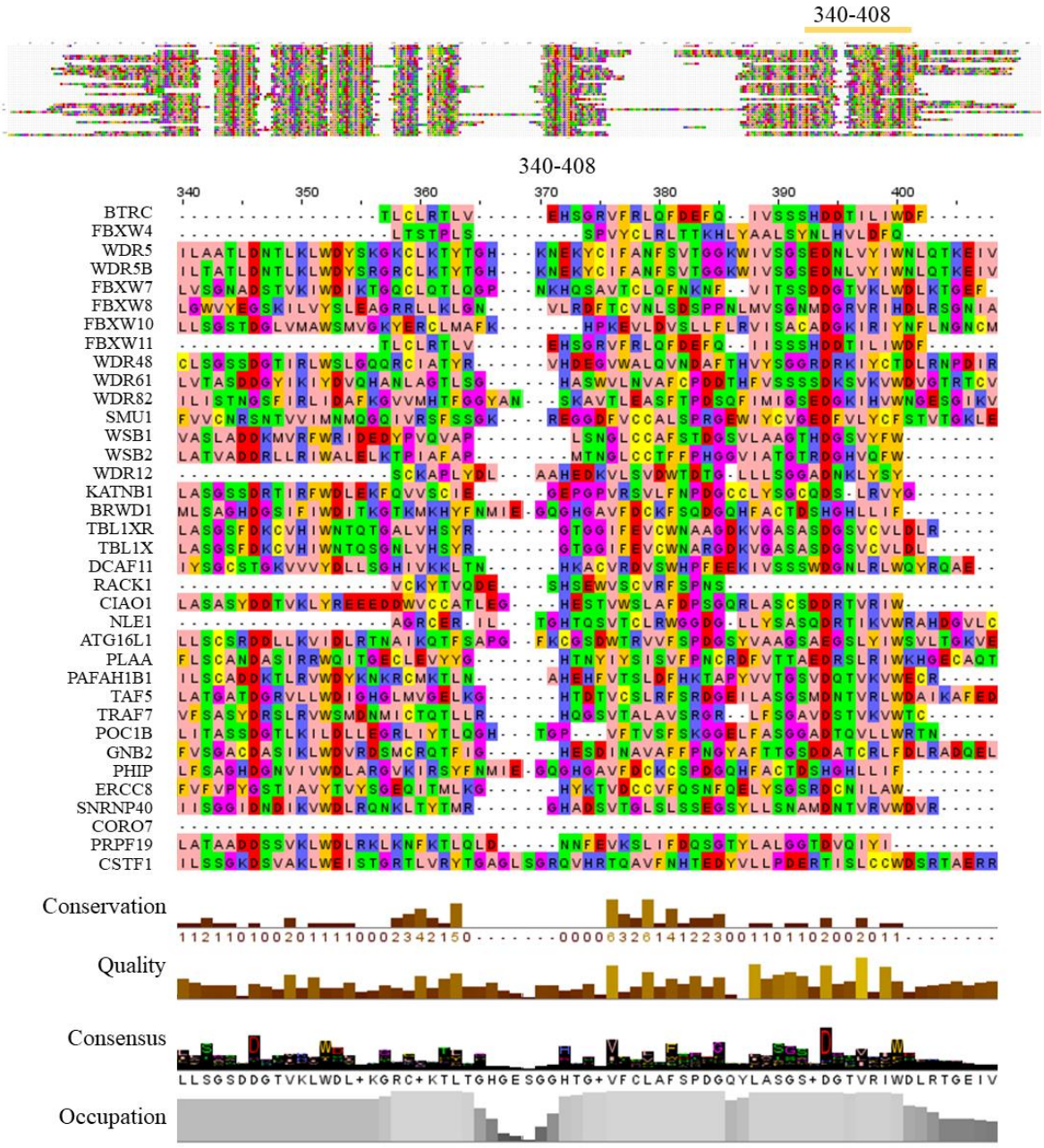


Fig.3.8.1.1. Structural, physicochemical and consensus alignments of human beta-Prp domains. 36 human beta-Prp domains were determined by cross database comparison, then aligned with Clustal W and X within Jalview. Zappo physicochemical properties were indicated by colour and alignment conservation, quality, suggested consensus and occupancy calculated automatically. Areas of interest were shown with relative amino acid positions and location within the full alignment indicated in the top panel. The position of the domain within each protein was indicated on the left by the number of the first amino acid and the relative position within the alignment at the top.

The ubiquitin-binding beta-Prp domains had an average of five WD repeats. Alignment by structure did not successfully identify these WD repeats and so the PROMALS3D data were discarded (Appendix A3.1.6). Refining the beta-Prp motif for greater ease of use and human-specificity identified several areas of high-quality physicochemical alignment (Fig.3.8.1.1). Each of these areas spanned a WD repeat, suggesting that further consensus information could be extracted. In general there was a hydrophobic/aliphatic patch approximately ten positions prior to the WD pair, although the exact positioning of this was variable and so could not be included in a consensus. More strikingly, the region between this patch and the WD repeat showed considerable conservation. This area began with nearly perfectly conserved Ser, and was followed by another Ser, a very well conserved Asp, a hydrophilic residue and an aliphatic-positive-aliphatic motif immediately before the WD. This motif was denoted as `SxSxDxwl+lWD` and was well conserved at each WD repeat.

3.8.2 A consensus motif for the SH3 family

The existing consensus for the SH3 domain was based around the chicken Spectrin SH3 domain identified in 1992 (Musacchio *et al.*, 1992). This helped identify many of the 248 human SH3 domain-containing proteins listed in SMART and Pfam. As only 13 human SH3 domains have been confirmed as UBDs, the existing consensus was reanalysed for refinement to a human and ubiquitin-specific motif (Fig.3.8.2.1).

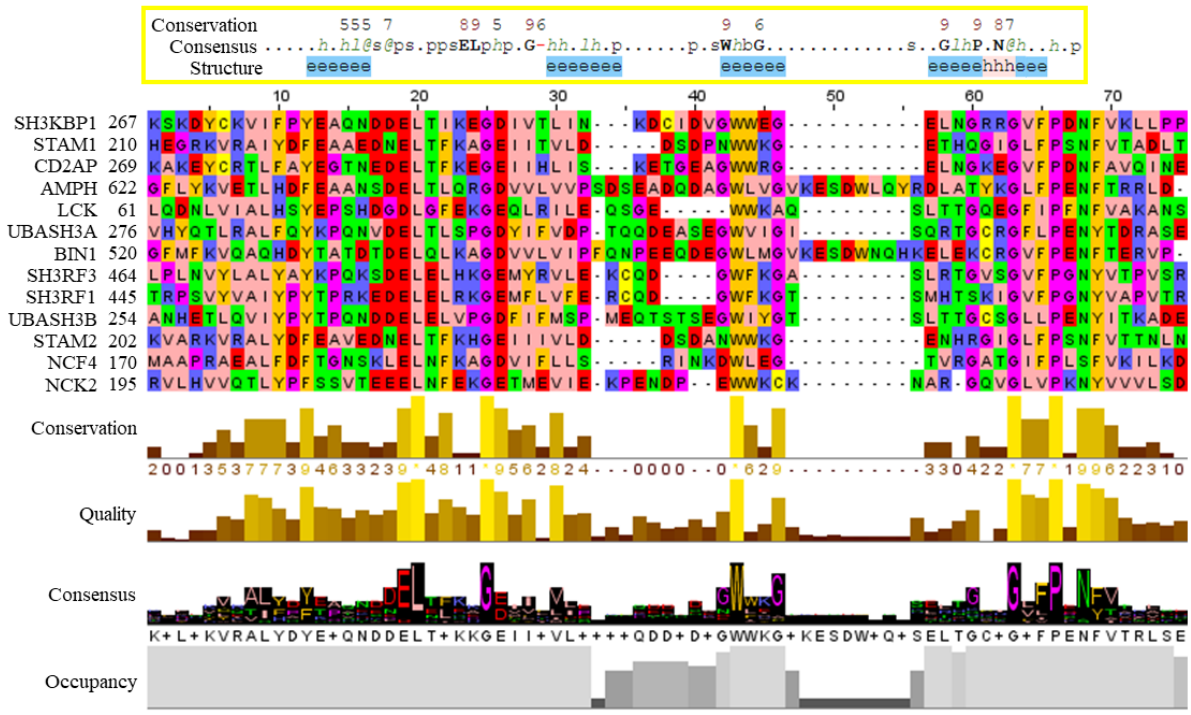


Fig.3.8.2.1. Structural, physicochemical and consensus alignments of human SH3 domains. 13 human SH3 UBDs were determined by cross database comparison, then aligned with PROMALS3D (upper, yellow panel) and Clustal W and X within Jalview (lower panel). Zappo physicochemical properties were indicated by colour within Jalview and alignment conservation, quality, suggested consensus and occupancy calculated automatically. The position of the domain within each protein was indicated on the left by the number of the first amino acid and the relative position within the alignment at the top.

The original SH3 consensus was defined as `A[LI][YF][DN][YF]xxxxxx[ED]lxxxxG[DE]xxxllxxxxxxW[WF]xxxxxxxxG@xPxx[YF]l` (Musacchio *et al.*, 1992). This was well conserved throughout the human ubiquitin-binding SH3 domains but was not comprehensive and did not include all the constraints. By analysing only the human ubiquitin-binding SH3 domains through PROMALS3D and Jalview, this consensus was updated and refined to `A[LI]@xYpxxpxDELpxxxG[DE]x#x#(x6-12)W[l@]xG(x6-16)G@[lF]PxN[YF]#` where (x6-12) represented anywhere from six to 12 random amino acids in a row. For direct comparison of the original and the refined consensus sequences were aligned (Fig.3.8.2.2).

```
Old A[LI][YF][DN][YF]xxxxxx[ED]lxxxxG[DE]xxxllxxxxxxW[WF]xxxxxxxx G@ x  Pxx[YF]l
New A[LI] @   x   Y  pxxpxD E  LpxxxG[DE]x#x#(x6-12)W[l@]xG(x6-16)G@[lF]PxN[YF]#
```

Fig.3.8.2.2. The consensus sequence for the human SH3 UBD. The original consensus sequence of SH3 domains (top line) (Musacchio *et al.*, 1992) compared to the new consensus identified through PROMALS3D and Jalview specific to human ubiquitin-binding SH3 domains (lower line). The shorter motif was underlined in yellow.

The full-length consensus sequence was targeted towards just ubiquitin-binding SH3 domains, rather than the larger family. To aid quick identification of a novel SH3 domain from a protein sequence or by manual identification as opposed to a sequence search, a shorter, less accurate but more easily identifiable motif was defined. This was chosen as the area of the UBD consensus with the greatest difference compared to the original consensus, i.e. the motif most likely to identify an UBD SH3 as opposed to a general SH3 domain. The motif was indicated by the yellow line in Fig.3.8.2.2 and comprised of `W[l@]xG(x6-16)G@[lF]PxN[YF].`

3.8.3 A consensus motif for the CARD family

The CARD family has already undergone considerable structural analysis in a variety of proteins and species, including identification of acidic and basic patches on the domain surface (Qin *et al.*, 1999; Srimathi *et al.*, 2008; Park *et al.*, 2007a; Lopez *et al.*, 2011). These domains were fairly long and formed a six-helix bundle (Park *et al.*, 2007b). The CARD domain was initially identified through structural homology, a trend which has continued when identifying novel members of the family, although alignments have been conducted (Hofmann *et al.* 1997; Park, 2019). As there were many human CARD domains that did not act as UBDs, the protein list from iUUCD and the literature was used to generate a motif rather than the broader family as annotated in the other databases (Fig.3.8.3.1) (Ver Heul *et al.*, 2014; Ver Huel *et al.*, 2013; Jiang *et al.*, 2012; Zeng *et al.*, 2010; Ferrage *et al.*, 2012). There were conflicting reports of NOD2-CARDs binding to ubiquitin. As one publication used Lys63-polyubiquitin and the other GST-monoubiquitin it was decided that the NOD2-CARD domains were at worst candidate UBDs and at best linkage-selective UBDs and so were included in the analysis.

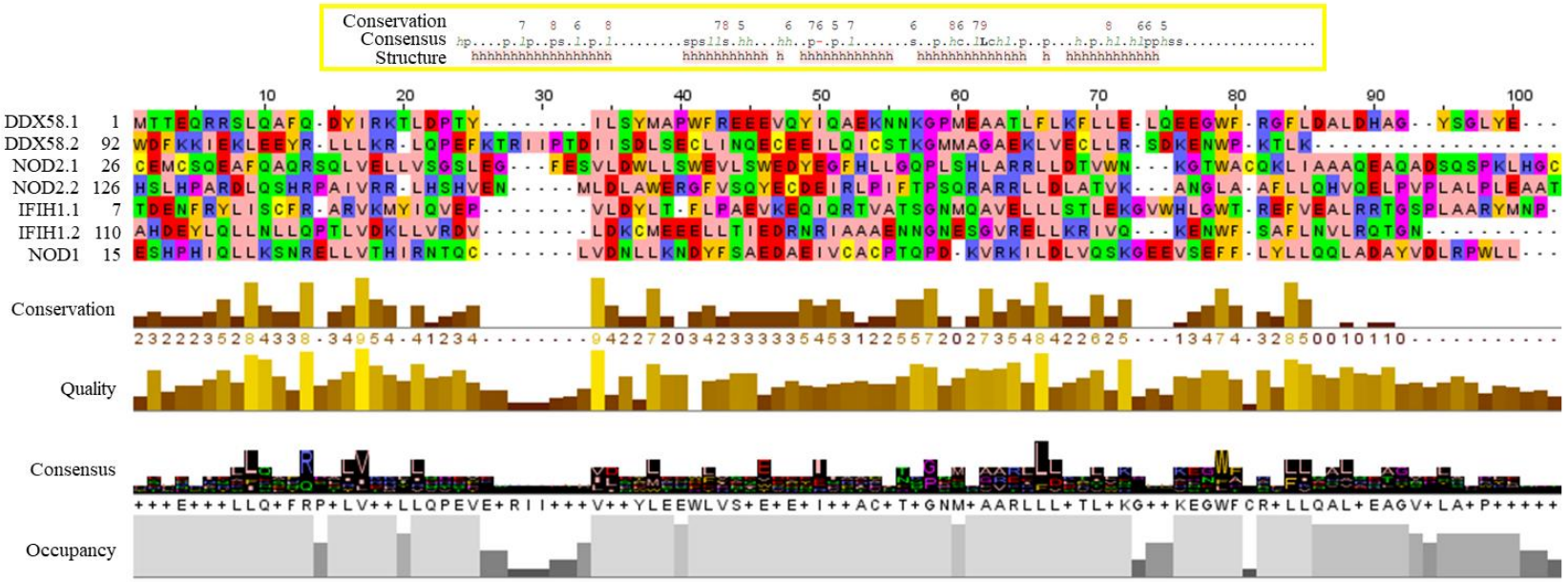


Fig.3.8.3.1. Structural, physicochemical and consensus alignments of human CARD domains. Seven human CARD UBDs were determined by cross database comparison, then aligned with PROMALS3D (upper, yellow panel) and Clustal W and X within Jalview (lower panel). Zappo physicochemical properties were indicated by colour within Jalview and alignment conservation, quality, suggested consensus and occupancy calculated automatically. The position of the domain within each protein was indicated on the left by the number of the first amino acid and the relative position within the alignment at the top.

As was to be expected from a structurally-determined domain, there was little sequence or physicochemical homology between the CARD domains, although the structural homology was highlighted by the PROMALS3D analysis (Fig.3.8.3.1). There were several well conserved residues, for example a Gly/Pro at relative position 58, Leu66, Trp79 and Phe/Leu84. Whilst there were also several points of physicochemical homology, none of these had sufficient conservation to create a usable motif.

3.9 CONSENSUS MOTIFS FOR HUMAN UBDS

To create accurate, human-specific and usable consensus SLiMs for the comprehensive annotation of human UBDs a standardised pipeline was created to identify human UBDs, classify them into the published families and assess them for amino acid, structural and physicochemical homology. This was designed to be easily replicated, allowing for revision of the data as new domains are discovered. The method identified 227 proteins containing 286 UBDs in 19 families across four groups. To ensure that all consensuses were up to date and comparable each family was reassessed, regardless of if a motif had already been published or not. Several of the families with well-known consensuses were found to no longer fit their consensus, either due to incomplete input data or an input of a single protein across multiple species (Table 3.9.1).

Table 3.9.1. Summary of reassessed human UBD consensuses.

Family	Consensus		
	None	Outdated	Suitable
α-Helix	3	4	3
ZnF	1	1	2
UBC-like	2	0	0
Other	1	2	0
Total	**7**	**7**	**5**

Comparison of the popular protein annotation databases SMART, Pfam, iUUCD, Prosite and UniProt identified 19 UBD families with at least two validated human proteins. Literature searches identified those with known (published) consensus sequences or motifs. These were compared to those created through Jalview and PROMALS3D analysis.

It was surprising how many UBD families required a new motif: 74% were either not fit for purpose or non-existent. The families with consensuses judged to be suitable were UBA, MIU, UBAN, NZF and ZnF_A20. These were all members of the α-Helix or ZnF groups, which was to be expected given that they were the most numerous and the most well studied. The number families with no published motif was larger than expected. Some of them were very small families, such as UBM and UEV; the consensuses identified here were considered low-confidence, as two reference points could not reliably describe the trend, i.e. the current pool of proteins with the domain was too small to confidently predict new instances. However, families such as VHS, GAT and UBP contained multiple proteins but had no reliable or accepted consensus. The up to date consensus for each family was presented below, alongside any pre-existing consensus (Table 3.9.2). It is the hope of this author that this pipeline and table form the basis of a resource that will be useful to many busy researchers, providing a quick, reliable and comparable source of motifs and consensuses without requiring additional work to identify the source material, species or domain boundaries.

Table 3.9.2. Consensus sequences for all families of human UBDs.

Family	Old consensus	New consensus
UBA	LxQLxx.MG.FxxxEx.xx ALxxxx...xxAxxxLL (Hofmann, 1996)	Maintained
UIM	-xxhxxAhx[h-]Szx OR LALAL (Swanson *et al.*, 2003)	hxxA[ILM]
CUE	MFP(x24)LL (Prag *et al.*, 2003; Kang *et al.*, 2003; Ponting, 2000)	lxphFPp(x25)LLp
VHS	None	AT(x7)DW(x5)[FI]C[DE]
GAT	None	[DS]x[ED][LA][LM][KQ]x[LV] xxx[CG]
MIU	DφxLAxxLxx...x (Penengo *et al.*, 2006)	Maintained
UBAN	ExxxxLxxQxxxYxxDFxxER xxRE(x5)KxxL(x6)L OR DFxxERxxRE (Wagner *et al.*, 2008)	Maintained
DUIM	-x-xhxhAhAzSz[SA]- (Hirano *et al.*, 2006)	ExExLxLAxAxSxSE
UMI	None	ExxxxLLA
UBM	None	DPxVFxxLP
NZF	xxxxWxCxxCxxxN(x6)CxxC(x5) (Alam *et al.*, 2004)	Maintained
UBP	None	[CA]xx[CES](x4)(.11)pl[WH]x [CN]Lx..[CD]xxx[CQ][GF]x (.10)H#xpH
UBZ	CxxC(x_n)[CH]xDx[CH] (Rizzo *et al.*, 2014)	CP#C(xn)H#xx.[CH]
ZnF_A20	CxxxsCsxhxssxxxshCpxC@ (Wertz *et al.*, 2004)	Maintained
Ubc	None	LDILxxQW
UEV	None	KxxxDxYVFxD
beta-Prp	WD(x40-160) (Li and Roberts, 2001)	SxSxDxwl+lWD
SH3	A[LI][YF][DN][YF]xxxxxx [ED]lxxxxG[DE]xxxllxxxxxx W[WF]xxxxxxxxG@xPxx[YF]l (Musacchio *et al.*, 1992)	W[l@]xG(x6-16)G@[lF]PxN[YF]
CARD	None	N/A

The human UBD families were defined by iUUCD and populated through cross-validation of Pfam, SMART, Uniprot, Prosite and the literature. Consensus sequences and motifs were identified by analysis with PROMALS3D and Jalview and compared to those already published within the literature. The new consensuses presented were human-specific, up to date (as of July 2019), reviewable and, for families defined in contexts other than exclusively ubiquitin-binding (e.g. SH3), UBD-specific.

One family, the CARD domain, was unable to be defined by a strong consensus sequence or motif. The CARD domain was identified structurally: despite all members of the family displaying a conserved surface patches these areas of conservation did not align when considering the primary structures (Qin *et al.*, 1999; Srimathi *et al.*, 2008; Park *et al.*, 2007a; Lopez *et al.*, 2011). This suggested that either it was the physicochemical properties of the surface patches that were important rather than the amino acid themselves, or that the family consists of several sub-families that could be aligned more satisfactorily in isolation.

3.10 APPLICATION OF THE UBD CONSENSUS RESOURCE

As the resource was created to be useful rather than merely informative, a trial run of its suitability and ease of use was performed on three different levels; by scanning a public database for proteins aligning to a new motif, by alignment to a candidate ubiquitin-binding protein, and by scanning a large dataset created later in this thesis. This demonstrated the applicability of the resource to a range of common potential uses.

The first trial searched the human KEGG database via the MOTIF programme for the newly identified motif for UBM domains, `DPxVFxxLP`. This identified the two known proteins, Pol iota and REV1, as well as the tyrosine aminotransferase (TAT) protein. TAT has not been annotated with any ubiquitin-binding domain.

The second trial attempted to identify any known ubiquitin-binding domains within peflin, which was identified as a candidate ubiquitin-binding protein in Chapter 5 of this thesis. The full protein sequence, obtained from UniProt, was searched for each of the 19 UBD consensuses defined in this chapter. The search took only a couple of minutes to quickly scan one protein against 19 families. No homology to any of the 19 families was found, confirming that this candidate ubiquitin-binding protein did not contain a known UBD.

The third trial used a dataset that had been created by panning a randomised peptide library of 10^7 variants of 16 amino acid-long peptides against immobilised ubiquitin to identify novel ubiquitin-binding unbiased sequences. Through this panning and computer-based sequence selection the library was refined to 52 16mer peptides. From these 52 candidates a small number were selected for validation as ubiquitin-binding peptides (Chapter 4). This provided an excellent opportunity to use the resource developed above in a practical setting.

The 52 candidates were each scanned for homology to the 19 UBD families using the motifs and consensuses identified (Table 3.9.2). This scanning of a large dataset for quick and reliable identification of homology to known UBDs was the main purposes of the resource. Whilst the scientific output of this analysis will not be discussed here (c.f. Chapter 4.8), as a trial run of the resource from a user's point of view it was a great success. Searching for potential UBD motifs within the dataset took a matter of minutes to cover all 19 families and only required a simultaneous view of the list of UBDs (Table 3.9.2) and the table of physiochemical properties (Table 3.4.3). The search was conducted on a standard PC with no specialised software. As the search was conducted through the Windows Ctrl+F function, wildcards or searches according to physicochemical property were not possible, and instead all potential combinations of amino acid had to be searched individually. For example, when searching for sequences that aligned to the updated UIM motif, `hxxA[ILM]`, individual searches were conducted for 'AI'. 'AL' and 'AM' with a manual check for the presence of a hydrophobic residue (`h`: Trp, Phe, Tyr, Met, Leu, Ile, Val, Ala, Cys, Thr or His) three positions prior. Despite this, searching 52 sequences for homology to 19 families took only minutes to complete, as opposed to the hours or days previously required to find motifs or consensuses in the literature and confirm their timeliness and species relevance. Of the 52 phage sequences, four were found to show some homology to a known UBD domain, of which one sequence (#18) aligned to two separate UBDs (Table 3.10.1).

Table 3.10.1. Identifying homology within the Next-Generation Phage Display dataset (Chapter 4) using the bioinformatic UBD resource.

Peptide		Motif	UBD
03	ARWLWFTLLAFYKRAM	hxxA[ILM]	UIM
04	ATQAISSEEREIqRTW	hxxA[ILM]	UIM
09	DLMVRALTLAVEKqSL	hxxA[ILM]	UIM
18	HPqWCSLITCSRALLF	CxxxsCsxhxssxxxshCpxC@	ZnF_A20
18	HPqWCSLITCSRALLF	hxxA[ILM]	UIM

52 Next-Generation Phage Display-derived peptides were identified according to frequency and comparative analysis. These were interrogated for homology to the UBD consensus motif resource. Sequences with homology were included in the table with aligned amino acids underlined in both the peptide and consensus. The small q represented the read through from the amber stop codons. This table is a reproduction of Table 4.8.2.

The repeated alignment to UIMs was to be expected, as the UIM domain is typically 12 amino acids long which is neatly encompassed by the 16mer peptides. Whereas alignment to a longer domain or motif, such as that for the SH3 domain which contained an insert of up to 16 undefined amino acids, was highly unlikely given that the 16mer peptides could only ever cover a small part of the domain. Alignment to the UIM motif was also more likely as the motif itself only had three defined residues, the N-terminal hydrophobic, then two amino acids later an Ala followed by Ile, Val or Leu. A basic calculation puts the probability of this small motif occurring by random at 0.41%, where there is equal usage of all 20 amino acids, 11 of which are hydrophobic. The probability of a false positive UIM is calculated by;

$$P_{hydrophobic} = 11/20$$

$$P_{x} = 20/20$$

$$P_{I/V/L} = 3/20$$

$$P_{false} = ((11/20) \times (20/20) \times (20/20) \times (1/20) \times (3/20)) \times 100$$

$$P_{false} = 0.41\%$$

The quick filtering of 52 to four peptides for validation demonstrated the ease of use of this resource. Even whilst not being able to use wildcards for physicochemical searches, checking

52 sequences individually for homology to 19 UBD families took a matter of minutes. For searching larger datasets or for whole proteome searches, online tools such as MOTIF can accept wildcards, enabling quick and reliable identification of proteins aligned to the 19 UBD families.

3.11 CONCLUSIONS

The primary aim of this chapter was to create a standardised, reliable list of human ubiquitin-binding SLiMs as a resource. Whilst many of the UBD families have published consensus sequences or motifs, they are often partial or not updated in the face of new knowledge. As consensuses are only as good as the knowledge they are based upon, these consensuses were reassessed through a standardised pipeline to fill the gaps and create reliable, current and easily updatable human motifs.

As a first test of the resource, this was highly successful and neatly worked towards the hypothesis of this chapter; that a standardised method for the identification and bioinformatic validation of ubiquitin-binding SLiMs could be used to identify novel instances of UBDs. A systematic analysis of the human proteome using these new motifs was beyond the scope of this chapter as to do so would move the focus away from resource development, however, the new motif for the UBM domain was checked by searching the human KEGG database via the MOTIF programme and analysis of an existing dataset for putative UBDs was conducted as proof of concept.

As such the resource could be used to search for novel instances of known UBDs and potentially identify novel ubiquitin-binding proteins. It could be used for quick literature searches, for manual identification of a UBD within a given protein sequence or for highlighting potential UBDs or ubiquitin-binding proteins from a large dataset for future validation. This approach will no doubt yield a high number of false positives, especially for

the larger domains with greater variability, but in combination with existing knowledge and experimental data, it could provide a valuable resource for groups to identify viable candidates for validation. This would be particularly valuable when experiments have resulted in large datasets, such as so-called fishing exercises and large-scale prediction or interaction analyses e.g. shotgun proteomics.

However, there remains the biases discussed at the start of this chapter, namely that consensus sequences are only ever as accurate as the current knowledge and they can only identify sequences that share sufficient sequence identity. In other words, they are very poor at identifying truly novel or diverse sequences. The issue of being based on current knowledge can be mitigated by frequent review of the consensus motifs, such as was presented here. However there remains significant confirmation bias, in that a user will only ever be able to identify new instances of known domains, rather than any truly novel UBDs. In order to address this, a different experimental approach would be required, one with greater randomisation of input material.

4 NEXT-GENERATION PHAGE DISPLAY TARGETING MONOUBIQUITIN TOWARDS GENERATING NOVEL UBIQUITIN-BINDING CANDIDATE PEPTIDES

4.1 INTRODUCTION

The bioinformatics approach presented in Chapter 3 created a potentially useful resource for the identification of known UBD families within previously unannotated ubiquitin-binding proteins. However, bioinformatic approaches alone are not suited to identifying novel families of UBDs, as they are limited by the current knowledge of ubiquitin recognition. Prediction programmes or data processing strategies can only reliably identify domains that are already well-characterised, and in that manner are often self-affirming and restrictive. In order to identify novel ubiquitin-binding sequences with little to no investigator bias, the alternative approach of Phage Display was used.

The standard Phage Display experiment provided an attractive combination of low bias and high throughput. It has been used to great success in the identification of ubiquitin variants with enhanced binding to known UBDs, E3 ligases and DUBs (Manczyk *et al.*, 2017; Zhang *et al.*, 2016; Ernst *et al.*, 2013 respectively). It has also recently been used to identify the endogenous CUBAN (Castagnoli *et al.*, 2019) and CoCUN (Nepravishta *et al.*, 2019) domains. These domains, previously discussed in the Introduction (Section 1.4.1 and 1.4.2, respectively), were identified from a human cDNA library simultaneously targeting both the ubiquitin-like protein NEDD8 and ubiquitin itself. In recent years Phage Display has been combined with Next-Generation Sequencing (NGS) to provide broad sequence coverage of large peptide libraries.

This led to the hypothesis that reversing this experimental design could create a carefully designed NGPD experiment to identify novel ubiquitin-binding sequences from a high-diversity randomised peptide library. By targeting both ubiquitin and phosphoubiquitin binding

sequences selective for a particular ubiquitin modification could be identified. In contrast to the bioinformatics approach, any novel UBDs resulting from NGPD were far more likely to be artificial than endogenous, although it was possible to identify an endogenous domain through NGPD, for example through cross-referencing identified sequences with known proteomes.

Bacteriophages, or phages, are viruses that infect every known type of bacteria. They are the most numerous biological entities, being found in the most mundane to the most extreme environments. One commonly used phage is the non-lytic filamentous M13. It had several advantages as a research model; it is a small bacteriophage with single stranded DNA, it has a non-lytic lifecycle and so secretes its high titre of progeny phage without killing the host (Deng *et al.*, 1999). It infects bacteria with F pili, preferably gram-negative, including species frequently used in research, such as *Salmonella*, *Pseudomonas* and *Escherichia* (Russel and Model, 2006; Haq *et al.*, 2012).

In a NGPD experiment, a randomised peptide library is expressed on the surface of bacteriophage. Each bacteriophage only expresses a single sequence of peptide although the copy number, the number of displayed peptides per phage, varies with the system used. For the M13 phage the capsid protein commonly used to display peptides, pIII, has a copy number of five. This library serves as the pool of potential binding sequences and allows a direct link between the binding domain and its encoding DNA. The expressed library is grown in infected bacterial culture, harvested and incubated with the immobilised target protein, in this case ubiquitin or phosphoubiquitin, in a process known as biopanning. Any phages that express a binding peptide are retained on the immobilised surface whilst non-binding phages are removed by washing. The bound phages are recovered and amplified in bacterial culture for another round of biopanning. Through repeated rounds non- or weakly-binding peptides are removed, and stronger binding peptides are enriched within the population. Increasing the strength or number of washes increases selection pressure as the experiment progresses and

encourages selection of high affinity binding peptides. Once the desired number of biopanning rounds have been completed, the phage cultures are harvested and the peptide-encoding DNA isolated and sequenced through NGS. From this data, careful analysis can yield a shortlist of candidate peptide sequences that may be validated into novel binding domains (Fig.4.1.1).

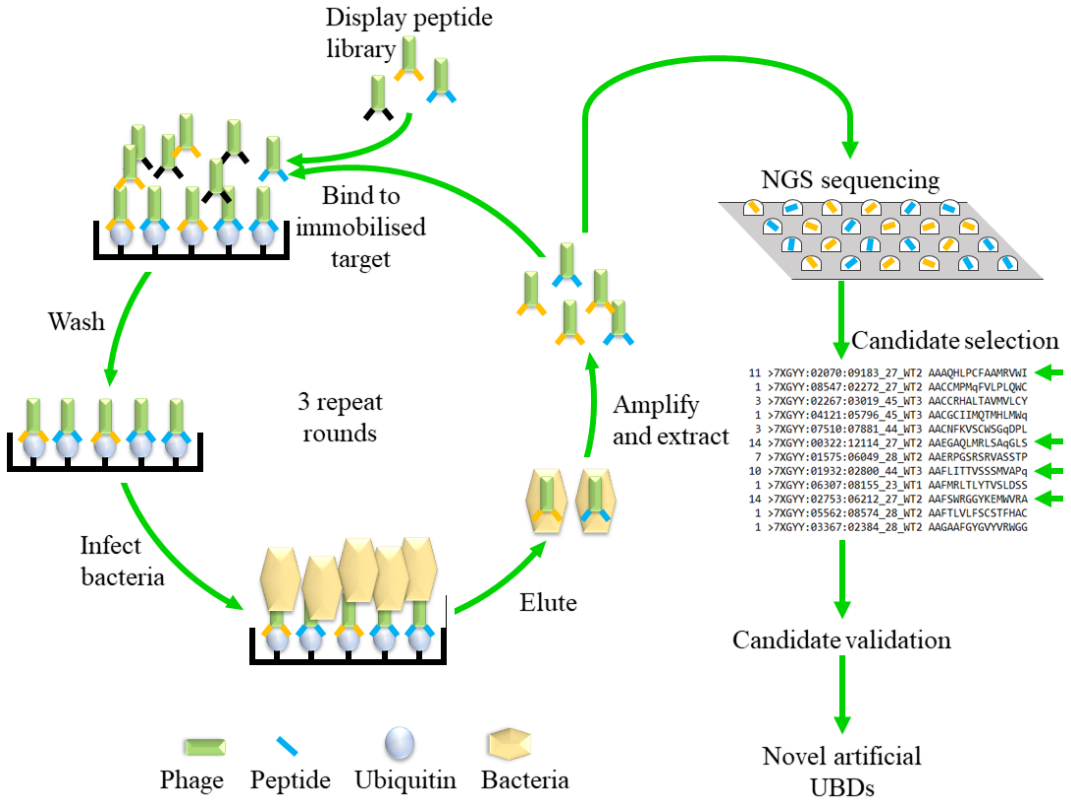


Fig.4.1.1. A flow diagram of the basic NGPD experiment. A randomised peptide library is displayed on the bacteriophage surface. Phage cultures are panned against target ubiquitin, and non-binding phages removed by washing. Binding phages are eluted by infection in bacteria and the process repeated to increase selectivity. Samples from each eluate are Next-Generation sequenced and converted to peptide sequences. Candidates are selected for biochemical validation as novel artificial UBDs.

The advantages of NGPD are numerous. The initial stages are economical and require little specialist equipment, being based around bacterial culture. The biopanning and selection process is comparatively rapid, taking just a few weeks from naïve library to selective candidate culture. There are also many libraries available, either commercially or through collaborations. Importantly, when conducting a NGPD experiment the investigator does not know which

sequences are displayed or enriched until after the experiment has been completed, removing any potential for unconscious or confirmation bias.

As is often the case with such experiments, it can be easy to justify any number of data processing approaches and so it is important to design the experiment to fit the analysis from the very beginning. A strong candidate peptide could be one that occurs multiple times within one sample, or it could occur infrequently in multiple experiments. The power of the selected analysis method depends on how the samples were generated and how each experiment was designed. For the purposes of this thesis, one experiment was defined as all samples and sequences that derived from one initial pan. To illustrate this in the example growth diagram below each experiment is in a separate colour, representing Ub *n*=3 and pUb *n*=3. (Fig.4.1.2).

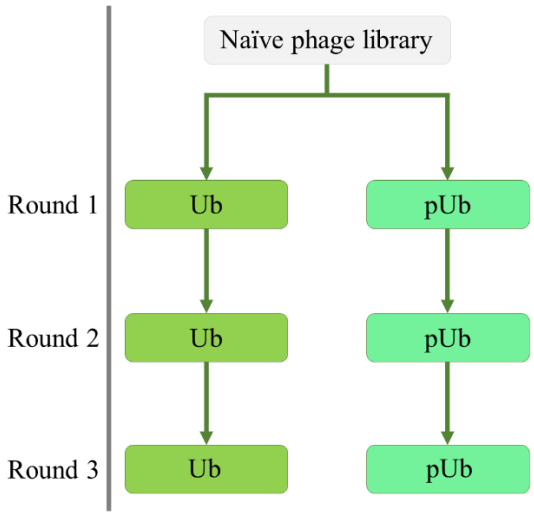


Fig.4.1.2. An example growth diagram for a hypothetical NGPD experiment. The naïve phage library (grey) was panned in Round 1 against two targets; ubiquitin (Ub) and phosphoubiquitin (pUb). The phages bound to each of these targets were recovered and panned again in Round 2 and so on. One experiment was defined as originating from the same initial Round 1 pan – each experiment was indicated by a different colour.

Phage Display is sensitive to single amino acid differences, allowing for identification of artificial sequences that are selective for a variety of mutants. In this instance, phosphomimetic S57E- and S65E-ubiquitin were used alongside pS65- and WT-mono ubiquitin to generate both wildtype- and phospho-selective novel UBDs. This allows the selection of novel UBDs that bind ubiquitin *via* domains not targeted to Ser57 or Ser65, which should theoretically appear

in pans targeting all ubiquitin variants. By using both phosphomimetic- and phosphoubiquitin the suitability of the former to act as a cheaper alternative to pS65-Ub can be assessed. At the time of experimentation, pS57-ubiquitin was not readily available and so no comparison of S57E- and pS57-ubiquitin was attempted.

To test the hypothesis that artificial novel UBDs could be extracted from a reversed NGPD experiment, the following aims were set;

1. To develop a reversed NGPD protocol, where ubiquitin was the target and the phage displayed peptides represented the pool of potential UBDs
2. To scan an appropriate naïve phage displayed library for UBDs
3. To select candidate UBDs from the resulting large NGS dataset
4. To validate these ubiquitin-binding candidates through alternative methods.

To address these aims ubiquitin variants were produced in *E. coli* and purified through FPLC and used as immobilised targets in several different NGPD protocols. The optimised protocol was then used to select seven candidate 16mer UBDs from a naïve phage displayed library, which underwent initial characterisation and ubiquitin-binding functional validation through NMR.

4.2 PRODUCTION OF UBIQUITIN

Throughout this thesis a variety of ubiquitin variants were used. pS65-ubiquitin was commercially sourced, and all other monoubiquitin variants were produced and purified in house. As an example, the growth and purification results for production of wildtype-ubiquitin (WT-Ub) were shown below (Fig.4.12.1). Cultures were grown from *E. coli* BL21 glycerol stocks containing plasmids for production of untagged human ubiquitin. These originally encoded yeast ubiquitin but were back mutated to the human sequence, which only differed by three amino acids (Sharp and Li, 1987), by Dr Jed Long, University of Nottingham. Bacterial

lysates containing expressed proteins were FPLC chromatography purified, then desalted prior to lyophilisation and use. In the absence of an affinity tag it was necessary to use at least two different chromatography purification steps. Clarified lysates were first precipitated at pH 4.0, separated through HiTrap™ SP HP column cation exchange, and the fractions of interest refined through HiLoad™ 26/600 Superdex™ 75 pg gel filtration gel purification, desalted and lyophilised prior to use. The standard yield was 5-10mg protein from 1L culture. This was slightly lower for all mutants and labelled proteins.

To confirm that ubiquitin was present in the final fractions, samples from each stage of the purification process were separated by SDS-PAGE and total protein Coomassie stained. To confirm that the final purified protein was ubiquitin, a sample was run on an SDS-PAGE gel and detected by Western blot (Fig.4.2.1).

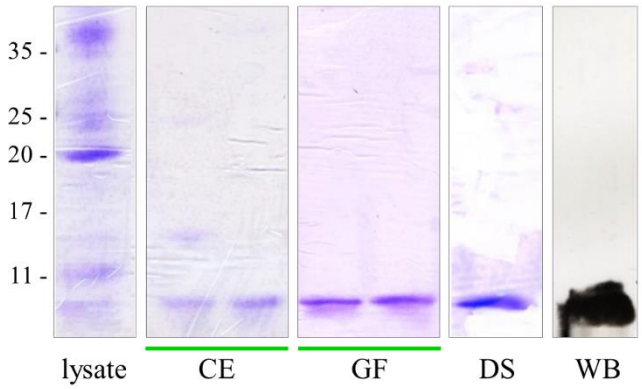


Fig.4.2.1. Confirming purification of ubiquitin. Samples from the fractions suspected to contain ubiquitin from the clarified lysate, cation exchange (CE), gel filtration (GF) and desalt (DS) were separated by SDS-PAGE and Coomassie stained. The purified protein was run on an SDS-PAGE gel and detected by Western blot (WB) with anti-ubiquitin. Ubiquitin was expected at ~8.6kDa.

4.3 SET 1: BIOPANNING AGAINST PHOSPHOMIMETIC UBIQUITIN

With the aim of identifying novel artificial UBDs specific to WT and phosphomimetic ubiquitin, a 10^7 M13 library of pIII-displayed randomised 16mer peptides was grown in TG1s and used as the naïve input for biopanning with the previously optimised (data not shown) ubiquitin-targeting biopanning protocol. The targets were WT-, S57E- and S65E-ubiquitin laid

at 1μg/mL on Amino plates. Correct immobilisation of ubiquitin on the plate was confirmed by ELISA.

Using the naïve library of $1.07x10^{14}$ pfu/mL as the starting input, three sequential rounds of panning were carried out against the three targets and control. Rounds 2 and 3 of the S57E- and S65E-ubiquitin experiments also included subtractive panning; a pan where the phages were pre-blocked against WT-ubiquitin to weight the experiment towards phage peptides that selectively bound S57E- or S65E-ubiquitin over WT-ubiquitin. Pre-blocking was achieved by incubating the prepared phages with 10μg/mL WT-ubiquitin (i.e., ten times the immobilised target ubiquitin concentration) in 200μl 18% Marvel in PBS per mL of phage supernatant during the standard phage blocking step. This blocked any displayed peptides that bound wildtype-ubiquitin, theoretically leaving only those peptides that bound to mutant ubiquitin variants available for biopanning. This subtractive panning was denoted as '-sub' in all diagrams included below.

Overall this biopanning set targeting wildtype and phosphomimetic ubiquitin resulted in 22 samples that were later subjected to NGS for candidate ubiquitin-binding peptide selection (Fig.4.3.1).

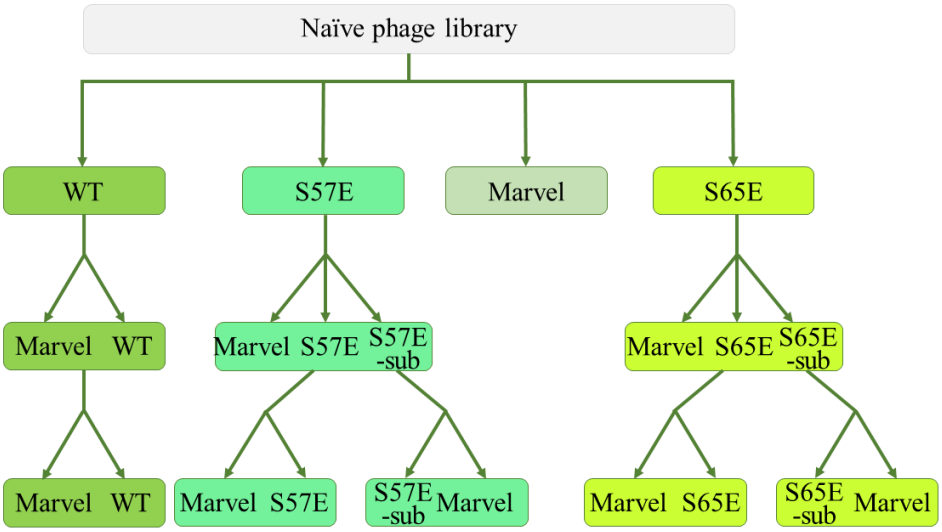


Fig.4.3.1. Generating Set 1 Phage Display peptides targeting WT-, S65E- and S57E-Ub. A naïve 10^7 pIII M13 phage library was panned against three ubiquitin targets for three sequential rounds. The eluate from one round became the input library for the next. Marvel non-specific controls were included at every round. Rounds 2 and 3 included pans with a subtractive block against WT-Ub before exposure to the target phosphomimetic ubiquitin (-sub). Samples from each pan were sent for NGS. WT *n*=3, WT marvel *n*=2, S57E *n*=3, S57E marvel *n*=2, S57E-sub *n*=2, S57E-sub marvel *n*=1, S65E *n*=3, S65E marvel *n*=2, S65E-sub *n*=2, S65E-sub marvel *n*=1, marvel *n*=1.

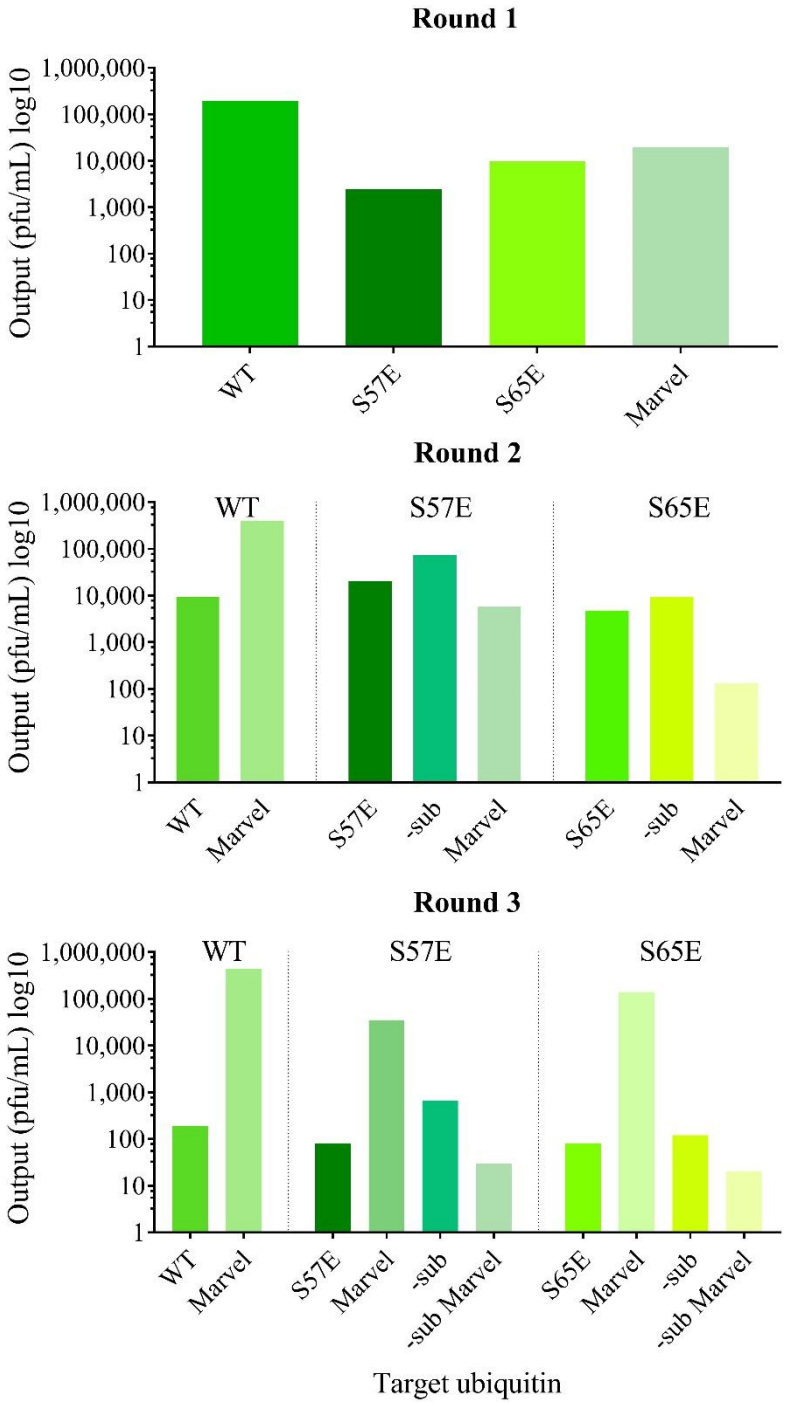


Fig.4.3.2. Set 1 biopanning targeting WT- and phosphomimetic-ubiquitin. A naïve 16mer phage library of 1.07 x10^{14} pfu/mL was panned for three sequential rounds against WT-, S57E- and S65E-ubiquitin with matched Marvel controls. Rounds 2 and 3 included subtractive blocking against WT-ubiquitin (-sub). The eluate from one round became the input for the next. Phages expressing binding peptides were eluted, growth titred and displayed as pfu/mL log10.

The levels of eluted TG1 growth were monitored at every stage of the biopanning by titration on agar plates and colony counting (Fig.4.3.2). Whilst it could be assumed that the pfu/mL of ubiquitin-targeting cultures should increase in titre relative to non-specific Marvel controls as the rounds progress, this was not always the case. Due to phage bias towards or against certain displayed peptides, the growth level was an indication only.

Phage bias can be broadly divided into two major categories; censorship, where certain sequences are supressed, and parasitic, where sequences are over-amplified. Phage libraries can be censored under certain conditions; Arg-rich sequences fused to the pIII protein of filamentous phage have been observed to supress expression by a factor of 10^2-10^6 (Peters *et al.*, 1994). M13 phage have also been linked with preferences for β-turn conformations, which may work against certain experimental aims (Steiner *et al.*, 2006; Rodi *et al*., 2002). Parasitism has been shown to increase abundance of a sequence from the expected 0.000,000,1% to 1% upon standard amplification (Matochko *et al.*, 2014) and can sometimes occur as a counter measure to censorship. Whilst resources such as PhD7Faster 2.0, PuLSE and PHASTpep can be used to identify such sequences and reduce the false-positive rate, these function on a library-specific basis (He *et al.*, 2019; Shave *et al.*, 2018; Brinton *et al.*, 2016). For this library, the skew caused by parasitic phages was mitigated by careful use of negative controls and comparative data analysis.

The level of growth was important to note, however, as it determined the amount of DNA present in a sample and the number of available reads during NGS. Accordingly, whilst preparing samples for NGS the levels of growth and DNA concentration were taken into account to correct skew in the reads (see Section 4.6 for more details).

With the NGPD protocol developed, further optimisations were made after the proof-of-concept pan against WT- and phosphomimetic-ubiquitin, particularly in regards to equalising growth levels between cultures.

4.4 PROTECTING AGAINST VARIATION IN COLONY GROWTH

After the first set of biopanning experiments it was noted that colony growth could vary considerably. Even when using the same phage preparations and keeping all other parameters equal, the eluate growth could change several fold from one day to another. Anecdotal evidence from the other members of the Gough laboratory suggested that one in every six phage cultures failed to grow. This could be due to some unknown factor during growth, contamination or variations in pipetting and washing procedures. However, it was generally accepted that this was impossible to control, prevent or correct.

To accommodate for this variability and to build resilience into the data, a repeat step was included in the second set of biopanning experiments targeting WT-, S65E- and pS65-Ub. Round 1 was conducted as normal, but from Round 2 onwards each pan was conducted twice from the same phage preparation on consecutive days. In this manner, if one pan failed to grow correctly then it could be compensated for by the repeat pan. The harvested eluates from each repeat were combined and used as a single input for the next round (Fig.4.4.1).

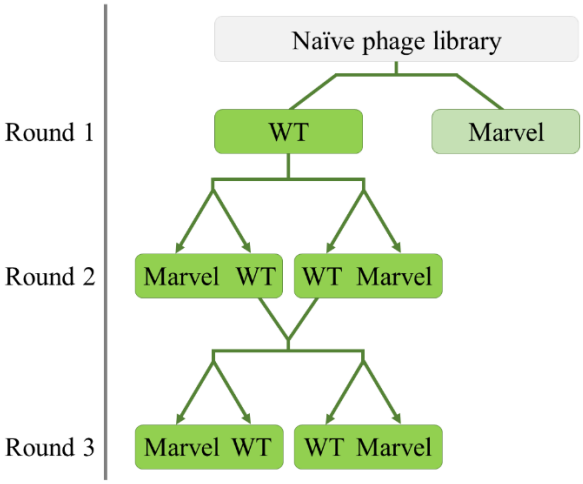


Fig.4.4.1. Growth flow chart for repeated biopanning against WT-ubiquitin. A schematic diagram of the growth from the naïve phage library to the final Round 3 samples. In Round 1 the naive library was panned once against WT-Ub and a non-specific Marvel control. The eluate was used as the input for Round 2, where it was panned twice against WT-Ub and Marvel controls. The eluates of both WT-Ub pans were combined and used as the input for Round 3, when it was again panned twice against WT-Ub and Marvel controls. WT n=5, WT marvel n=4, marvel n=1.

A proof of concept growth was conducted (data not shown) panning the same phage preparation against WT- and S65E-Ub twice on consecutive days. This demonstrated that the variable culture growth was not an artefact of the ubiquitin target and was not due to the phage culture only containing low affinity binding peptides. By pooling the eluates from each repeat pan the *n* number of experiments was quickly increased. It also demonstrated that the repeated panning protocol yielded more consistent growth levels, protecting the experiment from growth failure.

4.5 SET 2: BIOPANNING TARGETING PHOSPHOUBIQUITIN

Phosphomimetic ubiquitin is frequently used as a cheaper and more convenient alternative to native phosphoubiquitin (for example, see Lee *et al.*, 2017). After the optimised NGPD protocol was trailed against phosphomimetic-ubiquitin, a second set was panned with phosphoubiquitin to create more physiologically relevant data. At the time of conducting these experiments, pS57-Ub was not readily available and so only pS65 was targeted. WT- and S65E-Ub both acted as controls and to increase *n* number from Set 1.

The same naïve 10^7 pIII 16mer library was grown in TG1s and used as the naïve input for 3 sequential rounds of biopanning with subtractive and repeat panning. The targets were WT-, S65E- and pS65-ubiquitin laid at 1μg/mL on Amino plates. NGS samples were taken separately from each replicate, and the eluates from one round were combined to create the input for the next (Fig.4.5.1). Immobilisation of ubiquitin was confirmed at every stage by ELISA.

This gave a total of 36 samples for NGS. The growth levels were monitored and calculated as pfu/mL by titration as before (Fig.4.5.2).

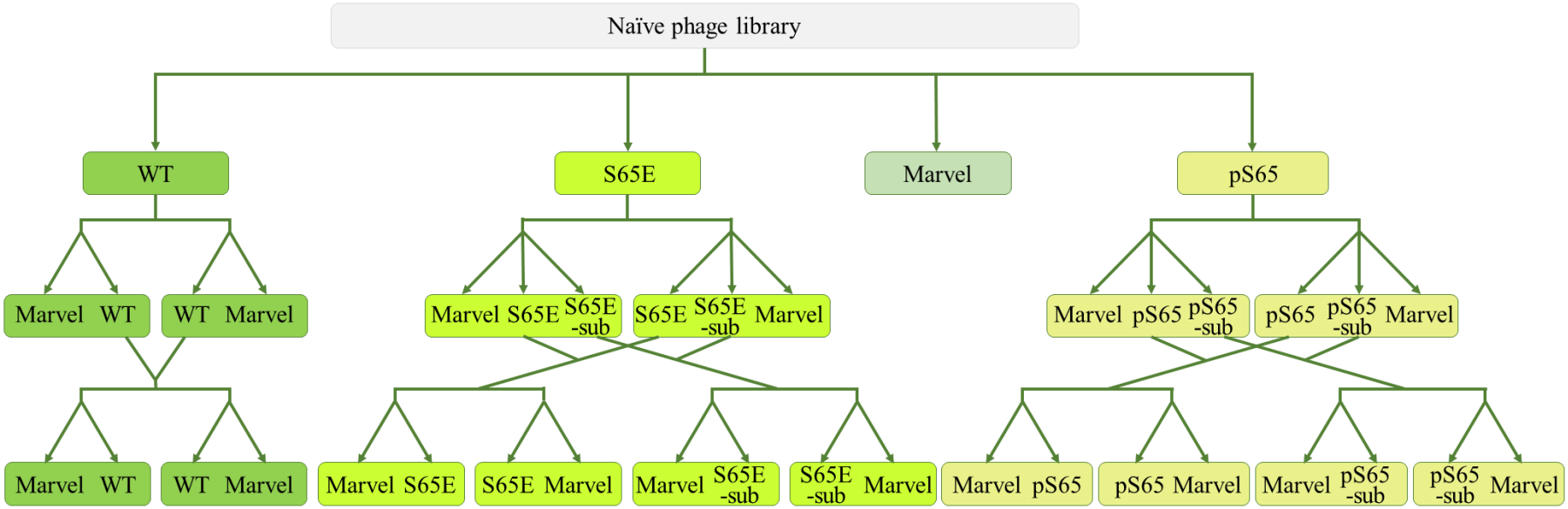


Fig.4.5.1. Generating Set 2 Phage Display peptides targeting WT-, S65E- and pS65-Ub. A naïve 10^7 pIII phage library was panned against three ubiquitin targets for three sequential rounds. The eluate from one round became the input library for the next. Marvel non-specific controls were included at every round. Rounds 2 and 3 included duplicates and pans with a subtractive block against WT-Ub before exposure to the target phosphomimetic ubiquitin (-sub). Samples from each pan were subjected to NGS. WT *n*=5, WT marvel *n*=4, S65E *n*=5, S65E marvel *n*=4, S65E-sub *n*=4, S65E-sub marvel *n*=2, pS65 *n*=5, pS65 marvel *n*=4, pS65-sub *n*=4, pS65-sub marvel *n*=2, marvel *n*=1.

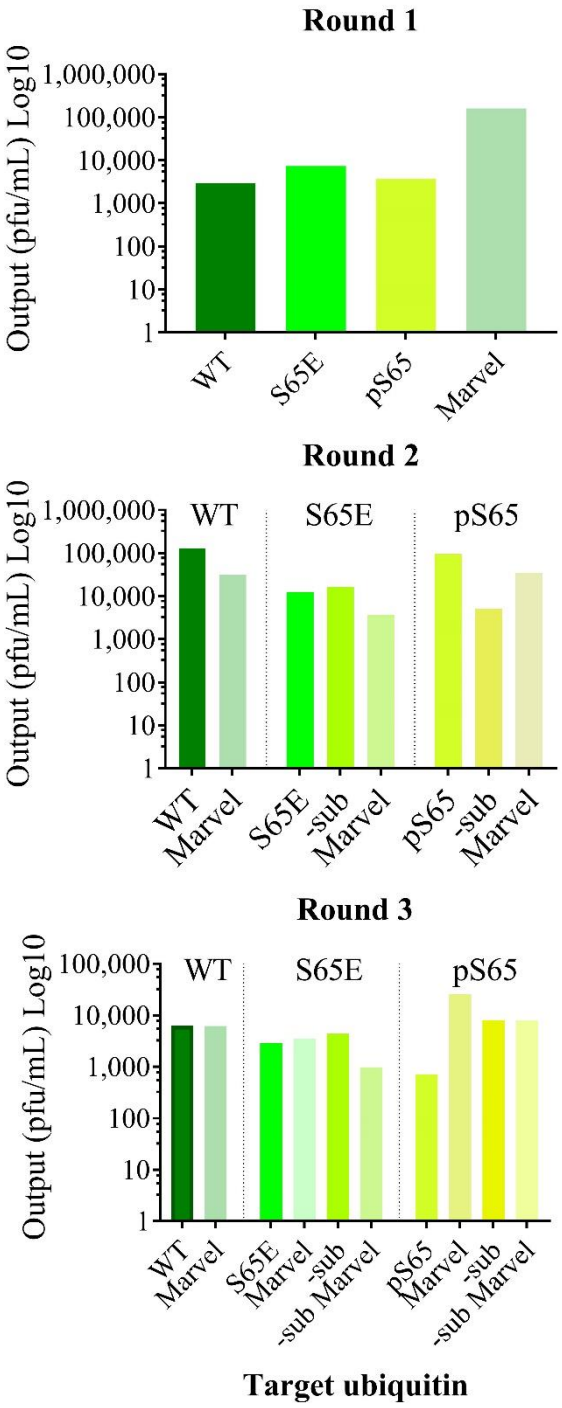


Fig.4.5.2. Set 2 biopanning targeting WT- and phosphoubiquitin. A naïve 16mer phage library of 1.05 x10^{14} pfu/mL was panned for three sequential rounds against WT-, S65E- and pS65-ubiquitin with matched Marvel controls, subtractive panning to WT (-sub) and replicate pans. The eluates from both replicates were combined and became the input for the next. Binding peptides were eluted with their expressing phage, growth titred and displayed as pfu/mL log10. Data are presented with replicates combined and averaged.

The same naïve phage library that was used for Set 1 was used here and grown to an input of 1.05 $x10^{14}$ pfu/mL, which is comparable to the input in Set 1. Not much information can be gleaned from the growth titres before NGS, but the even growth levels between conditions initially suggested that sufficient phages had survived to yield candidate ubiquitin-binding peptides. In part this was likely due to the inclusion of the repeated panning protocol. Samples from each of the 40 pans were subjected to NGS to work towards identifying the candidate peptides.

Combining Set 1 and Set 2 yielded six independent experiments, with each experiment being formed of all pans that originate from a single Round 1. These six experiments can be broken down into two WT-Ub, one S57E-Ub, two S65E-Ub and one pS65-Ub targeting experiments. Each experiment was formed of many pans from three rounds with and without replicates. Samples were taken from each pan and the phage peptide-encoding DNA isolated and processed before being sent for NGS. In total this represented 62 samples; 22 from Set 1, and 40 from Set 2.

4.6 NEXT-GENERATION SEQUENCING OF SELECTED PHAGE DNA

Samples from each biopanning round and experiment were taken, and the DNA extracted, amplified and cleaned for NGS (Fig.4.3.1 and Fig.4.5.1). Each of the resulting DNA sequences could be traced back to its originating sample, and therefore experimental condition, target ubiquitin and round, by means of labelling all DNA sequences within one sample with a unique barcode. The barcoding principle involved incorporation of an identifying DNA sequence into the phage dsDNA. As one barcode was used for each sample this encoded the experimental conditions (i.e., ubiquitin target, round number etc.) within the peptide DNA itself, tying the conditions directly to the resulting peptide. For a list cross-referencing the barcode number, sample and experimental conditions, please see Appendix Fig.A4.1.1.

As there is some overlap and redundancy of terms in the sequencing field, in this thesis the term 'read' is used to describe each time the sequencer detected a particular sequence. It should be noted that if a sequence was reported with 100 reads, it was impossible to know from the output data if that represented 100 reads of a single phage's DNA or if it represented 1 read from 100 different phages each encoding the same DNA. For this reason it was important to note the pfu/mL of the original phage-infected TG1 sample sent for sequencing, as 100 reads of a sequence from a culture of 100 pfu/mL would suggest that each phage was read once, whereas 100 reads from a culture of 5 pfu/mL would suggest multiple reads of each single sequence.

The cleaned and barcoded DNA from all 62 samples was pooled and subjected to NGS sequencing at the University of Pennsylvania Perelman School of Medicine's DNA Sequencing Facility. NGS was conducted on an Ion Torrent S5 with a 540 chip, giving 1,242,184 expected reads, each with a max read length of 200bp and basic data compilation was performed by the facility.

4.7 SELECTION OF NGPD-DERIVED CANDIDATE UBD SEQUENCES

To ensure the most appropriate and powerful analysis of the DNA sequences recovered from NGS, a tailor-made programme was written for basic sequence processing (courtesy of Dr Jed Long) (Fig.4.7.1). The NGS sequences were translated to amino acids, the non-peptide encoding sequences removed and the frequency of each sequence counted within each sample. The Ala-Ser N-terminal phage pIII leader sequence, which was present in the display of peptides on the phage surface and could have affected peptide-ubiquitin interaction, had been removed during sequencing so this appended back to the N-terminal end of the peptide sequence. The data set was split into each barcode and linked back to the originating sample and pan conditions for future reference.

The total number of NGS reads across the 62 barcodes presented here was 2,585,011, which averaged at 41,694 per sample and nine per sequence. This was double the expected total reads, so even though there was a low number of average reads per sequence this was likely due to the diversity of the panned library. As this panned library included Round 1 and non-specific Marvel samples this high diversity was to be expected. The dsDNA weighting conducted in section 4.6 meant that the majority of reads were likely to be conducted on samples targeting ubiquitin rather than non-specific Marvel samples. Removing the non-specific samples, there was an average of 39,242 reads per targeted sample. This was triple the 13,076 reads expected, suggesting good coverage and depth of sequencing and giving confidence in the accuracy of the resulting sequences (Table 4.8.1). The pfu/mL and dsDNA submitted to NGS and the reads and sequences per sample are included in the Appendices (Table A4.1.1).

In order to increase the probability of identifying a high affinity candidate ubiquitin-binding sequence, all sequences were grouped according to their target ubiquitin variant and were processed accordingly. For example, all sequences targeting pS65- and S65E-ubiquitin, with and without subtractive panning, were grouped for processing. Due to the unique barcoding sequences could still be traced back to their original experimental conditions as required, meaning that no loss of information or selectivity was incurred by this grouping of sequences.

Table 4.7.1. Successful NGS of NGPD-derived peptide sequences.

Target	*n*	Average pfu/mL per sample	Total Reads	Total 16mer Sequences
WT	8	5.92 $X10^4$	363,463	52,398
S57E	5	1.93$x10^4$	226,196	15,286
S65E/pS65	23	1.41$x10^4$	823,048	80,695
Total	36	3.44$x10^4$	1,412,707	148,376

Phage displayed peptides targeted at immobilised ubiquitin were sent for NGS. This yielded a high number of sequences per sample with good sequencing depth.

From this, peptides were selected as candidate artificial UBDs, including at least one sequence that targeted WT-, S57E- and pS65/S65E-ubiquitin. Samples that targeted pS65- and S65E-ubiquitin were combined to give a greater pool of potential selective peptides. There are multiple approaches to selecting candidates from NGS data, any of which can yield suitable candidates, so method choice was particularly important.

Calculating the Z-score of a given sequence is frequently used in candidate selection to estimate enrichment of a sequence compared to the expected background abundance. For a high-powered Z-score analysis the data must come from independent populations, with an underlying Gaussian distribution and high number of repeat samples per experiment, none of which were possible with the method developed here given the logistical constraints and need for duplicate pans (Naqid *et al.*, 2016; Yuan *et al.*, 2018).

Another common method is selection of the highest frequency sequences. However, this is highly susceptible to data skew by phage censorship or parasitism which can artificially increase or decrease the frequency of a given sequence. It is also highly dependent on enrichment of just one or two binding sequences being enriched, which is normally only achieved through experiments with more selection rounds and higher selection pressures. This method was not suitable, as this experiment was designed to select for UBDs, which have comparatively low affinities and so excessive rounds or selection pressures were avoided. Instead a combination of comparative and biophysical profiling approaches was chosen (Fig.4.7.1).

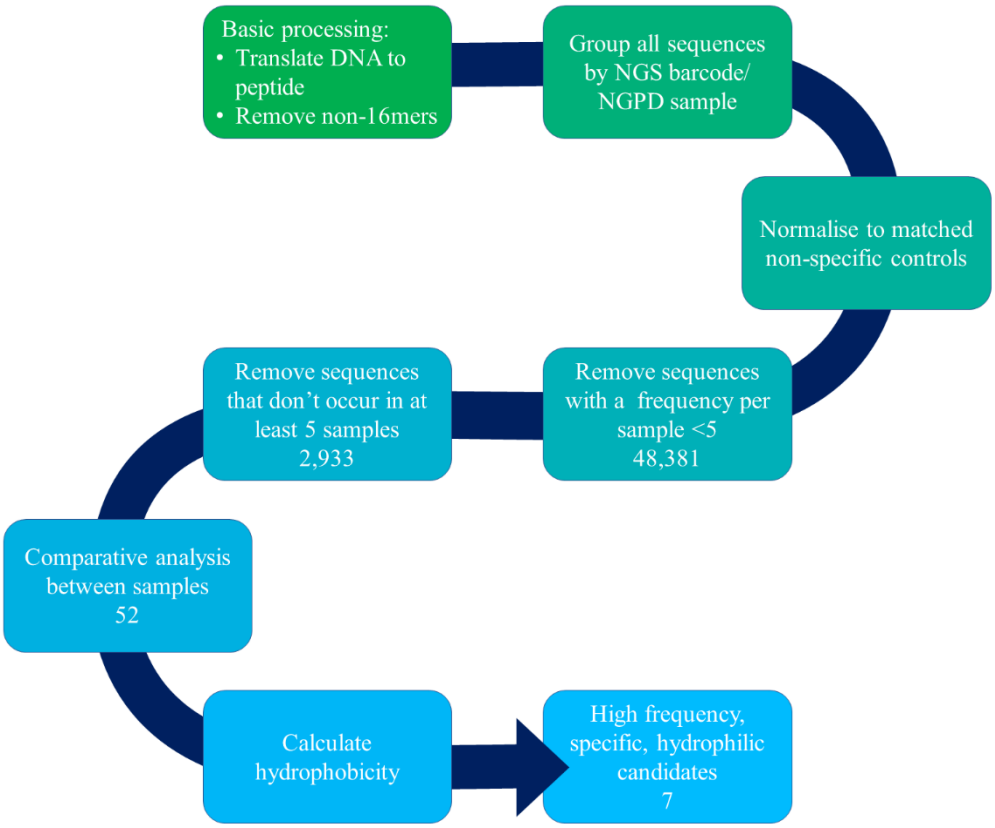


Fig.4.7.1. Flow chart of sequences processing from NGS to candidate selection. NGPD against immobilised ubiquitin yielded an excess of sequences representing potential UBDs. To filter these and select candidate UBDs for validation, an eight-stage processing stream was created, resulting in seven candidates. The number of sequences after each major processing step is indicated.

First, any sequences that occurred in both the experimental ubiquitin-targeting and the matched non-specific Marvel sample were discarded, as they presented non-specific binding. Any sequence that occurred less than five times in a sample was removed. Any sequence that did not occur in at least five separate samples targeting the same ubiquitin variant was also discarded. This filtered for sequences with a frequency above five in at least five samples, which were more likely to bind specifically. As seen in Fig.4.7.2 there was a sharp drop off in sample frequency: the vast majority of sequences occurred in fewer than five samples, and so these criteria quickly selected for the stronger sequences.

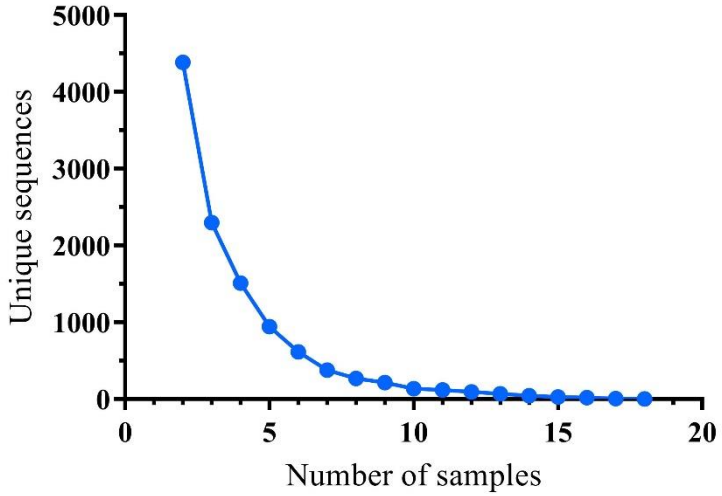


Fig.4.7.2. Most NGPD-derived sequences occurred in less than five samples. The number of samples a single sequence occurred in were counted. This was plotted against the number of sequences that occurred in *x* samples. For example, 945 sequences each occurred in five different samples.

Sequences were grouped according to which variant of ubiquitin was targeted. Groups were compared to highlight sequences that specifically bound to the Ser57 or Ser65 residue, and any sequences that only bound one variant were highlighted as likely to be variant-selective. As both S65E- and pS65-ubiquitin were used, an additional comparison to find sequences that occurred in both groups was conducted. This was to find sequences that showed strong, repeated binding to the phospho- or phosphomimetic-Ser65 variants. Four different comparisons were made;

- Sequences that bound S57E- but not WT-ubiquitin
- Sequences that bound S65E- or pS65- but not WT-ubiquitin
- Sequences that bound WT- but not S57E-, pS65- or S65E-ubiquitin
- Sequences that bound both S65E- and pS65- but not WT-ubiquitin.

This successfully identified 52 sequences that were considered to be high confidence, as they occurred in multiple samples at frequencies higher than would be expected through background

binding and were specific for their target ubiquitin. Of these, 12 sequences predominantly targeted WT-ubiquitin, 28 targeted S57E-ubiquitin and 12 targeted S65-ubiquitin.

This shortlist was then screened through biophysical prediction programmes and aligned to the bioinformatic UBD consensuses developed in Chapter 3. Whilst this application of the consensus resource could potentially identify viable phage peptide candidates that showed homology to short UBD-derived motifs, it more importantly acted as proof that the resource could be used to scan a large dataset for homology to UBD consensus motifs. Of the 52 sequences, four were found to show some homology to a known UBD domain, of which one sequence aligned to two separate UBDs (Table 4.7.2).

Table 4.7.2. Identifying homology within the NGPD dataset using the bioinformatic UBD resource.

Peptide		Target	Motif	UBD
03	ARWLWFTLLAFYKRAM	S57E	hxxA[ILM]	UIM
04	ATQAISSEEREIqRTW	S65E	hxxA[ILM]	UIM
09	DLMVRALTLAVEKqSL	S57E	hxxA[ILM]	UIM
18	HPqWCSLITCSRALLF	S65E	CxxxsCsxhxssxxxshCpxC@	ZnF_A20
18	HPqWCSLITCSRALLF	S65E	hxxA[ILM]	UIM

52 NGPD-derived peptides were identified according to frequency and comparative analysis. These were interrogated for homology to the UBD consensus motif resource developed in Chapter 3. Sequences with homology were included in the table with aligned amino acids underlined in both the peptide and consensus. The small q represented the read through from the amber stop codons. This table is adapted from Table 3.10.1.

The majority of the peptides aligned with UIM sequences. This was to be expected, both because the UIM is canonically about 12 amino acids long and so could be covered by the length of the 16mer, and because the UIM motif contained considerable room for alignment flexibility. Only three amino acids were defined, a hydrophobic residue, then two amino acids later an Ala followed by Ile, Val or Leu. This was not a very restrictive motif and so was more likely to align to the 16mers.

Whilst the comparison of the 52 phage-derived sequences to the UBD consensus resource did highlight some areas of interest, the peptides still had to pass biophysical checks before they could be considered for experimental validation as ubiquitin-binding peptides. Despite this, it was a successful test of the resource and proved its ease of use and speed compared to currently available methods.

Alignment to identify common structural and sequence motifs within the 52-peptide dataset, regardless of homology to known UBDs, failed to identify any common motif (data not shown). Hydrophobicity was estimated as the GRAVY score using the Sequence Manipulation Suite GRAVY Calculator (Stothard, 2000). Any sequence that was hydrophobic (positive GRAVY score) was discarded as it was likely to be insoluble and more difficult to handle than a hydrophilic sequence (negative GRAVY score). The hydrophobicity of each sequence was plotted against its average frequency across all containing samples (Fig.4.7.3). For ease of identification, each sequence was given a numerical identifier, 1-52, at random. All parameters were calculated on sequences with the Ala-Ser N-terminal adjunct and C-terminal amidation, to replicate the local conditions when presented on the phage surface.

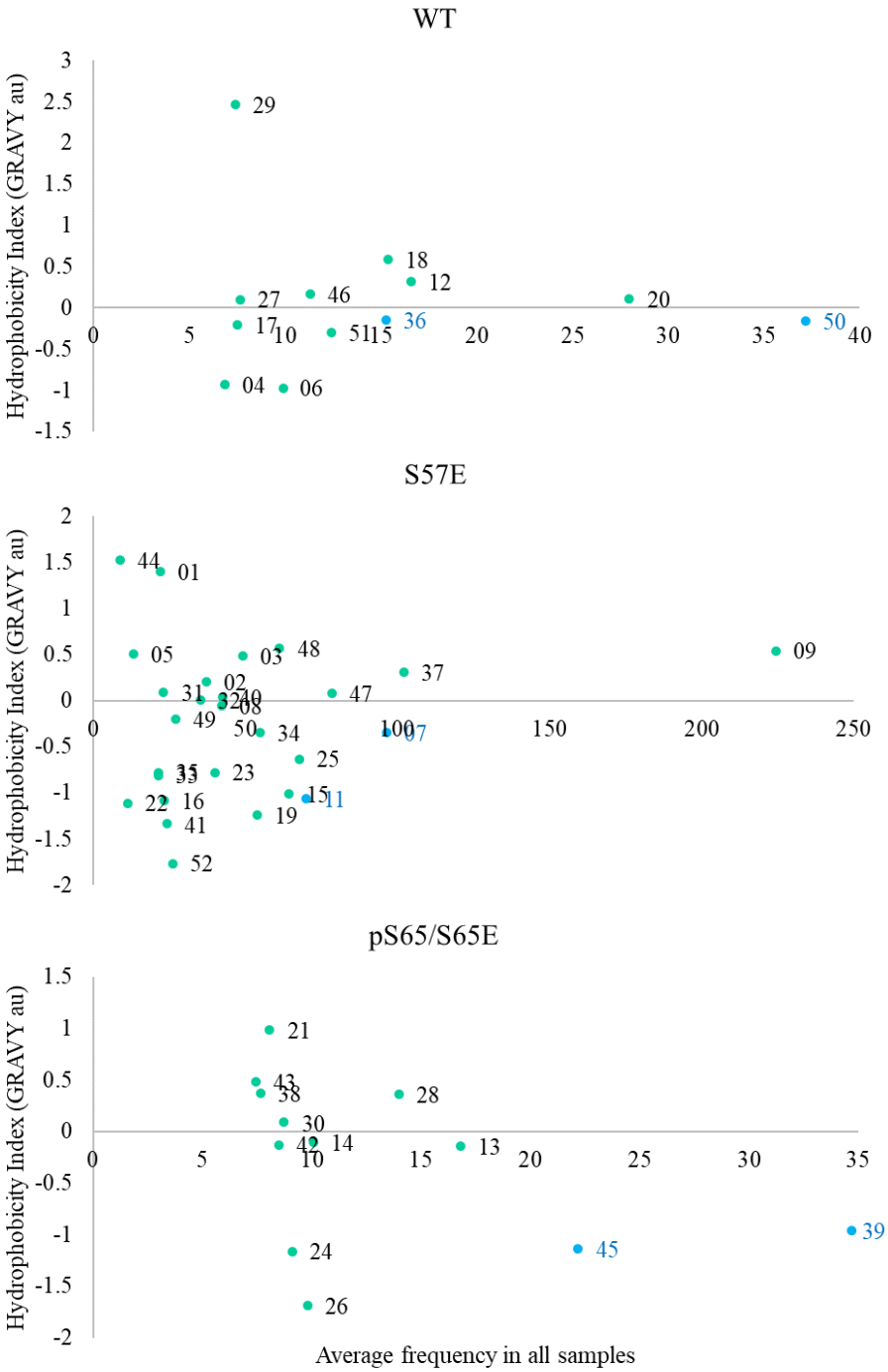


Fig.4.7.3. Candidate NGPD-derived peptides were selected according to frequency and hydrophobicity. Three sequential rounds of NGPD targeting WT-, S57E-, S65E- and pS65-ubiquitin were sequenced, and the binding peptides grouped by target. Sequences that met the threshold frequency criteria were compared to their hydrophobicity (green), and two candidates chosen per target as candidate UBDs for validation (blue). All sequences are shown with their Ala-Ser pIII leader adjunct.

From each group, the two hydrophilic peptides with the highest average frequency were chosen as candidates for validation. The hydrophilic nature of these peptides ensured that they would be usable within a laboratory environment, and the high average frequency suggested that the peptide was enriched during the biopanning as a high affinity binding peptide.

One additional candidate was included, peptide 10. This peptide targeted S57E-ubiquitin but is not shown on the graph as it was an outlier of exceptionally high frequency and fell beyond feasible representation on the axis. For reference, peptide 10 had an average frequency of 1,247 with a GRAVY score of -0.867.

The five peptides identified by applying the bioinformatic human UBD consensus resource developed in Chapter 3 (Table 4.7.2) were not included in the final list of candidates. This was due to their unsuitable biophysical properties and low frequency per sample, as indicated in the graphs above (Fig.4.7.3). Peptides 03, 09 and 18 were predicted to be hydrophobic and insoluble, and peptides 04 and 52 were of comparative low frequency.

This gave a total of seven NGPD-derived peptides, that occurred with a frequency ≥5 in at least five of the 36 different experimental samples, that were specific for their target ubiquitin and were useably hydrophilic for validation as candidate UBDs (Table 4.7.3).

Table 4.7.3. The seven NGPD-derived candidate UBD peptide sequences.

Peptide	Sequence	Target	Average Frequency	Hydrophobicity (GRAVY au)
07	ASCYFRYHRVCIRSWEVq	S57E	97	-0.350
10	ASEACSVqWRKTRVHRDV	S57E	1,247	-0.867
11	ASERRKPHGWTVIRCVRq	S57E	70	-1.072
36	ASRVSSRTAILTRNSAHV	WT	15	-0.150
39	ASSqWTRKQVqTSVMqSP	S65	35	-0.967
45	ASVHDAQWWqEMRLRANE	S65	22	-1.139
50	ASYLASPEALYRRLWTAP	WT	37	-0.161

After NGPD sequences were selected through frequency, reproducibility and hydrophobicity. Seven candidate sequences were chosen for validation and are displayed with their Ala-Ser pIII leader adjunct.

With no accepted protocol for NGPD experiments targeting immobilised ubiquitin being published at the time of experimentation, a new protocol was developed. The identification of sequences that occur at high frequency across several experiments suggests that this development was successful, and that this protocol could have use in other ubiquitin-targeting experiments, regardless of the success of the individual sequences as UBDs.

One great advantage of Phage Display was the very low bias, as the scientist cannot unconsciously skew the selection towards 'promising' sequences. Unfortunately, this also had the consequence that ability of the candidates to bind ubiquitin without the construct of Phage Display could not be determined until after candidates had been selected, produced, characterised and validated.

4.8 NMR VALIDATION OF NGPD-DERIVED UBD PEPTIDES

The seven candidate UBD peptides selected through ubiquitin-targeted NGPD and a positive control Rabex5-MIU peptide were synthesised by Genosphere Biotechnologies. Peptides were purchased with C-terminal amidation to mimic the Phage Display conditions and the N-terminal Ala-Ser adjunct representing the pIII leader tail. 2mg of each peptide was purchased at 70% purity, which was considered the minimum required for successful validation.

Cross-discipline validation was chosen to reduce any bias or experimental artefacts inherent in the NGPD protocol. Protein NMR spectroscopy was chosen above other potential validation methods as it is comparatively quick, taking just a couple of hours per peptide and is well-suited for the study of low affinity interaction. Compared to the common alternative of isothermal calorimetry (ITC) and surface plasmon resonance (SPR), NMR can detect a wider range of weaker affinities (K_d of 10^{-3} – 10^{-7} M for NMR compared to 10^{-5} – 10^{-8} M for ITC and 10^{-5} – 10^{-12} M for SPR) (Erba and Zenobi, 2011). Both fluorescence anisotropy and electrospray ionization (ESI-MS), whilst offering a good range of affinity detection with low

sample consumption and set up time, do not provide information on binding sites. Affinity purification by coupling of peptide to CNBr-Sepharose and incubating with ubiquitin similarly requires considerable amounts of peptide, both for coupling at 2mg/mL (a standard starting concentration) and to allow for repeat experiments during optimisation which can quickly become expensive.

By NMR it is possible to view the microenvironment of atomic nuclei. For proteins, ^{15}N HSQC allows the observation of the amide backbone resulting in one peak per residue (although in certain circumstances there may be more). Proline and the most N-terminal residue are undetectable due to the lack of amide protons and the presence of the -NH2 moiety.

To validate the seven NGPD-derived candidate peptides as artificial ubiquitin-binding sequences, each peptide was incubated at an 8:1 molar ratio excess with ^{15}N-ubiquitin (400μM:50μM) at 281K whilst collecting 2D ^{1}H-^{15}N HSQC spectra. Each spectrum was compared to that of ^{15}N-ubiquitin in the absence of candidate peptide and any chemical shift noted. As a positive control, a peptide of covering the confirmed MIU UBD from human Rabex-5 (total of amino acids 48-74) was purchased from the same supplier at >95% purity and spectra measured under the same conditions with ^{15}N-ubiquitin (Fig.4.8.1).

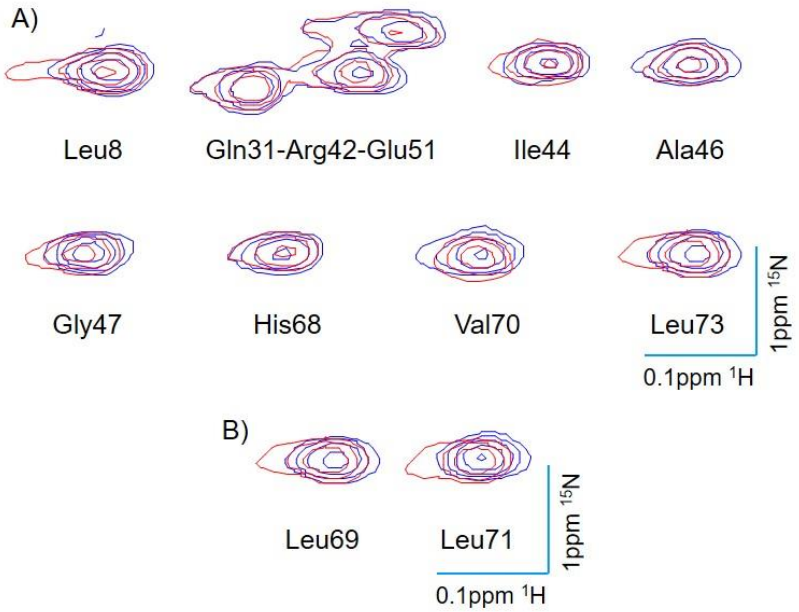


Fig.4.8.1. Peaks of interest from overlaid ^{1}H-^{15}N HSQC NMR spectra for ^{15}N-ubiquitin in the presence and absence of an MIU. Spectra for 50µM ^{15}N-ubiquitin sampled at 281K on a Bruker Ascend^{TM} 800 MHz spectrometer in the absence (blue) and presence (red) of 135µM Rabex-5-MIU. A) The residues of the published binding site are shown (Penengo *et al.*, 2006) with two extra overlapping peaks. B) Two additional peaks showed notable shift. Data was captured as 2D ^{1}H-^{15}N HSQC NMR in 10% D_2O pH 7.5. The reference point was the water line. For the full spectra, see Appendix A4.2.3.

The MIU had previously been shown to bind the Ile44 patch of ubiquitin across Leu8, Arg42, Ile44, Ala46, Gly47, His68, Val70 and Leu73 (Penengo *et al.*, 2006). However considerably less shift was seen than was expected. Minor shifts were seen across Leu8, Gly47, Val70 and Leu73 as well as Leu69 and Leu71 consistent with a small interaction between ^{15}N-ubiquitin and Rabex-5-MIU. That the peaks had begun to drift without beginning to properly separate suggests that there was a slow exchange interaction occurring. Unfortunately, there was not sufficient peptide to permit a repeat experiment with higher molar ratios between the MIU and ^{15}N-ubiquitin.

Given the lower than anticipated detection of MIU interaction with ^{15}N-ubiquitin, candidate peptides were run at a high concentration (400µM) and at the same low temperature (281K) to encourage ubiquitin binding. All seven candidate NGPD-derived peptides were measured under the same 2D HSQC NMR conditions in the presence of ^{15}N-ubiquitin. Sample temperature, concentration and buffer were maintained throughout. Each spectrum was

compared to the ^{15}N-ubiquitin spectrum in the absence of candidate peptide and any chemical shift noted. As all seven candidate peptides showed similar levels and patterns of peak shift, only Peptide 45 is shown as a representative spectrum (Fig.4.8.2).

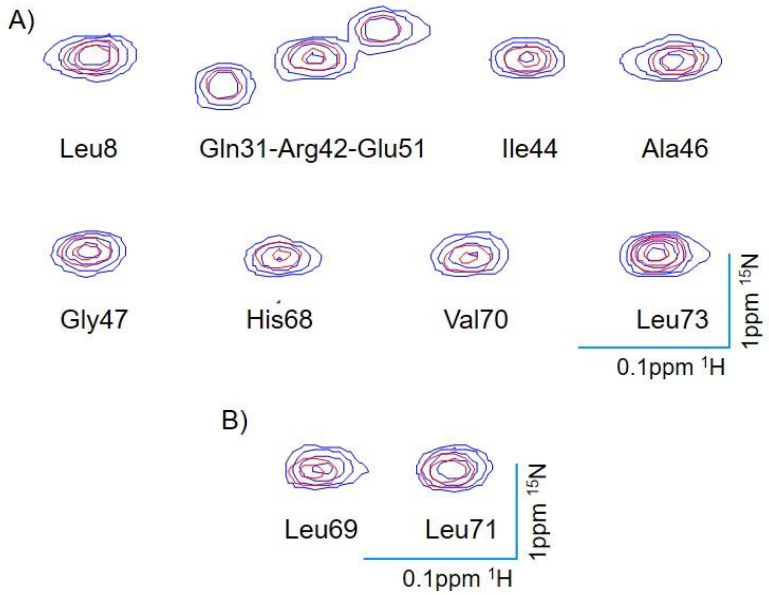


Fig.4.8.2. *Peaks of interest from overlaid ^{1}H-^{15}N HSQC NMR spectra for ^{15}N-ubiquitin in the presence and absence of Peptide 45.* Spectra for 50µM ^{15}N-ubiquitin sampled at 281K on a Bruker Ascend^{TM} 800 MHz spectrometer in the absence (blue) and presence (red) of 400µM Peptide 45. A) The residues of the Rabex-5 MIU published binding site are shown (Penengo *et al.*, 2006) with two extra overlapping peaks. B) Two additional peaks that showed notable shift for the MIU. Data was captured as 2D ^{1}H-^{15}N HSQC NMR in 10% D_2O pH 7.5. The reference point was the water line. For the full spectra, see Appendix Fig.A4.2.4.

The only 15N-ubiquitin residue to suggest a shift when in the presence of Peptide 45 was Leu69. A single minor shift was insufficient to confirm interaction. None of the other residues, either those show in the figure above or those in the full spectra (Appendix Fig.A4.2.4) demonstrated any notable shift. A sample of the peptide was subjected to circular dichroism (CD) and suggested a slight helical structure (data not shown).

These NMR and CD results were indicative of all seven candidate peptides and suggested that none of the peptides interacted with ^{15}N-ubiquitin as detected by NMR. When considering this with the ^{15}N-ubiquitin + MIU trace in Figure 4.8.1, it seems likely that there were not

sufficiently high proportions of full-length peptide present in the source material, as the MIU overlay should have yielded more significant differences based on previous (unpublished) data.

With little to no interaction seen between ubiquitin and the seven candidate peptides the NGPD-derived sequences could not be reasonably validated as novel UBDs.

4.9 CONCLUSIONS

To test the hypothesis that Next-Generation Phage Display could be redesigned to yield novel artificial UBDs, a naïve 16mer randomised peptide library was displayed on the surface pIII of the M13 phage and panned for three sequential rounds against immobilised WT-, S57E-, S65E- and pS65-ubiquitin.

At the time of submitting this thesis (September 2019), Phage Display had never been targeted towards ubiquitin and the method required considerable development. Several aspects of the method were adjusted to suit the low-affinity ubiquitin target. Previous uses of NGPD with the ubiquitin system involved creating a library of ubiquitin variants and panning against immobilised UBDs or ubiquitin-binding proteins. The protocol developed here reversed the interaction, with a library of potential ubiquitin-binding sequences targeting immobilised ubiquitin. The number of specific and frequently occurring sequences resulting from the NGS confirmed the success of the protocol development.

Seven peptides were chosen as candidate artificial ubiquitin-binding peptides. The selection process involved several approaches combined to complement the Phage Display experimental design. The aim was to select candidates that were high frequency and occurred in multiple samples – the repeated identification of a sequence across samples gave confidence that it was specific binding.

Cross-disciplinary methods were consciously chosen for candidate development and validation. This was to remove the chance of experimental artefact carrying through and maintaining false positives. By switching between solid-phase and solution-phase experiments any false positives created by the artificially high concentration of immobilised ubiquitin were removed. Unfortunately, all seven candidates were found to be false positives, as validation through NMR was not possible.

In summary, whilst development of the NGPD protocol was successful, validation was not, possibly due to the change from ubiquitin immobilised on a plate to in solution or possibly due to incorrect selection of candidates. With no novel ubiquitin-binding peptides resulting from this method an alternate approach was required to identify novel UBDs. Whilst the high throughput capability of NGPD is a significant advantage, the system is perhaps too far removed from endogenous conditions and too strictly controlled to allow use in multiple methods. A more physiologically relevant approach, whilst lower throughput, might yield more successful candidates.

5 PROTEOMIC IDENTIFICATION OF UBIQUITIN-BINDING DOMAINS AND UBIQUITIN-BINDING PROTEINS

5.1 INTRODUCTION

The power of approaches such as NGPD (Chapter 4) lies in the ability to screen high numbers of potential binding sequences with very low bias. However, validating the candidate peptide ubiquitin-binding sequences proved challenging. Indeed, none of the seven best candidate phage peptides could be functionally validated for ubiquitin-binding after moving from the solid-phase screening protocol to the solution-phase NMR. Whether these candidates were false positives, potentially occurring through library parasitism, or functional with a requirement for solid-phase immobilisation was unclear. A more consistent, physiologically relevant approach to identify ubiquitin-binding proteins was therefore required, one that was already closer in design to the endogenous environment.

Classical proteomic approaches to identifying binding partners are generally lower throughput but less likely to yield false positives. Previous experiments in the Layfield laboratory have used Sepharose affinity chromatography in conjunction with LC-MS/MS to investigate ubiquitin interactions (for examples, see Scott *et al.*, 2016; Goode *et al.*, 2016; Wright *et al.*, 2013). However historically such experiments have focused on the changes observed between wild-type and modified or mutated ubiquitin or on how changing conditions affected ubiquitin binding to known ubiquitin-binding proteins. This led to the hypothesis that novel ubiquitin-binding partners could exist in the background of such experiments, so far ignored in favour of studying those with known physiological relevance. Many ubiquitin-binding proteins could have been overlooked in previous assays due to their low abundance within a sample, or due to the focus of the experiment being on more conventional proteins that fit an existing scientific narrative. The latter obstructs the discovery of truly novel ubiquitin-binding proteins and UBDs, potentially discarding interesting ubiquitin-binding proteins or UBDs with the

"background" data. An affinity chromatography and LC-MS/MS identification of mammalian proteins biochemically binding ubiquitin could yield novel binding partners, which would either represent new examples of ubiquitin-binding proteins or, if no known UBD is present within the binding partner sequences, potentially novel UBDs. In order to address this, the aims for this chapter are;

1. To experimentally determine proteins that interact non-covalently with ubiquitin from mammalian tissue,
2. To identify common features in observed ubiquitin-binders,
3. To prioritise candidates for novel ubiquitin-binding proteins or UBDs,
4. To validate candidates through independent analyses,
5. And to begin to characterise molecular details of the interaction between candidate and ubiquitin.

To achieve these aims an affinity chromatography purification was conducted with no buffer- or condition-bias towards any known ubiquitin-binding protein and the resulting data searched for potentially novel binding proteins. Gene Ontology (GO) term analysis was used to identify enrichment of any biological functions common to the proteins. One candidate ubiquitin-binding protein was prioritised, validated and biochemically characterised. This led to the discovery of a novel aspect of ubiquitin structure and recognition.

5.2 AFFINITY CHROMATOGRAPHY AND LC-MS/MS IDENTIFICATION OF MAMMALIAN UBIQUITIN-BINDING PROTEINS

To identify previously overlooked ubiquitin-binding proteins, ubiquitin-Sepharose was used to affinity purify proteins from pig brain lysate. Pig brain was chosen as a readily available tissue with a diverse proteome with high homology to that of human tissue, which would be preferable but considerably more difficult to obtain. Briefly, pig brain tissue was lysed in

capture buffer (Appendix A.2.1), and proteins captured on ubiquitin- or control-Sepharose for 2hrs with end-over-end rotation at room temperature. Bound proteins were eluted, separated by SDS-PAGE gel, the complete mixture identified by LC-MS/MS at the Cambridge Centre for Proteomics Core Facility (University of Cambridge) and the resulting pig proteins mapped to their human homologues.

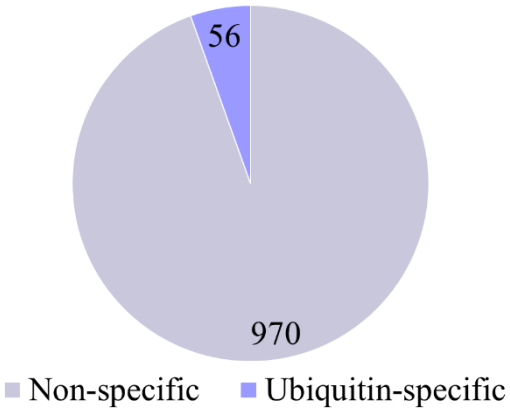


Fig.5.2.1. Identification of ubiquitin-specific human binding proteins from LC-MS/MS data. 6.7mg of total protein from pig brain lysate was captured on 140µl Sepharose coupled to human ubiquitin at 2mg/ml for 2hrs with end-over-end rotation at room temperature. Captured proteins were eluted and 50% of the eluate resolved by SDS-PAGE gel then identified by LC-MS/MS. 1,026 proteins could be mapped to unique human homologues, of which 970 were non-specific also binding to control-Sepharose (grey) and 56 specifically bound ubiquitin-Sepharose (blue). An identification of one peptide per protein threshold was applied to retain even low affinity or low occurrence proteins that may have been missed by previous experiments. The number of times a given peptide was identified within the sample was indicated by the total spectrum count. Data represent proteins from a single experiment.

Table 5.2.1. Ubiquitin-specific binding proteins identified from pig brain lysate.

Total spectrum count	UniProt	Gene	Protein	
13	Q7Z569	BRAP	BRAP	BRCA1-associated protein
11	Q9UBN7	HDAC6	HDAC6	Histone deacetylase 6
10	Q92995	USP13	UBP13	Ubiquitin carboxyl-terminal hydrolase 13
6	Q9Y5K5	UCHL5	UCHL5	Ubiquitin carboxyl-terminal hydrolase isozyme L5
6	Q9UBV8	PEF1	PEF1	Peflin
4	P46777	RPL5	RL5	60S ribosomal protein L5
4	P51665	PSMD7	PSMD7	26S proteasome non-ATPase regulatory subunit 7
4	O75663	TIPRL	TIPRL	TIP41-like protein

3	Q96DC9	OTUB2	OTUB2	Ubiquitin thioesterase OTUB2
3	Q9Y5Z4	HEBP2	HEBP2	Heme-binding protein 2 (placental)
3	P49748	ACADVL	ACADV	Very long-chain specific acyl-CoA dehydrogenase, mitochondrial
2	P11586	MTHFD1	C1TC	C-1-tetrahydrofolate synthase, cytoplasmic
2	P61086	UBE2K	UBE2K	Ubiquitin-conjugating enzyme E2 K
2	O14562	UBFD1	UBFD1	Ubiquitin domain-containing protein UBFD1
2	O00560	SDCBP	SDCB1	Syntenin-1
2	Q16401	PSMD5	PSMD5	26S proteasome non-ATPase regulatory subunit 5
2	Q8TBF2	PRXL2B	PXL2B	Prostamide/prostaglandin F synthase
2	Q92696	RABGGTA	PGTA	Geranylgeranyl transferase type-2 subunit alpha
2	Q96GG9	DCUN1D1	DCNL1	DCN1-like protein 1
2	A4D126	CRPPA	ISPD	D-ribitol-5-phosphate cytidylyltransferase
2	O75822	EIF3J	EIF3J	Eukaryotic translation initiation factor 3 subunit J
2	Q9UJ41	RABGEF1	RABX5	Rab5 GDP/GTP exchange factor
2	P40429	RPL13A	RL13A	60S ribosomal protein L13a
2	Q9P2E9	RRBP1	RRBP1	Ribosome-binding protein 1
2	P53007	SLC25A1	TXTP	Tricarboxylate transport protein, mitochondrial
2	Q9BXS5	AP1M1	AP1M1	AP-1 complex subunit mu-1
2	Q96P48	ARAP1	ARAP1	Arf-GAP with Rho-GAP domain, ANK repeat and PH domain-containing protein 1
1	Q9UBQ0	VPS29	VPS29	Vacuolar protein sorting-associated protein 29
1	P26639	TARS	SYTC	Threonine--tRNA ligase, cytoplasmic
1	P52788	SMS	SPSY	Spermine synthase
1	O00232	PSMD12	PSD12	26S proteasome non-ATPase regulatory subunit 12
1	P27824	CANX	CALX	Calnexin
1	Q9UDY4	DNAJB4	DNJB4	DnaJ homolog subfamily B member 4
1	P53621	COPA	COPA	Coatomer subunit alpha
1	Q08211	DHX9	DHX9	ATP-dependent RNA helicase A
1	P24588	AKAP5	AKAP5	A-kinase anchor protein 5
1	Q8N1G4	LRRC47	LRC47	Leucine-rich repeat-containing protein 47
1	P21695	GPD1	GPDA	Glycerol-3-phosphate dehydrogenase [NAD (+)], cytoplasmic
1	P26640	VARS	SYVC	Valine--tRNA ligase

1	O60763	USO1	USO1	General vesicular transport factor p115
1	Q86TU7	SETD3	SETD3	Actin-histidine N-methyltransferase
1	P30626	SRI	SORCN	Sorcin
1	O60784	TOM1	TOM1	Target of Myb protein 1
1	Q13501	SQSTM1	SQSTM	Sequestosome-1
1	P04632	CAPNS1	CPNS1	Calpain small subunit 1
1	Q6JQN1	ACAD10	ACD10	Acyl-CoA dehydrogenase family member 10
1	O15083	ERC2	ERC2	ERC protein 2
1	Q9NQW6	ANLN	ANLN	Anillin
1	Q13838	DDX39B	DX39B	Spliceosome RNA helicase DDX39B
1	Q9BS26	ERP44	ERP44	Endoplasmic reticulum resident protein 44
1	Q4G0N4	NADK2	NAKD2	NAD kinase 2, mitochondrial
1	Q14CZ8	HEPACAM	HECAM	Hepatocyte cell adhesion molecule
1	Q8NCW5	NAXE	NNRE	NAD(P)H-hydrate epimerase
1	P61758	VBP1	PFD3	Prefoldin subunit 3
1	P46782	RPS5	RS5	40S ribosomal protein S5
1	Q9P0L0	VAPA	VAPA	Vesicle-associated membrane protein-associated protein A

LC-MS/MS identification of specifically affinity chromatography purifying pig brain proteins to ubiquitin-Sepharose identified 56 unique proteins that could be mapped to human homologues. The number times a protein was identified within the sample is listed under 'Total Spectrum Count'. The human homologues are presented with UniProt identifier code, homologous gene name and the human protein abbreviation and full name.

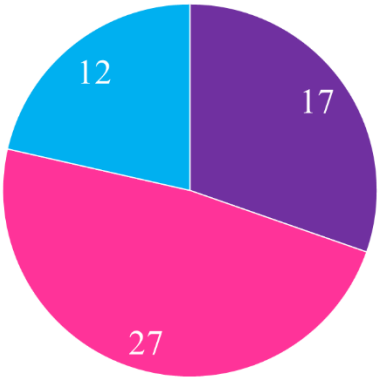
LC-MS/MS identified a total of 1,026 proteins with spectra matching those of known pig proteins (*Sus scrofa*) in the CCP_Suidae_genus_20190708 Mascot database (www.matrixscience.co.uk). These were mapped to their human homologues through gene and sequence homology for greater physiological relevance. Scaffold v4.8.9 (http://www.proteomesoftware.com/products/scaffold/) was used to visualise the "total spectrum count" with protein and peptide thresholds of 95% and a minimum peptide count threshold of one. The majority were non-specific as they were also retained on control-Sepharose, leaving 56 proteins that specifically bound ubiquitin- but not control-Sepharose (Fig.5.2.1, Table 5.2.1). Whilst this data represented a single affinity chromatography experiment, it was supported by work from other members of the Layfield group and all

conclusions were experimentally validated, including a repeat of this experiment, later in this chapter.

The resulting list was screened for previously confirmed ubiquitin interactions by cross referencing to established databases such as BioGRID and underwent analysis of GO term annotations to identify any significantly enriched GO terms in ubiquitin-binding proteins regardless of their current interest or physiological role.

5.3 TREND ANALYSIS OF AFFINITY CHROMATOGRAPHY PURIFIED UBIQUITIN-BINDING PROTEINS

The BioGRID 3.5 database (Oughtred *et al.*, 2019) is a manually curated repository of experimentally determined protein-protein interactions as identified by the literature, although care must be taken to assess the validity of each interaction. A combination of this and GO term annotations extracted from UniProt were used to assess if any of the 56 ubiquitin-binding proteins identified by LC-MS/MS had previously been linked to ubiquitin-binding.



Direct	Affiliated		Unknown
MTHFD1	VPS29	CAPNS1	PRXL2B
BRAP	TARS	ERC2	RABGGTA
UBE2K	SMS	ANLN	CRPPA
UBFD1	RPL5	DDX39B	HEBP2
USP13	CANX	ERP44	AKAP5
SDCBP	EIF3J	TIPRL	GPD1
UCHL5	DNAJB4	SLC25A1	RRBP1
PSMD12	COPA	AP1M1	SETD3
PSMD5	ACADVL	ARAP1	SRI
PSMD7	LRRC47	NAXE	ACAD10
DCUN1D1	RPL13A	VBP1	NADK2
OTUB2	PEF1	RPS5	HEPACAM
RABGEF1	VARS	VAPA	
DHX9	USO1		
TOM1			
SQSTM1			
HDAC6			

Fig.5.3.1. Novelty of ubiquitin-binding proteins identified from pig brain lysate. Proteomic identification of porcine brain ubiquitin-binding proteins identified 56 human homologues. These were assessed in the BioGRID 3.4 database and through ubiquitin-linked GO term annotations from PANTHER for previously identified interaction with ubiquitin. 17 proteins were confirmed as known direct interactors (purple), 27 were affiliated with ubiquitin-mediated processes or showed indirect or unconfirmed binding to ubiquitin (pink), and 12 had no known links (blue). The gene names are displayed for continuity within this section.

The majority of proteins were found to be affiliated with ubiquitin but had not been reported to bind directly. Affiliation was defined as involvement in a ubiquitin-mediated processes such as autophagy, or as a direct interaction with a known ubiquitin-binding protein like the E3 ligase Parkin or the autophagy receptor optineurin. Any instances where there was unreliability of the source data or inability to validate through other sources were ignored. This had

particular relevance to one protein, peflin, which the databases listed as a direct binding partner of ubiquitin, but closer examination of the sources (Matsumoto *et al.*, 2005; McGourty *et al.*, 2016) downgraded the interaction to indirect or unconfirmed.

Using the online esyN programme (Bean *et al.*, 2014) the 56 genes were presented as an interaction network according to the knowledge currently in the literature and BioGRID (Fig.5.3.2). Ubiquitin was represented by the UBB and UBC genes as it can be translated from both. Only human interactions were considered, with both high and low throughput physical interaction data. This allowed analysis of the broader interactome and visualisation of any possible protein networks or complexes. The Degree of each node was also calculated, which is defined as the number of edges connected to the node. A higher Degree is indicated by a larger node and represents a better connected, more central node which has a higher chance of forming multiprotein or ubiquitin-mediated complexes.

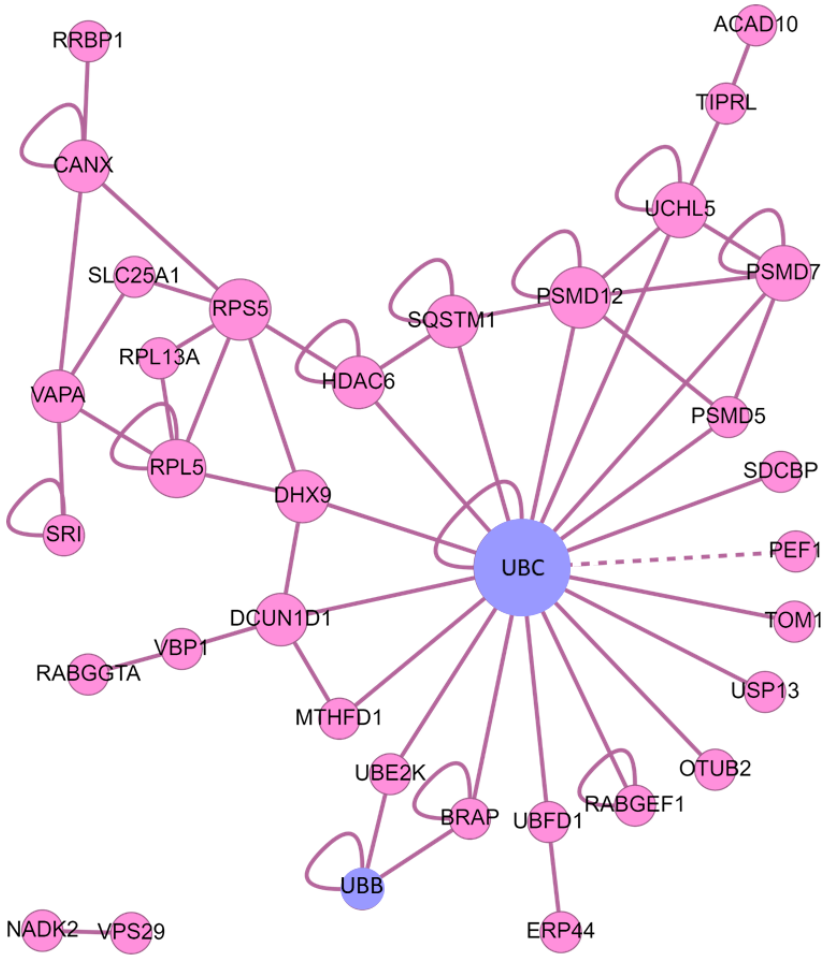


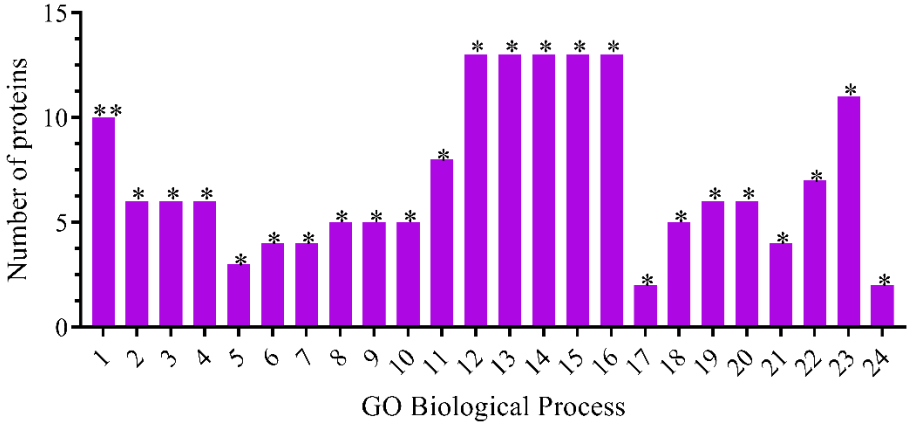
Fig.5.3.2. The interactome of ubiquitin-specific gene products identified in this study. The 56 human homologue genes identified from ubiquitin-Sepharose affinity chromatography of pig brain lysate were visualised as a network with both UBC and UBB ubiquitin-encoding genes (blue). Each node represents a gene, with its size being relative to the gene's average Degree. Each edge represents a known interaction, with loops indicating identical protein binding, and dotted lines indicating an unconfirmed or debated interaction. Genes with no confirmed interactions within the network were omitted from the figure.

Based on current knowledge, 13 proteins could have affinity chromatography purified in indirect complexes with ubiquitin as there is no known direct interaction. These are represented by the genes *ERP34*, *TIPRL*, *ACAD10*, *RPS5*, *RPL5*, *RPL13A*, *SLC25A1*, *VAPA*, *SRI*, *CANX*, *RRBP1*, *VBP1* and *RABGTA*. Several ubiquitin-mediated multiprotein complexes were possible, ranging between four and 22 proteins per complex. This does not preclude a yet undiscovered direct interaction.

The BioGRID and GO term analysis identified 27 proteins as affiliates of ubiquitin (Fig.5.3.1). As the network in Fig.5.3.2 shows, nine of these were known to bind ubiquitin indirectly (represented by genes *TIPRL, CANX, VAPA, SLC25A1, RPL13A, RPS5, RPL5, VBP1* and *ERP44*). 16 had no known interaction partners within the 56 proteins and were not presented here.

The remaining affiliated protein, peflin (*PEF1*), was of particular interest. The data presented by BioGRID, both in the manual searches (Fig.5.3.1) and the algorithm underpinning the network analysis (Fig.5.3.2), listed a direct ubiquitin-peflin interaction, however, closer analysis of the source papers revealed that direct interaction was unconfirmed in all published data (Matsumoto *et al.*, 2005; McGourty *et al.*, 2016). To indicate this, peflin was included with the proteins that are currently thought to interact with ubiquitin in an indirect manner only and was presented in the affiliated section of the pie chart and with a dotted line in the interactome.

The unbiased affinity chromatography to identify ubiquitin-binding proteins achieved 69% unconfirmed ubiquitin-binding proteins (novel and affiliated), with 31% already being known ubiquitin-binding proteins and a low probability of proteins co-precipitating due to the presence of protein complexes. As the trend towards novelty had already been determined, gene and total spectrum count data for the 56 ubiquitin-specific proteins were submitted to PANTHER14.1 Tools (Mi *et al.*, 2019; Thomas *et al.*, 2006) for GO term annotation and enrichment analysis (Fig.5.3.3).



1 protein modification by small protein conjugation or removal (GO:0070647)
2 regulation of proteolysis (GO:0030162)
3 negative regulation of cellular protein metabolic process (GO:0032269)
4 negative regulation of protein metabolic process (GO:0051248)
5 regulation of protein stability (GO:0031647)
6 anatomical structure morphogenesis (GO:0009653)
7 negative regulation of proteolysis (GO:0045861)
8 regulation of protein catabolic process (GO:0042176)
9 regulation of proteolysis involved in cellular protein catabolic process (GO:1903050)
10 regulation of cellular protein catabolic process (GO:1903362)
11 regulation of protein modification process (GO:0031399)
12 cellular protein modification process (GO:0006464)
13 macromolecule modification (GO:0043412)
14 protein modification process (GO:0036211)
15 regulation of cellular protein metabolic process (GO:0032268)
16 regulation of protein metabolic process (GO:0051246)
17 response to inorganic substance (GO:0010035)
18 negative regulation of catalytic activity (GO:0043086)
19 protein deubiquitination (GO:0016579)
20 protein modification by small protein removal (GO:0070646)
21 regulation of proteasomal protein catabolic process (GO:0061136)
22 protein modification by small protein conjugation (GO:0032446)
23 cellular response to stress (GO:0033554)
24 protein stabilization (GO:0050821)

Fig.5.3.3. Enrichment of GO Biological Process annotations in human homologues specifically binding ubiquitin from pig brain lysate. Only terms with significant enrichment compared to the human proteome are shown (**P,0.005, *p<0.05). Terms are arranged by significance, with most significant at position 1 and least significant at position 24.

Using the PANTHER v14.1 Gene List Analysis tool, the genes encoding the 56 ubiquitin-specific proteins were annotated with their associated biological processes. In total there were 1,689 processes represented by the 56 genes, minus generics such as 'cellular process'. As calculated by PANTHER, 24 processes were considered significantly over enriched compared to the human proteome ($p<0.05$). Of the 24 processes, 22 could be feasibly linked to ubiquitin or ubiquitin-mediated processes. The processes unrelated to ubiquitin were "#6 anatomical structure morphogenesis" and "#17 response to inorganic substance" (Fig. 5.3.3). The top 20 by total spectrum count all related to metabolic processes, response to stimuli or cell component organisation, however none of these showed significant changes in enrichment.

Trends in GO molecular function and protein class were also determined, using the same PANTHER specifications. There were 358 molecular functions in total, of which seven showed significant enrichment. However, these could all feasibly be related to ubiquitin-mediated processes (data not shown).

Including 'Unclassified' there were 36 protein classifications represented within the data. Only three showed significant enrichment. It is likely that they all refer to similar sets of proteins, as protease and hydrolase activities are often multiple annotations of a single process but these could all feasibly be linked to ubiquitin-mediated processes (data not shown).

Overall the unbiased affinity chromatography of pig brain lysate to ubiquitin-Sepharose identified proteins that were linked to ubiquitin-mediated processes but not always confirmed direct interactors of ubiquitin. This dataset includes several proteins with no known links to ubiquitin or ubiquitin mediated processes, suggesting that the attempt at an unbiased ubiquitin-binding protein selection was successful. One of the identified proteins, peflin, has already been highlighted within this chapter as being of interest. It has two published instances of interacting with known ubiquitin-binding proteins but no known instances of direct ubiquitin

binding. As confirmed by the interactome analysis it was unlikely that peflin precipitated as part of a ubiquitin-mediated complex, suggesting that this was a novel interaction.

5.4 IDENTIFYING PEFLIN AS A CANDIDATE NOVEL UBIQUITIN-BINDING PROTEIN

Peflin had previously appeared in several affinity chromatography and LC-MS/MS experiments in the Layfield lab under various buffer and experimental conditions and has been linked to multi-protein complexes with ubiquitin ligases in two independent papers (McGourty *et al.*, 2016; Kikuma *et al.*, 2019). To confirm that it was a novel ubiquitin-binding protein, detection of peflin within the affinity chromatography sample was first confirmed by alternate methods, and peflin produced and purified for use in interaction and characterisation experiments.

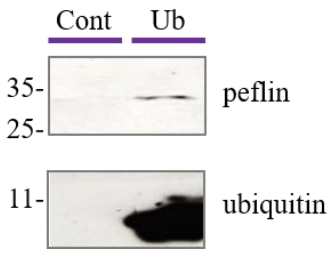


Fig.5.4.1. Peflin was detectable in affinity chromatography samples by Western blot. 6.7mg of total protein from pig brain lysate was captured on Sepharose coupled to human ubiquitin at 2mg/ml (Ub) or uncoupled control-Sepharose (Cont) for 2hrs with end-over-end rotation at room temperature. Captured proteins were eluted, and 50% of the eluate separated by SDS-PAGE gel then confirmed by Western blot with anti-peflin and anti-ubiquitin. Figure represents *n*=1.

Peflin interaction with ubiquitin-Sepharose from similar endogenous pull downs was confirmed by the alternative method of Western blot (Fig.5.4.1). It was only just detectable, however, suggesting that the interaction is easily out-competed by alternative binders, or that Western blot is less sensitive than LC-MS/MS.

Comparatively little is known about peflin in the literature; it is a penta-EF hand protein and is 284 amino acids long with a mass of 30-34 kDa, depending on PTMs (Kitaura *et al.*, 1999). It can heterodimerise with Apoptosis-Linked Gene 2 protein (ALG-2, also known as

Programmed Cell Death Protein 6, PDCD6) (McGourty *et al.*, 2016), with which it shares 41% identity and a general domain structure (Kitaura *et al.*, 1999), although the peflin structure has not been solved. Peflin contains no known UBD, according to the UniProt database entry (UniProt Consortium, 2018) and the UBD consensus resource developed earlier in this thesis (Section 3.10); if the peflin-ubiquitin interaction is shown to be direct then it would be a source of a novel endogenous UBD.

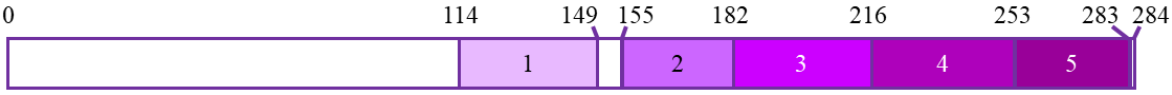


Fig.5.4.2. The basic domain structure of peflin. As a member of the penta-EF-hand family, peflin contains five EF hand domains, marked as #1-5 with the domain boundaries indicated. Domain annotations and boundaries were taken from UniProt and displayed running left to right from N- to C-terminus.

As an EF-hand-containing protein peflin can bind calcium and is involved in BCR(KLHL12) (BTB domain-Cullin 3-E3 ubiquitin-protein ligase-RING-box protein 1 and Kelch-like protein 12, respectively) complex-mediated mono-ubiquitination of sec31 in neural crest specification. It is also involved in COPII coat size regulation in transport between the endoplasmic reticulum and Golgi body (McGourty *et al.*, 2016). Peflin's exact role in these processes is yet to be determined, as is the effect of calcium binding on its function, although calcium certainly plays a role in peflin-mediated regulation (Kitaura *et al.*, 2001). Given the high number of calcium-binding domains and the literature on calcium-mediated function, it seemed plausible that high calcium concentrations were needed to promote the interaction with ubiquitin.

The LC-MS/MS data presented so far only indicated co-precipitation with ubiquitin, which could be through direct or indirect interaction. The frequent detection of peflin in multiple LC-MS/MS datasets under varying conditions over several years in the Layfield laboratory (data not shown) suggests that this interaction was either direct or part of a very robust complex. In order to decipher the interaction, recombinant peflin was produced and purified for further study.

5.5 PRODUCTION OF PURIFIED PEFLIN

To confirm that peflin did not precipitate with ubiquitin as part of an indirect complex or due to E1/E2/E3 ligase-mediated covalent peflin modification with ubiquitin, purified peflin was produced to observe binding in the absence of any other cell components.

The peflin insert from a pOTB7_PEF plasmid was subcloned into the pGEX-4T2 vector with GST tag and ampicillin resistance. This was transformed into Rosetta *E. coli*, expressed under IPTG, purified by capture on glutathione-Sepharose and thrombin cleavage then run on SDS-PAGE for confirmation of peflin expression (Fig.5.5.1).

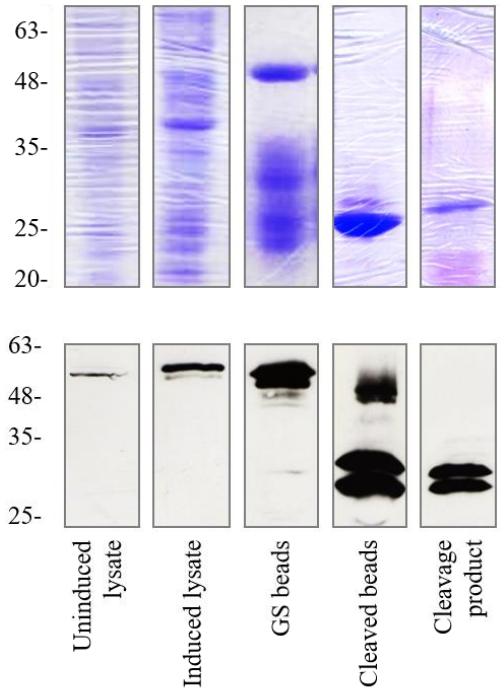


Fig.5.5.1. Purification of recombinant peflin. GST-peflin was purified from 25mL induced Rosetta *E. coli* by capture on 250µL glutathione-Sepharose (GS), and peflin cleaved from GST by thrombin. The top panel represents Coomassie staining to confirm purification and the bottom panel represents detection by Western blot with anti-peflin to confirm expression.

Excised Coomassie bands were also subjected to LC-MS/MS at the Cambridge Centre for Proteomics (University of Cambridge) and the spectra confirmed the detected band was peflin with good coverage and protein identification probability (37% and 100% as determined by the Centre) with minimum number of peptides threshold of two (Fig.5.5.2).



MASYPYRQGC PGAAGQAPGA PPGSYYPGPP NSGGQYGSGL
PPGGGYGGPA PGGPYGPPAG GGPYGHPNPG MFPSGTPGGP
YGGAAPGGPY GQPPPSSYGA QQPGLYGQGG APPNVDPEAY
SWFQSVDSDH SGYISMKELK QALVNCNWSS FNDETCLMMI
NMFDKTKSGR IDVYGFSALW KFIQQWKNLF QQYDRDRSGS
ISYTELQQAL SQMGYNLSPQ FTQLLVSRYC PRSANPAMQL
DRFIQVCTQL QVLTEAFREK DTAVQGNIRL SFEDFVTMTA
SRML

Fig.5.5.2. LC-MS/MS confirmation of peflin expression. GST-peflin was purified from 25mL induced Rosetta *E. coli* as described above and peflin cleaved from GST by thrombin. The cleavage product was subjected to LC-MS/MS at the Cambridge Centre for Proteomics. Spectra were visualised with Scaffold v4.8.9 with a minimum threshold value of two peptides. Peflin was identified with 37% coverage (purple). Coverage across both the whole protein (top) and the amino acid sequence (bottom) is shown.

The production of recombinant human peflin was reliable enough to permit further biochemical experiments with the aim of exploring the potential ubiquitin-peflin interaction.

5.6 PEFLIN AS A NOVEL CALCIUM-SENSITIVE UBIQUITIN-BINDING PROTEIN

To determine if the detection of peflin from pig brain lysate was due to direct or indirect interaction with ubiquitin, GST-peflin was purified from 25mL of bacterial culture to a yield of 50µg as determined by BCA; as bacteria lack a ubiquitination system this precluded interference from interaction with other ubiquitination-related proteins. This was incubated with ubiquitin-Sepharose in the presence and absence of additional calcium (Fig.5.6.1).

Many ubiquitin-binding proteins and UBDs bind zinc as a requirement for binding (Matthews and Sunde, 2002), and, as peflin is a calcium-binding protein and both calcium and zinc are inorganic divalent cations, it was possible that high calcium would promote efficient binding. Calcium chloride was included to a concentration of 250µM to mimic the high physiological concentrations found at the endoplasmic reticulum (ER) membrane (Varadi and Rutter, 2002), where peflin is thought to accumulate (Rayl *et al.*, 2016).

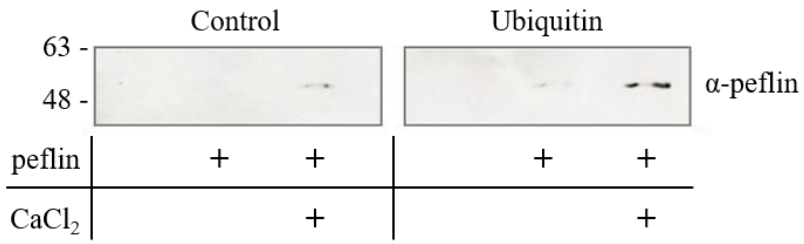


Fig.5.6.1. Purified recombinant human GST-peflin binds to ubiquitin-Sepharose. 50µg of recombinant human GST-peflin was incubated with 50µL ubiquitin-CNBr Sepharose (coupled at 2mg/mL) or control-CNBr Sepharose for 2hrs at room temperature with end-over-end rotation in capture buffer with or without addition of 250µM $CaCl_2$. Beads were washed, bound proteins eluted into GLB and 100% of the eluate analysed by SDS-PAGE and anti-peflin Western blot. Figure represents *n*=1.

Low level non-specific binding to control-Sepharose were seen, but as this was considerably lower than binding to ubiquitin-Sepharose it was not considered obstructive. GST-peflin directly bound ubiquitin-Sepharose, confirming that the binding seen from pig brain lysate pull downs was not indirect via a multiprotein complex.

In contrast to GST-tagged protein, thrombin-cleaved recombinant PEF1 was found to show variable ubiquitin-Sepharose binding and response to calcium, and for clarity these blots are not shown. Whilst peflin binding was visible in all ubiquitin samples, in the presence of calcium binding was near abolished in the thrombin cleaved untagged condition (not shown) and clearly potentiated in the GST-tagged condition. The effects of calcium remained unclear and required further comparison.

It should be noted that the initial peflin purification displayed as two discrete bands (Fig.5.5.1). The lower band had the expected molecular weight for peflin (30kDa), was confirmed as peflin through LC-MS/MS and was consistently produced through all purifications and experiments shown in this thesis. The upper band was only seen in these initial experiments and had a molecular weight greater than that of peflin. The growth conditions were identical between all samples and all cultures were taken from the same single glycerol stock, so it is unlikely that there was a change in peflin expression. Instead it is likely that this additional band resulted from off-target thrombin activity during peflin purification, as the band was not seen after

purchasing a new batch of thrombin. As the enhanced interaction was seen with the single peflin product in multiple subsequent experiments with various purification systems and experimental designs it was decided that the initial cleaved product contained some purification artefact that had obstructed binding. The wealth of positive interaction data collected after cleavage had been resolved confirmed the interaction regardless of the initial discrepancy.

The presence of E3 ligases within pig brain lysate and the high concentrations of protein in the purified samples prevent confirmation of peflin as a ubiquitin-binding protein; the interaction could be due to ubiquitination of peflin or artificially forced by high concentrations. To address this, GST-peflin was over-expressed in bacterial culture and captured from the lysate with ubiquitin-Sepharose. It is possible that addition of any divalent cation could potentiate binding, not just calcium specifically. To clarify this, additional controls were conducted, including a GST-p62, a known ubiquitin-binding protein as positive control, and also capture in the presence of $MgCl_2$ as an alternative divalent cation (Fig.5.6.2).

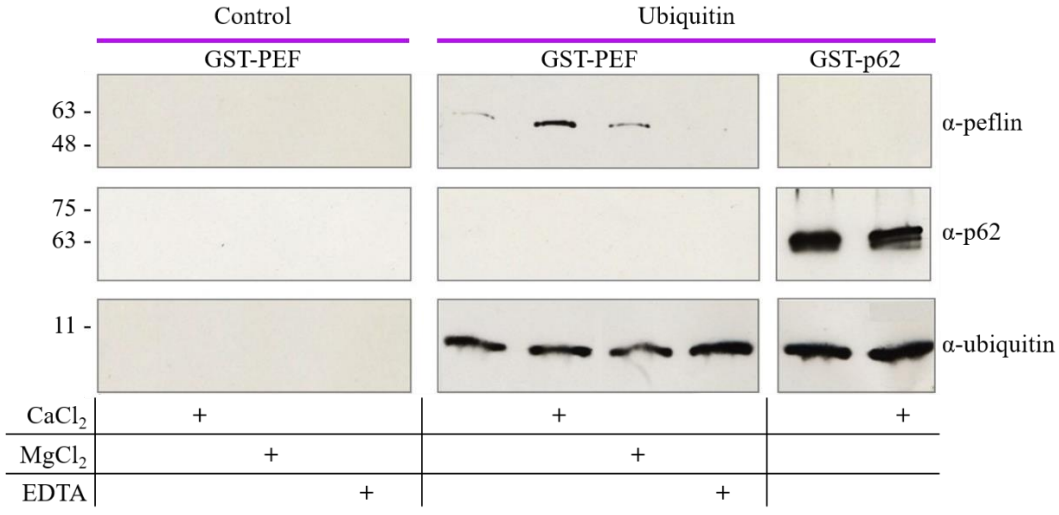


Fig.5.6.2. Overexpressed GST-peflin specifically binds ubiquitin-Sepharose and is potentiated by calcium. 25mL of GST-PEF Rosetta or GST-p62 XL10 gold overexpression was lysed in capture buffer with or without 250µM $CaCl_2$, 250µM $MgCl_2$ or 1mM EDTA. The supernatant was captured by 35µL ubiquitin-CNBr Sepharose (coupled at 2mg/mL) or to control- CNBr Sepharose for 2hrs at room temperature with end-over-end rotation. Beads were washed, bound proteins eluted into GLB and 100% of the eluate analysed by SDS-PAGE and Western blot with anti-peflin, anti-p62 and anti-ubiquitin. Figure represents *n*=2.

As bacteria lack a ubiquitin system, successful capture of GST-peflin from overexpressed bacterial lysates precluded the possibility of an E3 ligase covalently conjugating ubiquitin to peflin, indicating that the peflin-ubiquitin interaction is similar to that of a UBD and ubiquitin. No binding from a lysate expressing GST only was seen (data not shown), indicating that the interaction was not mediated by the GST tag.

The peflin-ubiquitin interaction was enhanced by the presence of 250μM $CaCl_2$ which was consistent with peflin being a calcium-binding protein. The addition of calcium-chelating EDTA to remove endogenous calcium supported this, however EDTA can have other effects outside of chelation and so should be taken with a note of caution. The addition of equal concentrations of $MgCl_2$ had only a very minor effect on ubiquitin-peflin binding, indicating that the calcium-induced enhancement of binding was not due to the addition of any divalent cation. It was also not a generalised effect of calcium on ubiquitin-interactions as binding to GST-p62, which has no calcium-binding EF hands, was unaffected

Taken together, the data presented in this section suggested that peflin directly interacted with ubiquitin and was potentiated by the presence of physiologically-relevant calcium concentrations. As peflin contains no known UBD and the peflin-ubiquitin interaction has been shown to be direct and not mediated by E3 ligases, peflin may contain a yet undiscovered UBD.

5.7 CHARACTERISING THE PEFLIN-UBIQUITIN BINDING INTERFACE

UBDs most commonly bind across the canonical Ile44 hydrophobic patch on the surface of ubiquitin, which also encompasses other adjacent non-polar residues such as Leu8, Ala46 and Val70 (Sloper-Mould *et al.*, 2001; Hurley *et al.*, 2006; Hicke *et al.*, 2005). To support the possibility of peflin binding through a UBD, the ability of endogenous porcine peflin to bind ubiquitin with mutations in the canonical patches was assessed (Fig.5.7.1).

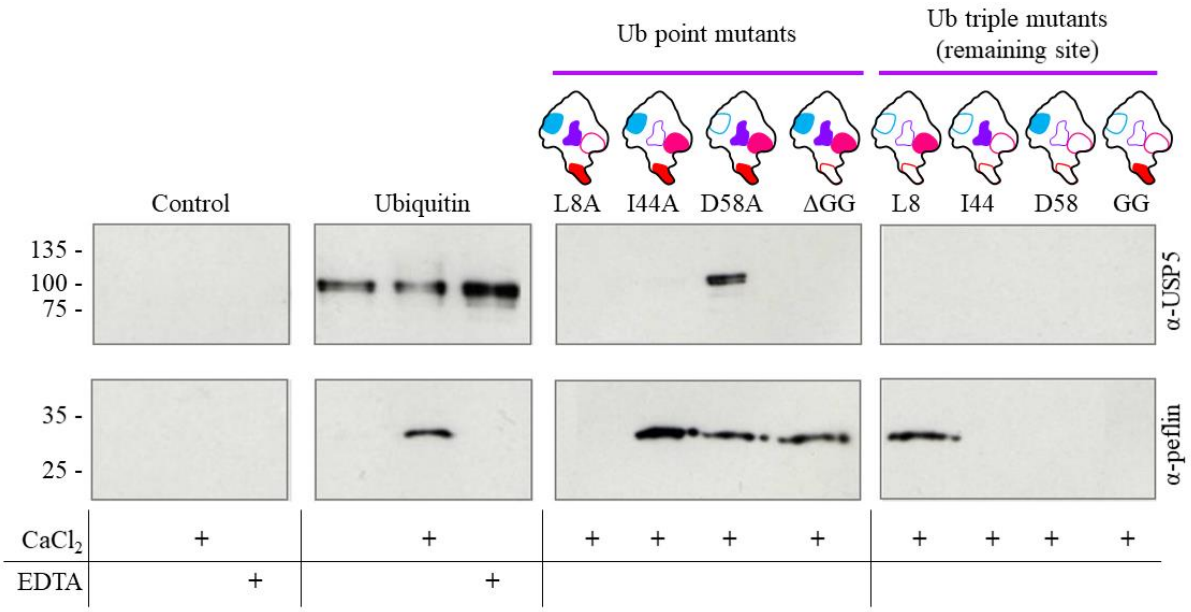


Fig.5.7.1. Peflin binding of ubiquitin-Sepharose is mediated by Leu8. 6.7mg of total protein from pig brain lysate was captured on 140µL Sepharose coupled to mutant human ubiquitin at 2mg/ml for 2hrs with end-over-end rotation at room temperature. Capture buffer was supplemented with 250µM $CaCl_2$ or 1mM EDTA. Captured proteins were eluted then 50% of the eluate separated by SDS-PAGE gel and detected by Western blot with anti-USP5 (positive control) or anti-peflin. Sites are coloured as; Leu8 pink, Ile44 purple, Asp58 blue and Gly76 red. Mutated sites are shown as outline only. Figure represents *n*=1.

Ubiquitin proteins with mutations across the canonical binding patches had previously been produced in-house and provided by Dr Jed Long (University of Nottingham). Proteins were either point mutated, replacing a single residue with Ala, or triple mutated, replacing three residues with Ala to leave a single site remaining. Alanine was chosen as the replacement amino acid as it lacks side-chain atoms beyond the β-carbon and so represents a sidechain truncation which is sterically readily accommodated. No anti-ubiquitin blot is shown as unpublished observations have frequently found that the antibody shows varying sensitivity to different ubiquitin mutants and is unreliable as a control. To help compensate for this, the positive control of USP5 is included. USP5 is known to bind ubiquitin across the residues in question and so acts as both a positive and loading control.

A very distinctive and clear binding profile was observed. Endogenous porcine peflin specifically bound ubiquitin-Sepharose in the presence of calcium but not EDTA,

corroborating evidence presented in this chapter. Interestingly the binding of peflin to the point mutated and triple mutated ubiquitin proteins were highly complementary; peflin bound to all point mutants except L8A and bound to none of the triple mutants except L8. This is a clear indication that peflin binds ubiquitin in a calcium-dependent manner *via* Leu8 of the canonical hydrophobic patch of ubiquitin and supports the presence of a peflin-UBD.

The figures presented here consist of only a small part of the Western blots conducted in this investigation. The blots not presented here included optimisations and small increments in knowledge that were neatly summarised by the handful of blots shown. However, whilst conducting these additional blots, peflin binding to ubiquitin-Sepharose in the absence of calcium showed some apparent variability. It was at the lowest boundary of detection by Western blot but consistently seen by the more sensitive LC-MS/MS. This suggested that the apparent variable binding simply reflected the normal day to day variation in Western blot sensitivity.

In an attempt to determine the site of the potentially novel peflin-UBD, structure simulations were predicted using the I-TASSER server (Iterative Threading ASSEmbly Refinement) (Yang and Zhang, 2015) (Fig.5.7.2) and the domain annotation extracted from UniProt (Fig.5.7.3). I-TASSER is an online platform for automatic structure prediction from a given amino acid sequence input. The sequence is first checked for existing or related structures in the Protein Data Base, and sequence and any partial structures combined into a tens of thousands of 'decoy' structures. Any steric clashes are removed, and C-score analysis is used to predict the most likely structures. The C-score is an indicator of confidence in the fidelity of the model and typically ranges from low confidence at -5 to high confidence at +2. The peflin structure prediction should be treated with caution, as the C-score was just -2.63.

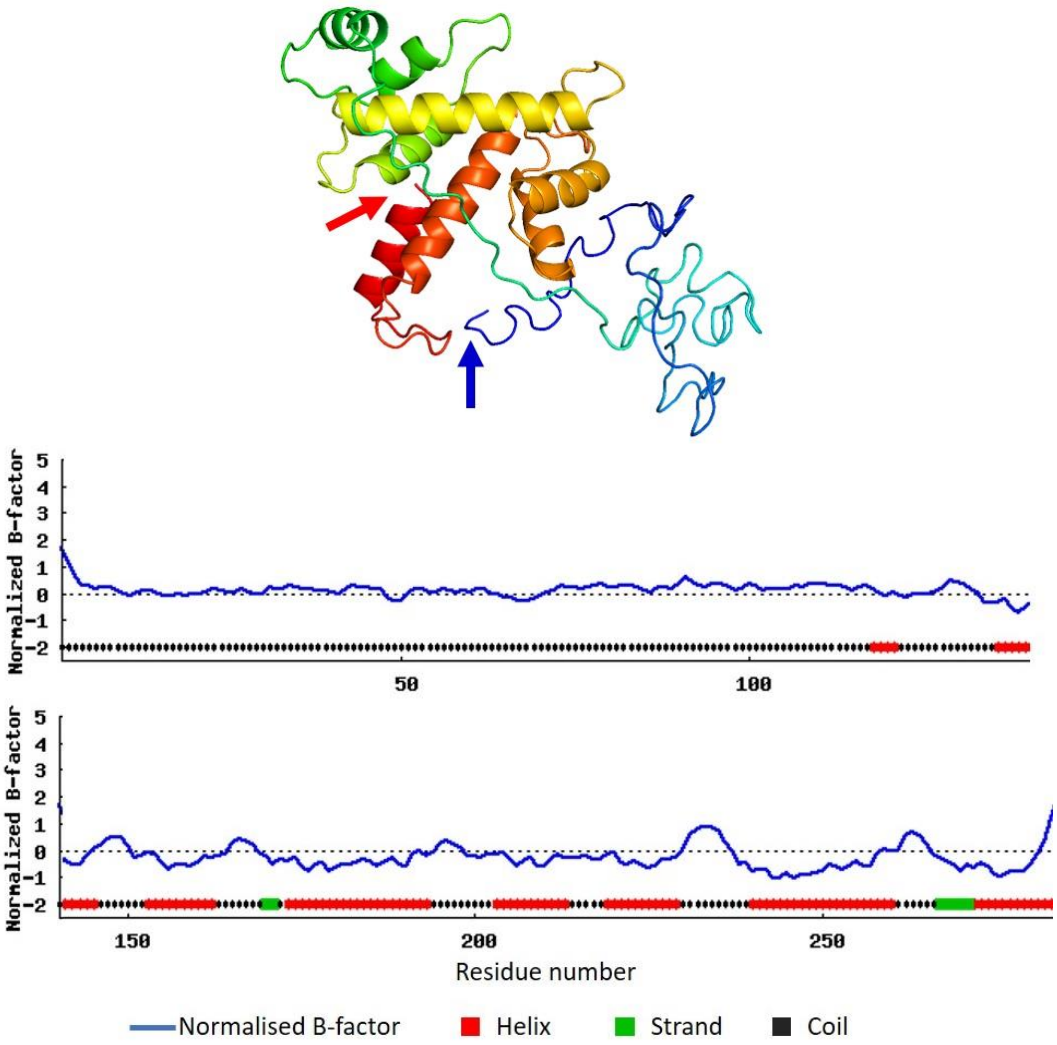


Fig.5.7.2. Structure prediction for human peflin. The full-length sequence of human peflin was submitted to the I-TASSER server to generate secondary structure predictions and models. Of the five possibilities, only the highest confidence model is shown. The N-terminus is indicated by a blue arrow and the C-terminus by a red arrow in the top panel.

Multiple C-terminal helices were predicted, consistent with the presence of the five EF-hands, which maintain a helix-turn-helix structure. Accordingly, the helical regions were likely to be more stable, as indicated by the normalised B-factor shown in the lower panel. The normalised B-factor is representative of the predicted stability of experimental structure: residues with a normalised B-factor >0 are likely to be unstable compared to residues with a B-factor <0. Despite the relatively low confidence of this model, predicted binding sites for calcium and peptides were annotated using I-TASSER (Fig.5.7.3).

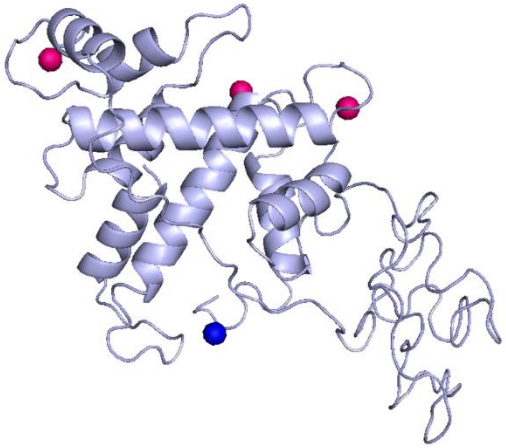


Fig.5.7.3. Binding sites for calcium and peptides are predicted on opposite faces of peflin. The full-length sequences for human peflin was submitted, calcium (pink) and peptide ligand (blue) binding sites predicted by the I-TASSER server.

Three calcium binding sites were predicted (Fig.5.7.3 pink). The server could only predict binding of a 'generalised peptide ligand' as opposed to a specific protein, however models suggested that calcium and ubiquitin would bind peflin on opposite binding faces. Therefore, it seemed unlikely that calcium directly participated in ubiquitin recognition by peflin. An alternative hypothesis might be that calcium could stabilise a specific binding conformation, or act as a molecular switch to stimulate recognition, or compete for a common binding site on ubiquitin with other ligands.

The data presented so far not only identified a novel ubiquitin-binding protein with the potential for a novel family of UBD, but also provided evidence for overlap between calcium and ubiquitin pathways. A number of these hypotheses were then tested by NMR.

5.8 INVESTIGATING INTERACTIONS OF UBIQUITIN, PEFLIN AND CALCIUM BY NMR

As a step towards validation of peflin as a novel ubiquitin-binding protein and to further investigate the binding site on ubiquitin, NMR titrations were conducted with ^{15}N-ubiquitin. In total four 2D HSQC spectra were acquired, all n=1.

Table 5.8.1. The HSQC spectra acquired to investigate the interaction between ubiquitin, peflin and calcium.

Spectra colour	^{15}N-ubiquitin	GST-peflin	Calcium
Blue	x		
Black	x		x
Purple	x	x	
Red	x	x	x

GST-peflin was produced and purified from 3L of growth culture into 600µL minimal capture buffer with 10% D_2O to give a final concentration of 400µM. The presence of folded protein was confirmed by 1D ^{1}H NMR. A 50µM ^{15}N-ubiquitin sample was prepared in the same buffer and confirmed by 1D NMR. A 2D HSQC spectrum was acquired for baseline control.

To observe the interaction between ubiquitin and peflin, the HSQC was repeated at an 8:1 molar ratio of 400µM GST-peflin to 50µM ^{15}N-ubiquitin in the absence of calcium. Spectra were assigned and CSP calculated. Significance was determined as number of standard deviations (SD) from the mean CSP across the spectra of interest, and a threshold for significance set to a CSP greater than one SD from the mean.

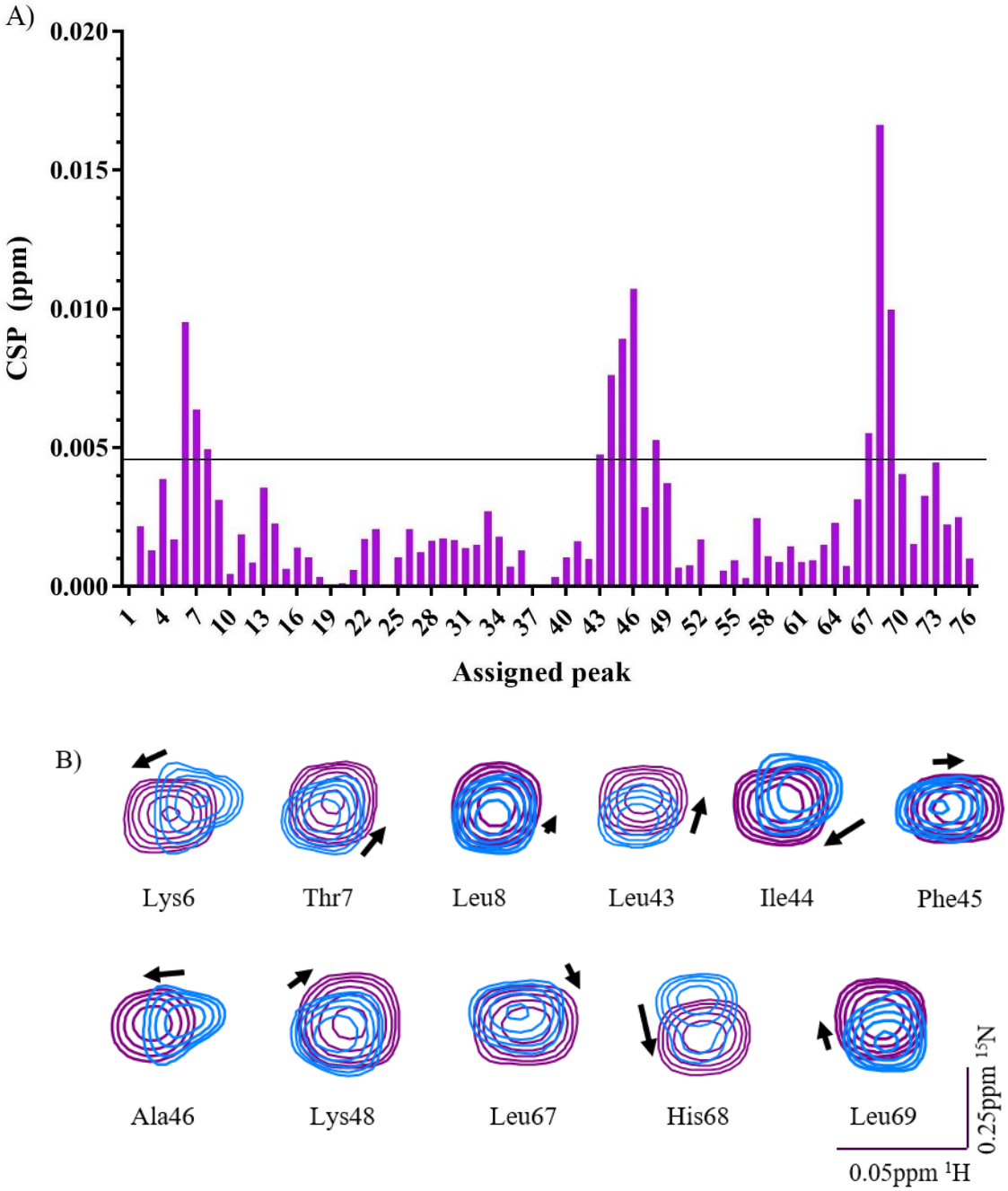


Fig.5.8.1. The effect of GST-peflin on ^{15}N-ubiquitin NMR spectra. HSQC spectra of 50µM ^{15}N-ubiquitin were acquired at 298K in minimal capture buffer with 10% D_2O in the absence and presence of 400µM GST-peflin. The mean CSP with 1SD was 0.0023ppm ±0.0023. A) The CSP of each residue is shown after addition of GST-peflin. Significance threshold is indicated by the black line. B) Residues with significant CSP from control are shown in the absence (blue) and presence (purple) of peflin, with the direction and qualitative amount of shift indicated by arrows.

In the absence of calcium 11 ^{15}N-ubiquitin residues displayed a significant CSP after addition of peflin, indicating that there had been a change in the residue microenvironment and that binding had taken place (Fig.5.8.1). Of these 11, five are canonically involved in ubiquitin-

binding interactions; Lys6, Leu8, Ile44, Ala46 and His68. It was notable that the interacting residues were clustered together, both in the linear amino acid sequence and on the folded protein surface, suggesting a discrete binding interface rather than a generalised conformational shift (Fig.5.8.2). This was clearly centred around His68, which also showed the greatest degree of CSP. Unexpectedly, Ile44 was not required for ubiquitin and peflin interaction in the affinity chromatography experiments (Fig.5.7.1) and yet showed significant shift here. It is important to note that whilst Leu67 was in close proximity to the other significant residues and it did demonstrate a significant CSP, it did not contribute to the surface of ubiquitin and was entirely internal. This may have been due to minor structural changes as a consequence of binding. The whole spectra can be found in the Appendices, Fig.A5.2.1.

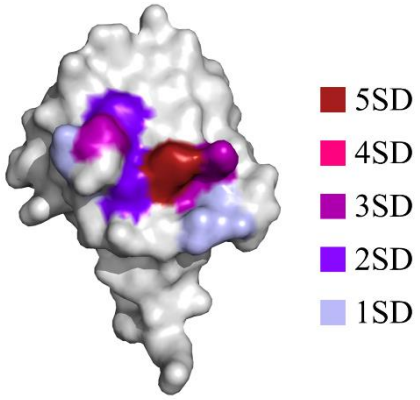


Fig.5.8.2. The binding footprint of GST-peflin on ^{15}N-ubiquitin. NMR spectra of 50μM ^{15}N-ubiquitin and 400μM GST-peflin in minimal capture buffer with 10% D_2O was acquired at 298K. Interaction was inferred by CSP and number of SD from mean indicated by colour. Image was made with PyMOL using PDB 1UBQ (Vijay-Kumar *et al.*, 1987).

The above experiment was next repeated with the addition of 250μM $CaCl_2$ to both samples to investigate the effect of physiologically high calcium seen in the pull-down experiments (Fig.5.8.3). The same samples of GST-peflin and ^{15}N-ubiquitin were used between experiments to prevent any difference in buffer conditions or protein concentration.

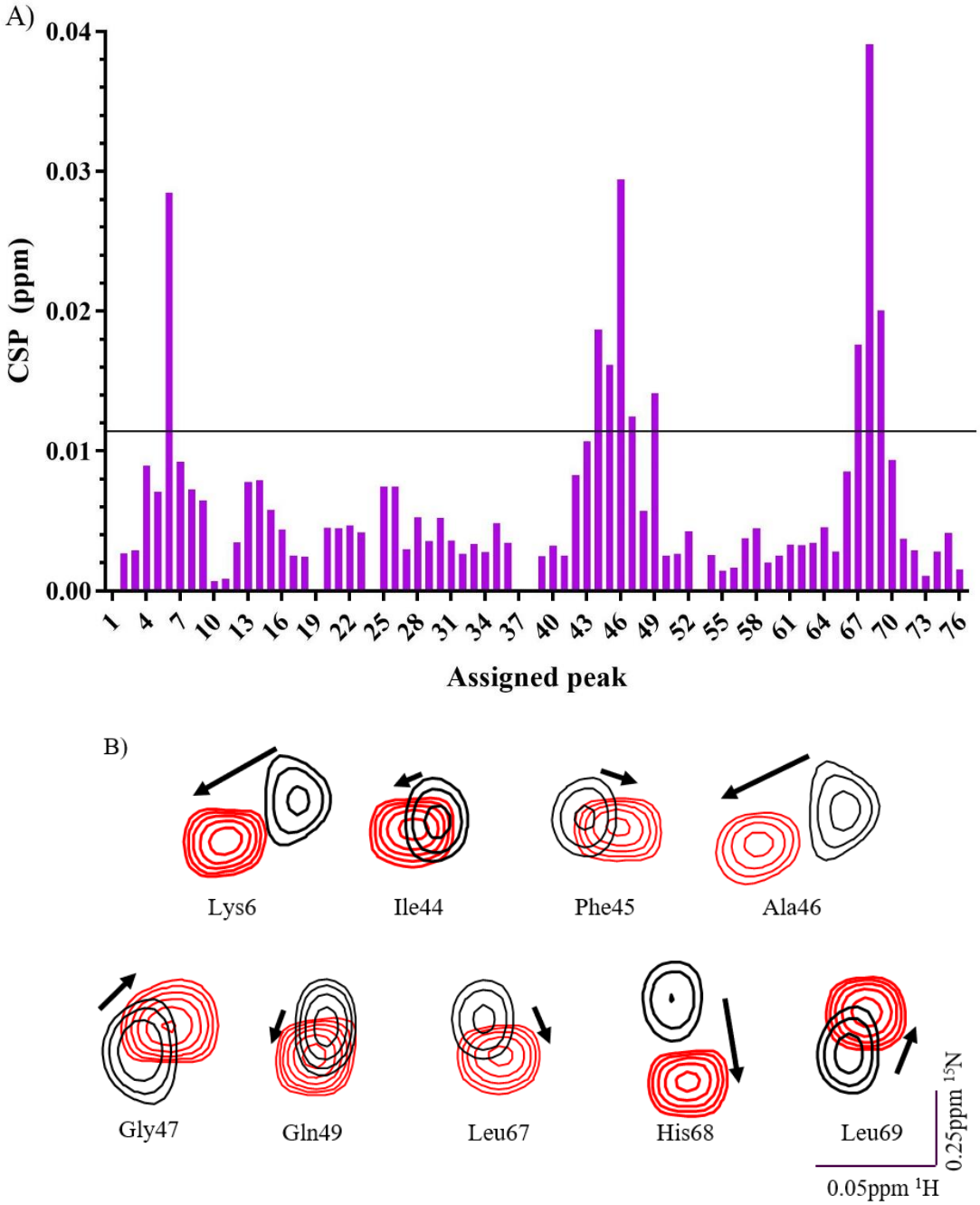


Fig.5.8.3. The effect of GST-peflin on ^{15}N-ubiquitin NMR spectra in the presence of calcium. HSQC spectra of 50µM ^{15}N-ubiquitin were acquired at 298K in minimal capture buffer with 10% D_2O and 250µM $CaCl_2$ in the absence and presence of 400µM GST-peflin. The mean CSP with 1SD was 0.006ppm ±0.0056. A) The CSP of each residue is shown after addition of GST-peflin. Significance threshold is indicated by the black line. B) Residues with significant CSP from control are shown in the absence (black) and presence (red) of peflin, with the direction and qualitative amount of shift indicated by arrows.

When in the presence of calcium there was a stronger average shift of ^{15}N-ubiquitin residues due to addition of GST-peflin, as indicated by the values depicted in Fig.5.8.3. Three peaks had separated entirely between the ^{15}N-ubiquitin (black) and ^{15}N-ubiquitin + GST-peflin (red) conditions, Lys6, Ala46 and His68. This indicated a considerable change in residue microenvironment and definite interaction between the two proteins in the presence of calcium. Two additional peaks showed significant CSP in the presence of calcium compared to in the absence; Gly47 and Gln49. However, four residues that showed a significant shift in the absence of calcium, Thr7, Leu8, Leu43 and Ly48, were not significant here.

Surprisingly, Leu8 did not show any significant CSP, despite its requirement in interaction between peflin and ubiquitin in affinity chromatography experiments (Fig.5.7.1) and its significance in the NMR spectra without calcium (Fig.5.8.1). Again, Ile44 had no effect in the pull down yet showed significant CSP effects.

Overall, the interaction between peflin and ^{15}N-ubiquitin caused a greater degree of shift over more residues when in the presence of calcium compared to the absence (Fig.5.8.4). Whilst the average level of significance was lower, the overall magnitude of the CSP effects was higher, indicating that more residues had minor CSP changes. Again, all the residues clustered around the same single patch centred around His68, indicating localised binding. The whole spectra can be found in the Appendices, Fig.A5.2.2.

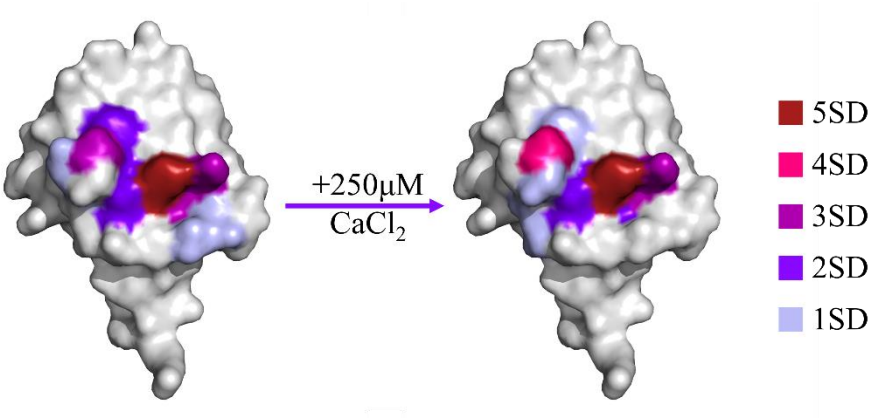


Fig.5.8.4. The binding footprint of GST-peflin on ^{15}N-ubiquitin. NMR spectra of 50μM ^{15}N-ubiquitin and 400μM GST-peflin in minimal capture buffer with 10% D_2O was acquired at 298K. This was repeated in the absence (left) and presence (right) of 250μM $CaCl_2$. Interaction was inferred by CSP. Amount of significant CSP was determined by standard deviations (SD) from the mean. Image was made with PyMOL using PDB 1UBQ (Vijay-Kumar *et al*., 1987).

In general, physiologically relevant high calcium concentrations both enhanced the shift of residues that already interacted and encouraged interaction at new sites. Most of the sites localised around a surface footprint that overlapped with the canonical Ile44 hydrophobic patch and stretched towards, but did not cover, the Asp58 site.

Importantly in the previous experiment calcium was present in both the ±GST-peflin conditions to ensure that GST-peflin was the only variable. If instead a comparison is made with the same protein sample before and after addition of calcium, then one can observe the effect of calcium on the ^{15}N-ubiquitin spectra. Unexpectedly, it appeared that the presence of calcium affected the ubiquitin structure in a GST-peflin-independent manner (Fig.5.8.5).

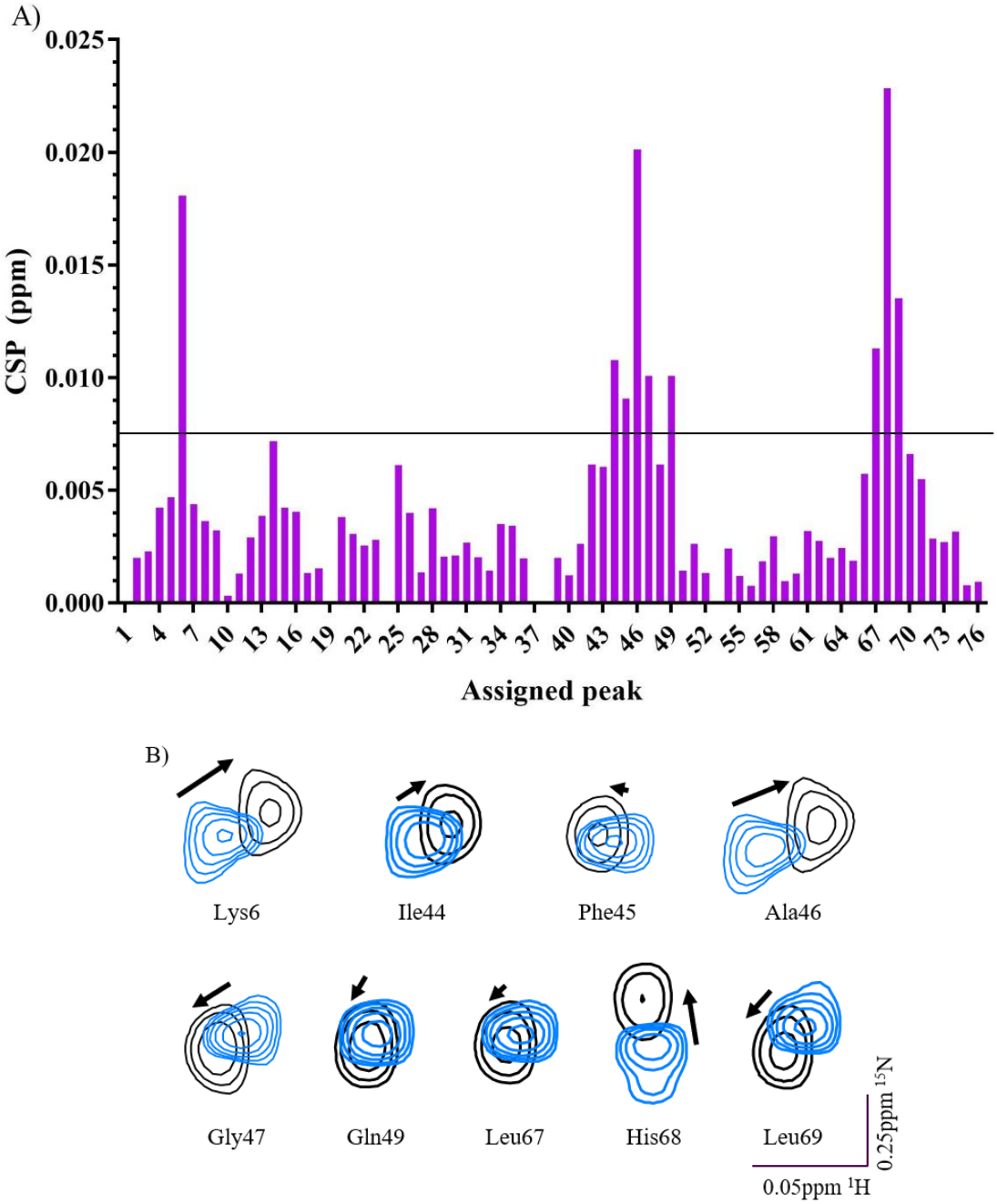


Fig.5.8.5. The effect of calcium on peak movement in ^{15}N-ubiquitin NMR spectra. HSQC NMR spectra for 50µM ^{15}N-ubiquitin in minimal capture buffer with 10% D_2O at 298K were captured in the absence (black) and presence (blue) of 250µM $CaCl_2$. The mean CSP with 1SD was 0.0039ppm ±0.0037. A) CSP were calculated for all peaks and the significance threshold determined by SD from the mean (black line). B) Residues with significant CSP from control are shown in the absence (blue) and presence (black) of calcium, with the direction and qualitative amount of shift indicated by arrows.

The majority of residues showed no response to calcium. Nine residues, however, demonstrated a shift after addition of calcium; Lys6, Ile44, Phe45, Ala46, Gly47, Gln49, Leu67, His68 and Leu69. These shifts were of varying direction and distance, indicating that it was not merely an effect of pH or buffer change. The affected surface was identical to that seen by interaction between GST-peflin and ^{15}N-ubiquitin in the presence of calcium, although perturbations were lower on average (Fig.5.8.6). The whole spectra can be found in the Appendices, Fig.A5.2.3.

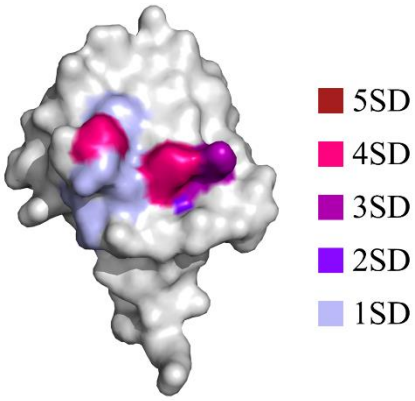


Fig.5.8.6. The structural effects of calcium on ^{15}N-ubiquitin. NMR spectra of 50μM ^{15}N-ubiquitin with 250μM $CaCl_2$ in minimal capture buffer with 10% D_2O was acquired at 298K. Structural changes were inferred by CSP compared to ^{15}N-ubiquitin alone. Ubiquitin residues with significant CSP are coloured according to the degree of shift, calculated as SD from the mean. Image was made with PyMOL using PDB 1UBQ (Vijay-Kumar *et al*., 1987).

It appeared that calcium caused a change in ^{15}N-ubiquitin structure, perhaps creating or stabilising a micro-conformation unseen without calcium or perhaps modulating the physicochemical properties of the canonical binding site, for which no precedent could be found within the literature.

By comparing the CSPs seen between conditions (Fig.5.8.1, Fig.5.8.3 and Fig.5.8.5) it was clear that the CSP after addition of calcium were not the same as those caused by the addition of GST-peflin: they were of different direction and distance. In other words, calcium and GST-peflin were causing two separate effects on ^{15}N-ubiquitin (although they may be able to act cooperatively).

This effect was confirmed by independent HSQC NMR results obtained by Dr Jed Long, although these experiments were still in progress at the time of submitting this thesis and so the data are not shown. Anecdotally, this preliminary data showed that 250µM $CaCl_2$ could alter the spectra of ^{15}N-ubiquitin, particularly around the canonical binding patch. This effect was observed in both human and yeast ^{15}N-ubiquitin samples, suggesting that it was evolutionarily conserved. Interestingly this appeared to show some buffer dependence, which could explain the contradictory data of calcium shown at the start of this chapter, although the exact details of this are still under investigation.

To determine if this ^{15}N-ubiquitin modulation by calcium was seen whilst bound to GST-peflin, GST-peflin-bound ^{15}N-ubiquitin spectra were compared in the absence and presence of calcium (Fig.5.8.7).

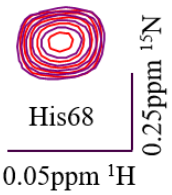


Fig.5.8.7. There was no difference in ^{15}N-ubiquitin with GST-peflin spectra with or without calcium. HSQC spectra of 50µM ^{15}N-ubiquitin with 400µM GST-peflin was acquired at 298K in the absence (purple) and presence (red) of 250µM $CaCl_2$ in minimal capture buffer with 10% D_2O. A single peak typical of the general spectra is shown.

Unexpectedly, it appeared that calcium had no effect on the ^{15}N-ubiquitin + GST-peflin spectra. How then, did the data shown in Fig.5.8.1 to Fig.5.8.4 show a difference in binding between calcium conditions? To investigate this further, all four spectra were compared (Fig.5.8.8).

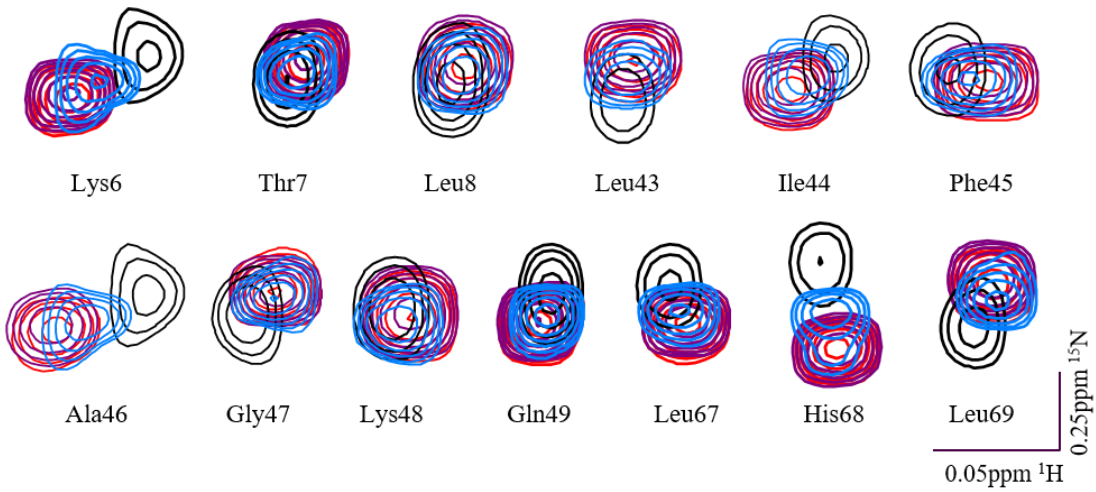


Fig.5.8.8. The effect of both calcium and GST-peflin on peak movement in ^{15}N-ubiquitin NMR spectra. HSQC spectra for 50µM ^{15}N-ubiquitin in minimal capture buffer with 10% D_2O at 298K were captured under four conditions; ^{15}N-ubiquitin only (blue), +250µM $CaCl_2$ (black), +400µM GST-peflin (purple) and +400µM GST-peflin with 250µM $CaCl_2$ (red).

Overlaying the spectra from all four experiments shows that GST-peflin caused shifts regardless of the presence of calcium (for the full spectra, see Appendix Fig.A5.2.4). Therefore, the shifts seen after addition of GST-peflin were valid and were not simply the effect of calcium on ^{15}N-ubiquitin. It still suggests that calcium modulates ubiquitin and GST-peflin binding as, whilst there was little difference seen between bound conformations with and without calcium, when compared to the matched ^{15}N-ubiquitin control the CSPs were greater in the calcium condition. Put simply, whilst the end point may have been the same the +calcium condition had further to travel before reaching the same end point, as illustrated by Fig.5.8.9. One possible rationalisation for this set of observations is that calcium and peflin binding are competitive and independent processes that involve some part of the same Ile44 canonical hydrophobic patch. The pattern of CSP effects can be explained if peflin is first required to displace a more weakly bound calcium ion from the same site on the ubiquitin surface. This would explain why the end points of the titration between ^{15}N-ubiquitin and peflin are very similar whether conducted in the presence or absence of calcium ions.

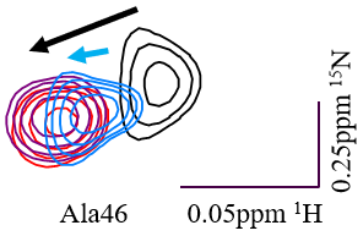


Fig.5.8.9. The effect of both calcium and GST-peflin on Ala46 peak movement in ^{15}N-ubiquitin NMR spectra. HSQC spectra for 50µM ^{15}N-ubiquitin in minimal capture buffer with 10% D_2O at 298K were captured under four conditions; ^{15}N-ubiquitin only (blue), +250µM $CaCl_2$ (black), +400µM GST-peflin (purple) and +400µM GST-peflin with 250µM $CaCl_2$ (red). The CSP due to addition of GST-peflin is indicated by arrows; blue for the absence of calcium and black for the presence of calcium.

As can be seen from the example of Ala46, there was a greater shift between the ^{15}N-ubiquitin + GST-peflin condition in the presence of calcium (black to red peaks; black arrow) compared to in its absence (blue to purple peaks; blue arrow). This reflects the fact that the nature of the CSP effects induced by the two different ligands is slightly different, or even in opposite directions in some cases. The angle of shift was also slightly different. If addition of calcium had affected both ^{15}N-ubiquitin and ^{15}N-ubiquitin + GST-peflin peaks in the same manner then it could have been due to a general buffer or pH change affecting the protein as a whole. However, addition of calcium shifted the ^{15}N-ubiquitin peak towards the top right, whereas addition of GST-peflin shifted the peak towards the bottom left. As such, it can be assumed that peak shift was seen due to a molecular interaction between ubiquitin, GST-peflin and calcium through independent binding events of the two ligands and was not due to a pH change. A similar effect was seen for all peaks.

By comparing all four spectra together it appears that two independent binding events were occurring; that between ubiquitin and calcium, and between ubiquitin and peflin. The data suggest that calcium ions can modulate binding to the ubiquitin hydrophobic patch, for which no precedent could be found within the literature.

Within the GST-peflin conditions there were also three additional peaks observed. As no reasonable explanation could be found for these and as they did not appear to affect the experimental data, the spectra of these peaks are included in the appendices (Fig.A5.2.5).

Due to the unexpected effect of calcium on ^{15}N-ubiquitin, no comparison of peflin binding should be made where calcium is also a variable. As previously discussed, it was unlikely that a pH change was responsible for the observed shifts. However, whilst all buffers and samples were pH adjusted before use, due to a lack of access to a NMR pH probe sample pH was not determined after addition of calcium. Direct comparisons where GST-peflin was the variable, as in Fig.5.8.1 and Fig.5.8.3, were still possible as any pH change would have been equal for them both.

In summary, the work presented in this chapter so far has identified a novel ubiquitin-binding protein, peflin. This interaction has been observed through several independent experiments under different conditions and so is unlikely to be a false positive. Peflin bound across the Ile44 patch of ubiquitin. NMR experiments confirmed that ubiquitin interacted with GST-peflin and, when combined with the affinity chromatography of cleaved peflin, gave a clear indication that cleaved peflin would also bind to ^{15}N-ubiquitin, although there was not sufficient time to conduct this experiment by NMR. Binding was potentiated by the presence of physiologically relevant levels of calcium but not other divalent cations such as magnesium chloride, as determined by affinity chromatography. Unexpectedly, calcium was also found to affect ^{15}N-ubiquitin in the absence of any protein binding partner during NMR. It was suggested that ubiquitin could present with slightly different micro-conformations, and that one of them was promoted by the presence of calcium, or that calcium modulated the binding patch of ubiquitin in some as of yet unknown manner.

5.9 THE EFFECT OF CALCIUM ON GLOBAL UBIQUITIN BINDING

Given the effects of physiologically relevant levels of calcium on ubiquitin structure and binding, it was possible that calcium-mediated regulation ubiquitin interactions could be a more wide-spread mechanism. To investigate this, an affinity chromatography of pig brain lysate to ubiquitin-Sepharose was conducted under identical conditions to that already discussed, with the addition of calcium to both control- and ubiquitin-Sepharose captures and the ubiquitin-specific proteins identified by LC-MS/MS (Fig.5.9.1).

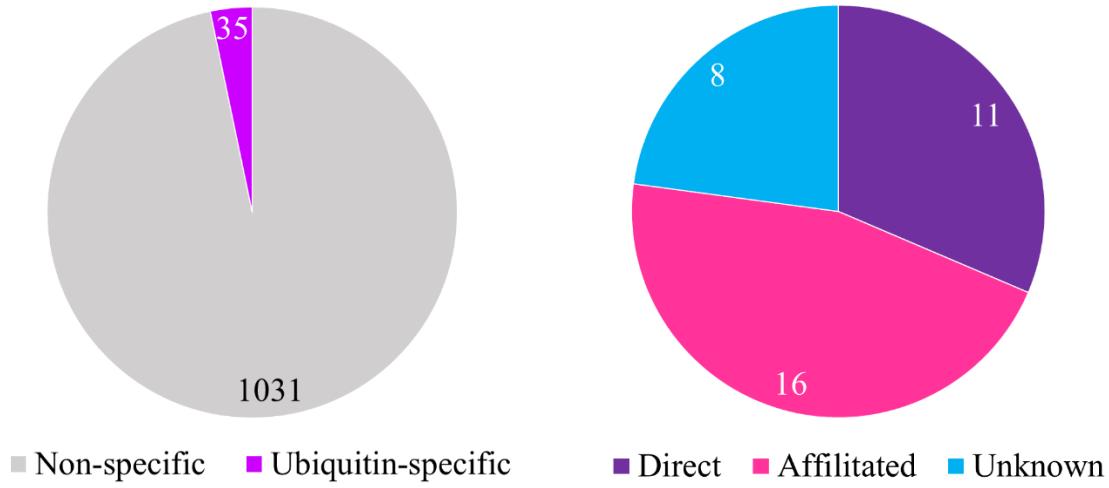
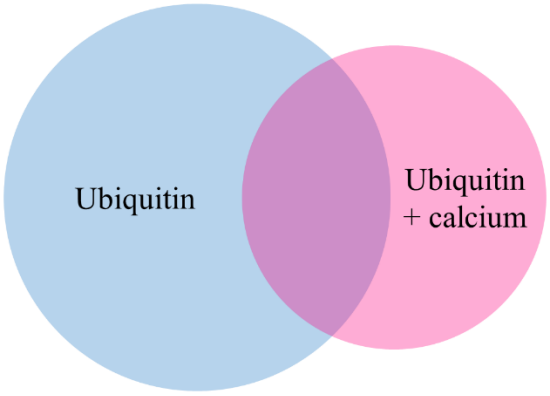


Fig.5.9.1. Identification of ubiquitin-specific human binding proteins in the presence of calcium. 6.7mg of total protein from pig brain lysate was captured on 140µL Sepharose coupled to human ubiquitin at 2mg/ml for 2hrs with end-over-end rotation at room temperature in the presence of 250µM $CalCl_2$. Captured proteins were eluted and 50% of the eluate identified by LC-MS/MS and mapped to human homologues. Left panel: proteins that specifically bound ubiquitin-Sepharose (purple) were normalised to non-specific binding to control-Sepharose (grey). Right panel: ubiquitin-specific genes were assessed for known interactions with ubiquitin though BioGRID 3.4 and GO term annotations. Genes were grouped into those with direct interactions (purple), affiliated with ubiquitination (pink) or with no known links (blue). Figure represents *n*=1.

There was good indication that the addition of calcium did not adversely affect the experiment as a whole, as similar numbers of ubiquitin-specific proteins were identified compared to the pull down in the absence of calcium (Fig.5.2.1, Fig.5.3.1 and Fig.5.9.1). It also provided a second independent identification of endogenous porcine peflin interacting with ubiquitin-Sepharose, repeating the results presented at the start of this chapter. Therefore, conditions were sufficiently controlled to make the two datasets comparable.

To determine if calcium changed the identity of the ubiquitin-interacting proteins, the two datasets of proteins binding in the absence and presence of calcium were compared, and any areas of overlap highlighted (Fig.5.9.2).



Ubiquitin + Calcium	
Gene	Protein
RAB15	Ras-related protein Rab-15
AKR7L	Aflatoxin B1 aldehyde reductase member 4
GAS7	Growth arrest-specific protein 7
APPL1	DCC-interacting protein 13-alpha
RAP1A	Ras-related protein Rap-1A
HRNR	Hornerin
ECI2	Enoyl-CoA delta isomerase 2, mitochondrial
KIF1A	Kinesin-like protein KIF1A
SCN2B	Sodium channel subunit beta-2
LYZ	Lysozyme C
GPX4	Phospholipid hydroperoxide glutathione peroxidase
CAPN2	Calpain-2 catalytic subunit
UBL7	Ubiquitin-like protein 7
MT3	Metallothionein-3
NDUFA7	NADH dehydrogenase
ENOPH1	Enolase-phosphatase E1
EIF5	Eukaryotic translation initiation factor 5
SMAP1	Stromal membrane-associated protein 1
ABCE1	ATP-binding cassette sub-family E member 1
PGM3	Phosphoacetylglucosamine mutase
RPS7	40S ribosomal protein S7

Fig.5.9.2. Some ubiquitin-binding proteins only affinity chromatography purified in the presence of calcium. The proteins specifically binding to ubiquitin-Sepharose in the absence or presence of 250μM $CaCl_2$ were compared. 42 proteins only bound in the absence of calcium (blue), 21 only in the presence of calcium (pink), and 14 occurred in both conditions (purple).

Addition of calcium to the protein capture buffer did not significantly change the number of bound proteins but it did change the individual proteins that interacted with ubiquitin-Sepharose (Fig.5.9.2). Of the 77 proteins that were ubiquitin-specific across the two experiments, the binding of 14 were unaffected by the presence of calcium. 42 of the proteins only appeared in the absence of calcium and 21 only affinity chromatography purified in the presence of calcium. The pig brain lysate input sample was the same for both experiments, so it is unlikely that this was due to changes in the protein profile of the source material and therefore appeared to be due to the presence of high calcium concentrations.

For a list of the proteins that were only present in one of the two conditions, and therefore may be completely abolished or facilitated by the presence of calcium, please see Appendix 5.1.

Despite appearing in both datasets, the 14 proteins that precipitated in both the absence and presence of calcium were not necessarily consistent between conditions. To determine if the presence of calcium had more subtle effects on protein binding, the total spectrum count of commonly bound proteins was compared across experiments (Fig.5.9.3).

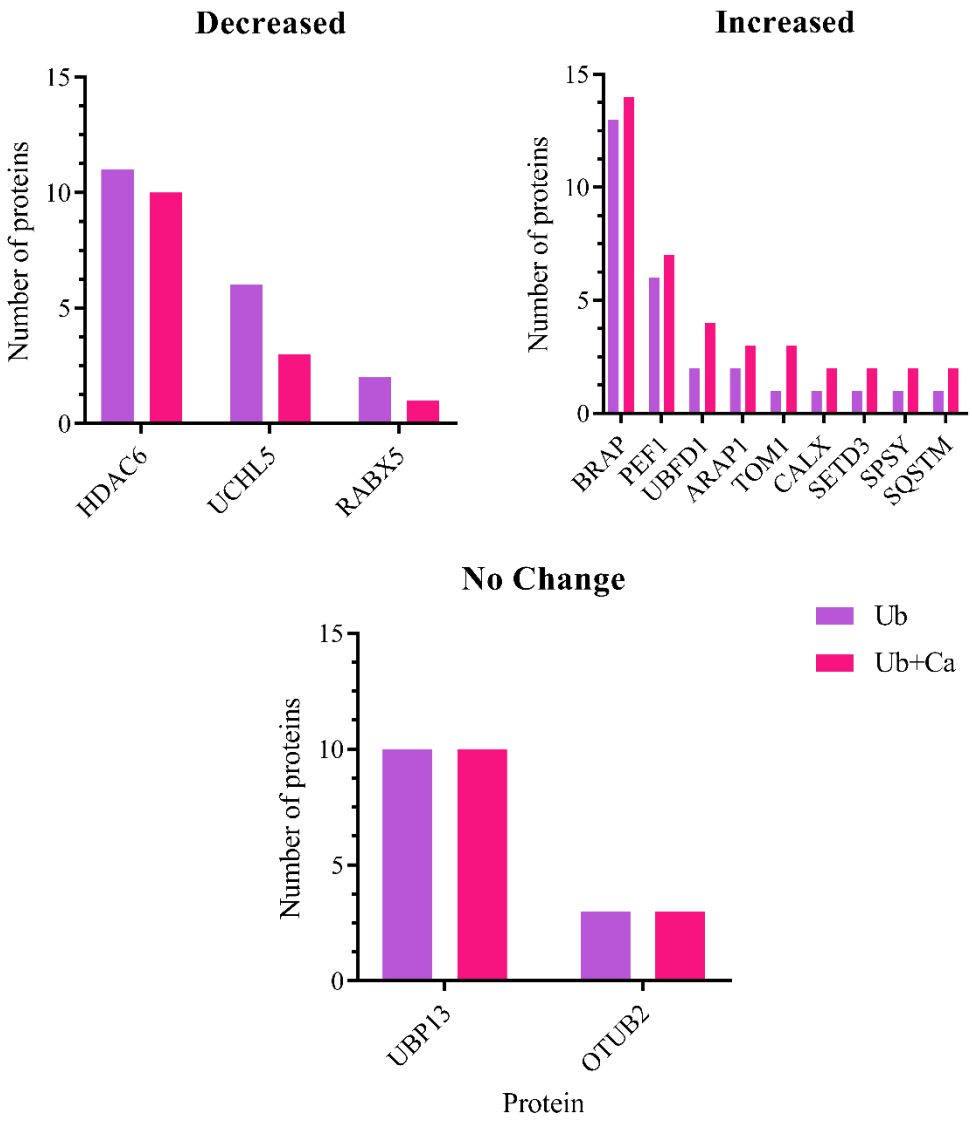


Fig.5.9.3. Total spectrum count of common ubiquitin-interacting proteins changed after addition of calcium. The total spectrum count of 14 proteins specifically binding to ubiquitin-Sepharose in both the absence (purple) and presence (pink) of 250μM $CaCl_2$ were compared. Addition of calcium decreased the total spectrum count of three, increased nine and had no effect on two proteins.

The majority of proteins (nine) increased after addition of calcium, which importantly included peflin (*PEF1*). Of those nine proteins, two were known to bind calcium; peflin and calnexin (*CALX*). Three proteins, HDAC6, UCHL5 and RABX5, decreased after addition of calcium, and two showed no change. The total spectrum count changes observed were very small but given the low initial values still represented large proportional changes.

Analysis of the entire dataset of the 35 proteins that bound ubiquitin-Sepharose in the presence of calcium did not show any significant enrichment of GO molecular functions or protein

classes compared to the levels expected in the human proteome. Four GO biological processes were identified with significant enrichment compared to the expected human proteome, however these all related to ubiquitination or ubiquitin-mediated processes and so are unlikely to be a consequence of calcium addition. Interestingly, two processes were significantly underrepresented when in the presence of calcium. These were "protein targeting" and "protein targeting to membrane" (GO:0006605 and GO:0006612, respectively). As ubiquitin normally aids protein targeting one would expect them to be enriched after ubiquitin-Sepharose capture. The significant underrepresentation of these processes when in the presence of calcium was unexpected. This suggested that calcium could work to inhibit ubiquitin-mediated protein targeting, although this is highly speculative as no follow up experiment was conducted.

Four proteins had known links to calcium, as determined by the presence of an EF-hand domain, confirmed calcium-binding, or calcium-related GO biological processes, molecular functions or protein classes. These proteins were; Peflin, Calnexin, Hornerin (*HRNR*) and Calpain-2 (*CAPN2*). The latter two proteins only bound in the presence of calcium.

Whilst this could be taken as an indication that calcium regulated the ability of multiple proteins to bind ubiquitin as a general mechanism, it should not be taken as certain due to lack of repeat experiments and the low differences in total spectrum count between conditions. It was also possible that addition of high calcium concentrations caused a large enough change in ionic strength as to affect protein binding in general. In order to confirm that this was an effect of calcium, the experiments should be repeated with an alternative metal ion, such as magnesium.

5.10 CONCLUSIONS

Based on the inability to validate NGPD-derived candidate novel UBD peptides, more physiologically relevant methods were used to identify novel ubiquitin-binding proteins with the potential for the identification of novel UBDs. This worked towards confirming the

hypothesis that novel ubiquitin-binding partners could exist in the background of affinity chromatography experiments, so far ignored in favour of studying those with known physiological relevance. Accordingly, the experiments were designed to avoid bias towards known ubiquitin-binding proteins: mammalian tissue was screened for proteins that interacted with ubiquitin-Sepharose with trend analysis based on interactions and biological processes. This strategy was successful in identifying proteins that had no known direct or indirect interaction with ubiquitin.

One of these proteins, peflin, was confirmed as a directly binding ubiquitin-binding protein, interacting with the canonical hydrophobic binding patch of ubiquitin. This interaction was modulated by calcium, which to the author's knowledge is the first instance of crosstalk between calcium and ubiquitin recognition. Domain annotation confirmed the lack of any known UBD within peflin, suggesting that peflin not only represents a novel ubiquitin-binding protein, with a novel manner of calcium-regulated binding, but also the potential for a novel family of UBD.

Further, ubiquitin was seen to react to physiologically relevant levels of calcium. It was suggested that calcium modified the canonical Ile44 hydrophobic patch in ubiquitin, and that this could affect ubiquitin recognition on a wider scale.

6 GENERAL DISCUSSION

6.1 THE DARK MATTER OF UBIQUITIN-RECOGNITION – WHY DOES IT MATTER?

Selective recognition of ubiquitin modifications by ubiquitin-binding domains (UBDs) underpins a diverse array of processes, many with biomedical relevance (for comprehensive reviews, see Rape, 2018; Gilberto and Peter, 2017; Li *et al.*, 2016; Rittinger and Ikeda, 2017; Courtois and Fauvarque, 2018; Sharma *et al.*, 2016; Mansour, 2018). The study of the specific creation, removal and recognition of ubiquitin modifications therefore informs many other areas of biological research and therapeutic development. The selectivity inherent in this system can be broadly achieved via mechanisms based on sub-cellular localisation, or recognition of ubiquitin chain length, linkage topology and post-translational modification. The number of functionally-verified endogenous UBDs is strikingly low compared to the still expanding variety of polyubiquitin conformations and ubiquitin-mediated processes, suggesting the existence of undiscovered UBDs.

The study of ubiquitin-mediated processes has described the space that novel UBDs can fill and yet little significant progress has been made in discovering them. Isolated examples of novel UBDs continue to be identified, each contributing something new to the understanding of ubiquitin recognition. This led to the major hypothesis of this thesis; that as of yet undiscovered families of UBDs and aspects of specific ubiquitin-recognition exist. It may be that we lack sufficiently specialised tools to study that which we do not know, or it may be that there are in fact entire new aspects to ubiquitin-recognition beyond UBDs such as crosstalk with other major cell signalling mechanisms. Three approaches were taken;

1. Bioinformatics: creating a resource for the discovery of known UBDs within novel proteins

2. NGPD: a high throughput method to identify truly novel artificial ubiquitin-binding peptides
3. Proteomics: assessing endogenous mammalian proteins for interaction with ubiquitin to identify physiologically relevant ubiquitin-binding proteins.

6.2 DEFINING CONSENSUS SEQUENCES FOR HUMAN UBDS – A BIOINFORMATICS APPROACH

Identification of novel instances of known ubiquitin-binding domains is often achieved by sequence alignment to a UBD consensus. During the creation of this thesis, the existing literature on UBD consensuses was found to be incomplete, out of date or generally not fit for purpose. No comprehensive published list could be identified, leaving an obvious opportunity for the creation of a consensus or SLiM resource. A systematic pipeline was created to standardise the creation of these SLiMs, and it was used to generate up to date, human-specific, and where possible, functionally relevant motifs for each of the 19 human UBD families which contained more than one human protein.

Unfortunately, many of the unique or newly identified UBDs have not been reported in any bioinformatic database and so their inclusion in this resource was reliant on finding the original publications online. Whilst comprehensive literature searches were conducted throughout the length of the PhD and the advice of many ubiquitin experts was sought at all stages, if certain key words were not included in the publication then it was unlikely to turn up in a literature search. As such, whilst every effort was made to create a comprehensive resource, there exists the possibility that domains could have been missed.

New or updated consensus SLiMs were created for 74% of the families, with the remaining 26% already being suitable (14/19 and 5/19, respectively). Proof of concept for the usability of the resource was conducted by interrogation of a publicly available database, a large peptide

library generated in Chapter 4, and the candidate ubiquitin-binding protein, peflin, identified in Chapter 5. Searching all these examples for homology to the 19 families took minutes as opposed to the hours or days that would have previously been required to search the literature for relevant, up to date, verified and species-relevant consensuses prior to the creation of this resource.

As a general rule, bioinformatics databases and annotations are only as good as their input, which is often low-throughput and can lead to biased and self-affirming data. Alignments should not be used as the sole domain annotation method and should always be experimentally validated. It must be acknowledged that whilst the alignments and SLiMS created here were accurate at the time of thesis submission (September 2019), they are unlikely to remain so. As new UBD-containing proteins are identified and existing UBDs are re-examined the population of a given UBD family will drift away from the defined 'consensuses'. It is for this reason that the reproducible pipeline was developed; to allow revision of the source data and review the SLiM in light of new knowledge.

As high throughput and datamining methods develop, reliable bioinformatic data will be required to interrogate large datasets and accurately identify potential ubiquitin-binding domains or proteins for further study. It is expected that this resource will be highly useful to other ubiquitin-recognition studies, providing an accessible list of collated consensus sequences to allow easy cross-referencing and rapid screening of large datasets.

6.3 NEXT-GENERATION PHAGE DISPLAY TARGETING MONOUBIQUITIN TOWARDS GENERATING NOVEL UBIQUITIN-BINDING CANDIDATE PEPTIDES

The application of NGPD to the identification of novel ubiquitin-binding peptides had great potential and was supported by several recent successful uses of Phage Display in the development of E3 ligase inhibitors (Gorelik *et al.*, 2018) and of ubiquitin-variants with

enhanced affinity towards the endogenous UIM target (Manczyk *et al.*, 2019) or as specific binders for DUBs (Spiliotopoulos *et al.*, 2019). It seemed the next logical step to use this high-powered method to identify novel ubiquitin-binding domains. Surprisingly, there had been no mention of this in the literature – even the CUBAN (Castagnoli *et al.*, 2019) and CoCUN (Nepravishta *et al.*, 2019) domains had originally been identified as NEDD8-binders.

The potential of this approach was not realised, however, as candidate peptides resulting from the NGPD could not be functionally validated using alternative methods. In hindsight, all the published data surround Phage Display and ubiquitination used ubiquitin-based libraries or targeted bait proteins other than ubiquitin (e.g., E3 ligases) – this seeming opportunity may in fact have been conspicuously left unaddressed as other research groups attempted and failed to use ubiquitin as the sole bait protein for NGPD-based novel UBD identification.

A number of factors could have contributed to the inability to functionally validate the seven candidate peptides chosen. These range from the underlying construction of the phage library to the final validation method chosen.

The coverage of a displayed peptide library is defined as the number of peptide variants displayed compared to the number of peptide variants possible. For example, assuming equal amino acid usage a library of 16mer peptides would have to contain 20^{16} variants for full coverage. The library used here, however, only contained a diversity of 10^7 variants. Whilst this may seem to be unfeasibly small library coverage, it must be remembered that the library still contained 10,000,000 different potential ubiquitin-binding peptides, which is still very high throughput for a protein binding experiment. Further, it does not follow that higher library coverage will yield high affinity binding peptides. Rather than increasing the coverage of the naïve library it is better to increase the complexity of the library; i.e. in order to identify high

affinity candidates it is better to achieve a low coverage analysis of a longer peptide library than to thoroughly analyse a small library (Muñoz and Deem, 2008).

A generalised NGPD protocol was available from other members of the Gough lab; however, this protocol was optimised for use in antibody Phage Display and involved comparatively harsh experimental conditions as antibodies typically present much higher binding affinities than UBDs (picomolar vs. micromolar K_d ranges, respectively) (Sokratous *et al.*, 2012; McVoy *et al.*, 2012). The NGPD protocol underwent extensive optimisation for ubiquitin-targeted biopanning before the final experiments were conducted, including optimisation of wash buffers, selection pressure and 96 well plate choice in an effort to reduce non-specific binding (data not shown). Despite this, the majority of NGS reads were conducted on peptides that only occurred one to five times in a sample, i.e., to low abundance peptides that were likely to be non-specific or weak binders. It was possible that these high levels of low abundance peptides skewed the candidate selection. Indeed, the same 96 well plates were later used by another member of the laboratory for a related, but distinctly different, NGPD protocol and yielded similarly high and variable non-specific binding.

It is known that many UBDs adopt a helical structure (Mattern *et al.*, 2019), which may have been discouraged in the TG1 hosts as *E. coli* have a reduced ability to introduce protein posttranslational modifications, many of which are required for correct protein folding (Georgieva and Konthur, 2011). The average predicted helicity content of the seven candidate peptides was low (data not shown). This was not initially considered to be an issue as it was of less importance when using a completely naïve or synthetic peptide library with no constraints on required protein folding, but it may have skewed the phage culture away from high affinity UIM-like ubiquitin-binding peptides. As such, whilst the candidates chosen were the best available, they may still have been low affinity binders or false positives.

This tied into broader issues around differences in peptide presentation between NGPD and NMR. The target ubiquitin was immobilised on the 96 well plate and the peptide presented on the phage surface during NGPD. It may be that this allowed for multi-peptide complexes to form, where one phage peptide bound between several ubiquitin molecules or vice versa. This would have contributed to avidity effects and the increased affinity would not have translated to the solution-based NMR experiment.

Alternatively, presentation of ubiquitin or phage peptide could have affected the protein conformation, changing the surface exposure or opening the globular structure to expose the hydrophilic core and promote binding. If this were the case, then it would explain the inability to validate binding through solution-based NMR. This could also work towards explaining how interaction partners can differentiate between ubiquitinated proteins and unanchored ubiquitin, an endogenous form of ubiquitin that is not conjugated to any target yet still brings about physiological effects (Di Rienzo *et al.*, 2019). This concept is supported by literature discussed below (Section 6.4.3) on the existence of hidden ubiquitin micro-conformations with differing binding selectivity. In this scenario, conjugation to the target would act analogously to immobilisation, slightly altering the ubiquitin structure and creating a distinct signal from free ubiquitin. Intriguingly, this also ties in with the later work on calcium-induced changes in ubiquitin recognition. Whilst this will be explored further in the next section, it does support the notion that minor conformational changes can affect ubiquitin recognition, such as may occur after immobilisation of ubiquitin on a plate surface. This would explain why candidates from NGPD were unable to be validated through NMR.

This could be relatively easily investigated, by obtaining NMR spectra of ^{15}N-ubiquitin bound to a known ubiquitin-binding protein in the absence and presence of low concentrations of urea. In this experiment, the urea would slightly denature the ubiquitin, mimicking the potential conformation change discussed above. If differences in interaction-induced CSP were seen

between conditions then it could indicate that minor conformational changes or unfolding alter ubiquitin-recognition. A similar experiment was conducted several years ago by Holstein *et al.* (2013); 2M urea was suggested to shift the predicted ubiquitin ligand binding site towards the C-terminal tail, altering affinity and presumably selectivity towards the protein ligand. The paper investigated binding of a predicted ligand rather than any endogenous protein so the direct physiological relevance of the paper is limited, however it serves as proof of concept that slight structural changes could affect recognition. Alternately, solid-phase NMR could be conducted on candidate interaction with ^{15}N-ubiquitin-Sepharose.

Finally, validation through NMR was not conducted with the most appropriate ubiquitin target. The candidate peptides had been chosen according to comparison between target groups, to find those peptides that bound S65E-ubiquitin but not wildtype-ubiquitin, for example. However, all NMR experiments were conducted on ^{15}N-wildtype-ubiquitin. It had been expected that, even though peptides showed selectivity during NGPD, the artificially high concentrations and molar ratios present in NMR would overcome this, and that minor binding would still occur through the rest of the peptide-ubiquitin interface. After this had been confirmed, binding to the phospho- or phosphomimetic-ubiquitin would be assessed and compared to wildtype to determine selectivity. However, as no binding was seen to wildtype ubiquitin for any peptide these secondary experiments were not conducted. It is possible that a peptide might have an absolute requirement for the S57E, S65E or pS65 residue, however this is highly unlikely.

It should be clarified that, whilst recent papers have been published on the inability of phosphomimetic-ubiquitin to fully recreate the biochemical effects of phosphoubiquitin (for example, Dong *et al.*, 2017), this was not as issue here as no peptides were selected based on their ability to discriminate between mimetic and phosphoubiquitin. Also, as none of the

peptides could be validated, this never became an issue as interaction against endogenous ubiquitin was never assessed.

Whilst a few false positives were to be expected, it was unusual that all seven tested candidates were unable to be validated. There are many possible explanations for this, it is currently unknown which, if any of them, caused the high false positive rate. The use of NGPD to identify novel UBDs is still a highly attractive idea, but it will remain unfeasible until the challenges discussed here are overcome.

To more successfully exploit this opportunity, a similar but alternative approach could be used. For example, combining Phage Display with protein biotinylation improves target immobilisation and epitope display, which would overcome the possible presentation issues described above (Koide *et al.*, 2009). Traditionally, the pIII and pVIII bacteriophage coat proteins are used for library display. As an alternative to this, display on the both the pIX and pVII proteins has been investigated but not widely adopted (Løset *et al.*, 2011). In general, both pIX and pVII coat proteins were found to tolerate display of fusion peptides well and provided an efficient model for Phage Display. Although it must be noted that the copy number was lower compared to traditional display peptides.

6.4 PROTEOMIC IDENTIFICATION OF UBIQUITIN-BINDING DOMAINS AND UBIQUITIN-BINDING PROTEINS

The use of proteomics to identify novel ubiquitin-binding proteins from pig brain lysate was highly successful; not only did it identify a novel ubiquitin-binding protein, peflin, which could contain a novel UBD, it also opened up a new line of research into crosstalk between ubiquitin and calcium signalling.

Peflin interaction with ubiquitin was demonstrated through several different, independent experiments. It was purified from pig brain lysate with ubiquitin-Sepharose beads on three

separate occasions under different conditions, it was shown to interact with ubiquitin-Sepharose as both a GST-tagged and cleaved protein, from overexpressed bacterial lysates, and finally through two NMR experiments with ^{15}N-ubiquitin (Fig.5.2.1, Fig.5.7.1., Fig.5.9.3, Fig.5.6.1, Fig.5.6.2., Fig.5.8.1. and Fig.5.8.3, respectively). This shows that the interaction occurred at physiological levels of peflin, directly without the need for adaptors or scaffold proteins, and in the absence of any E1/E2/E3 ligases or ATP, precluding the possibility that peflin was ubiquitinated rather than interacting with ubiquitin as a ligand. Taken together, these experiments suggest with a high degree of confidence that peflin bound ubiquitin non-covalently and indicated that this interaction could occur in physiological systems.

Whilst no investigation into the peflin binding site was conducted, no known UBD could be found within the protein by proteomic database searches, literature searches or alignment to known UBD families. Whilst the possibility exists that peflin does contain a known form of UBD that is simply too divergent from the consensus sequence to be currently identified, this was highly unlikely, and it was more feasible that peflin contained a novel UBD.

Experiments conducted through both protein NMR and affinity chromatography with Western blot suggested that the peflin-ubiquitin interaction was modulated by physiologically high concentrations of calcium. Initial data suggest that this modulation took the form of potentiated binding, which was supported by its near completed abrogation in the presence of the calcium chelator EDTA, but further experiments are required to confirm this.

With regards to the protein NMR investigations into the effect of calcium on peflin-ubiquitin interactions and ubiquitin micro-conformations, the effects of a pH change could not be ruled out due to the unavailability of a functioning pH probe that could accommodate the small 600μL NMR sample size. However, a pH change is highly unlikely to be the cause of the observed effects for two reasons. First, the same pH adjusted buffer was used to prepare all

samples with and without calcium, and second, a pH change tends to affect a broad range of dispersed residues, which was not the case as only a handful of localised peaks showed significant movement. This suggested a specific, localised effect of calcium on the ubiquitin conformation and suggested that ubiquitin could react to changes in local calcium concentrations to modify interactions with binding partners.

6.4.1 Peflin, a little studied candidate ubiquitin-binding protein

Very little is known about peflin – at the time of submitting this thesis (September 2019) a PubMed search for 'peflin' resulted in just 19 publications, most of which focused on other proteins. It is 284 amino acids long, ~30kDa and expressed in a wide range of animal tissues (reviewed, Maki *et al.*, 2002). It is a member of the wider penta-EF hand (PEF) protein family along with calpain, sorcin, grancalcin and programmed cell death protein 6/apoptosis-linked gene 2 (PDCD6/ALG-2) (reviewed, Kawasaki et al, 2019). Each of these proteins contains eight α-helices and five EF-hands formed of a helix-loop-helix motif, through which they can dimerise (Maki *et al.*, 2011). Of note, peflin is often found either as a homodimer or, when in the absence of calcium, a heterodimer with its closest paralog ALG-2 (Kitaura *et al.*, 2001; Kitaura *et al.*, 2002), implying that the homodimer forms in the presence of calcium.

The PEF family proteins are also known to contain multiple binding sites for their various peptide ligands, referred to as hydrophobic pockets. ALG-2 is known to contain two binding pockets when in a homodimer (Maki *et al.*, 2011), suggesting that the dimer of peflin may do likewise. This opens up the possibility for avidity and selectivity in peflin polyubiquitin binding, with one ubiquitin moiety binding at each pocket. Whilst some research has been published into the binding pockets of other PEF proteins, none has yet been published referring to peflin specifically (Kawasaki *et al.*, 2019).

It may be possible that calcium affects both peflin and ubiquitin, which would certainly tie in with peflin's function as a calcium binding protein. This could be supported by work on the calcium dependence of ALG-2; interaction between calcium and ALG-2 causes a conformational change in the side chain of ALG-2 Arg125, which is located in a loop between helices 5 and 6. This opens the binding site to expose the hydrophobic patch and promote ligand binding. Although caution must be taken as the binding pockets of the PEF family are not particularly conserved and peflin may differ from ALG-2.

Peflin has been seen to localise to the nucleus of hepatocellular carcinoma cells (HCCs) (Yang *et al.*, 2010), the cytosol in the absence of calcium and to membranes, potentially the endoplasmic reticulum (ER), in the presence of calcium (Maki *et al.*, 2002; Kitaura *et al.*, 2001).

The function of peflin has yet to be determined, although preliminary data has been published on its involvement as an adaptor for the Cullin-3 (CUL3) ubiquitin ligase and as a regulator of ER-to-Golgi transport. Whilst the majority of the work has described peflin as a mediator of CUL3 and vesicle formation, it is known to be involved in other contexts including cell cycle control through interaction with Histone 2AX (Yang *et al.*, 2010) and calcium-dependent interaction with the mechanosensitive (TRPN1) channel in sensory neurons and hair cells (Wiemuth *et al.*, 2012). Very little is known about these processes, however, and so they are not discussed here.

The E3 ligase CUL3 ligase is a regulator of neural crest specification during craniofacial development and vesicular secretion of collagen during bone formation, and has been linked to autophagy, maintenance of normal sleep (Freeman *et al.*, 2013), mitosis, endocytosis, hypertension, autism, myopathy and schizophrenia (Jin *et al.*, 2012; Werner *et al.*, 2015; reviewed, McGourty *et al.*, 2016). CUL3 acts as a complex with various adaptors, and the

adaptor determines the ubiquitination target. For example, the adaptor KBTBD8 directs CUL3 towards nucleolar protein ubiquitination (Werner *et al.*, 2015), whereas the adaptor Kelch-like protein 12 (KLHL12) promotes sec31 ubiquitination (McGourty *et al.*, 2016). Transient rises in intracellular calcium increase the activity of the CUL3-KLHL12 complex; this is mediated by the target-specific co-adaptor formed of peflin and ALG-2 and leads to a rise in monoubiquitination of the coat protein complex II (COPII) component sec31 to drive COPII coat formation and vesicular secretion of collagen from the ER (McGourty *et al.*, 2016; Jin *et al.*, 2012; reviewed Guo *et al.*, 2017). Peflin was required for stable interaction between CUL3-KLHL12 and sec31 and was found to act as a scaffold or adaptor protein; peflin bound the KLHL12 subunit of CUL3-KLHL12 and ALG-2 simultaneously. ALG-2 then recruited sec31 as the CUL3-KLHL12 substrate to promote E3 ligase activity and sec31 ubiquitination.

As an extra layer of complexity, peflin was inferred to be covalently ubiquitinated in a CUL3-dependent manner on Lys137 of EF hand 1. Ubiquitination altered the interaction between peflin and ALG-2. As described above, unmodified peflin forms a heterodimer with ALG-2 when in the absence of calcium, and the dimer dissociates when in the presence of calcium. Ubiquitinated peflin demonstrated the opposite binding behaviour, however, with ALG-2 heterodimerisation only occurring at calcium concentrations of at least 130nM. This may create some form of ubiquitination cascade, where both increases in intracellular calcium and peflin ubiquitination are required for CUL3-mediated ubiquitination of Sec31 and vesicle formation.

In this manner, peflin and ALG-2 were suggested to regulate ubiquitination by CUL3 in a calcium-dependent manner, ultimately altering protein ubiquitination and vesicle release in response to varying cellular calcium levels (McGourty *et al.*, 2016). This was the first known instance of calcium-dependent specific ubiquitination by an E3 ligase, although the related E3 ligases of the RSp5/NEDD4 family are thought to translocate in response to calcium (Wang *et al.*, 2010).

This work coincided with the independent discovery of peflin as a negative regulator of ER-to-Golgi transport in rat kidney cells (Rayl *et al.*, 2016). Calcium involvement in ER-to-Golgi transport is intricate, having many different roles in multiple different transport steps. Of relevance to peflin, it is thought that release of luminal calcium into the cytoplasm is involved in vesicle budding, docking or fusion. This increase in cytoplasmic calcium is presumed to cause dissociation of ALG-2 and peflin and encourage peflin localisation to membranes. This would remove the peflin-mediated inhibition from ALG-2 and allow it to interact with ER exit sites (ERES) and sec31A to promote ER-to-Golgi transport (reviewed, Rayl *et al.*, 2016).

Peflin has also recently been shown to be required for retrograde presynaptic homeostatic potentiation expression in postsynaptic neurons of the *Drosophila* neuromuscular junction (Kikuma *et al.*, 2019). The authors speculated that peflin might regulate postsynaptic compartment signalling, whereby peflin detects calcium fluctuations resulting from synaptic signalling and translate the calcium signal into regulation of CUL3. This would alter CUL3-mediated mono-ubiquitination of targets involved in synaptic signalling. Of course, this is all preliminary but makes for an interesting picture of the relationship between peflin and ubiquitin.

Whilst not much is known about peflin, it is clear that localisation and interaction with binding partners is tightly controlled by calcium, supporting the data presented here that calcium potentiated interaction with ubiquitin. Both calcium-control of ubiquitin micro-conformations and calcium-dependent peflin subcellular localisation (Kitaura *et al.*, 2001) regulate the peflin-ubiquitin interaction.

One important commonality between all the discussed studies on peflin and ubiquitin is that, whilst several papers have shown peflin to regulate ubiquitination and *vice versa*, none have

demonstrated direct interaction between the two proteins, which is shown for the first time here.

Overall it is difficult to conjecture about the physiological consequences of a peflin-ubiquitin interaction given the lack of knowledge around peflin; even those few publications that do exist tended to focus on interaction partners like ALG-2 or CUL3, with peflin as a side note. One clear model of calcium-mediated peflin interactions can be derived from the literature and data presented here, however. When in the absence of calcium, peflin dimerises with ALG-2. Upon an increase of calcium, ALG-2 dissociates to promote ER-to-Golgi transport (reviewed, Rayl *et al.*, 2016), leaving peflin free to associate with ubiquitin, which was shown here to be potentiated by the presence of calcium. This suggests that calcium signals act to switch peflin interaction from ALG-2 to ubiquitin.

If a direct interaction can be validated *in vivo*, then this could redefine the pathways discussed above. It could add an extra layer to the control of CUL3 activity, for example through peflin imparting a requirement for both high intracellular calcium and pre-existing target ubiquitination before triggering a CUL3-mediated ubiquitination feedback loop or cascade. It would be interesting to see how peflin-ALG-2 interaction affected ubiquitin-recognition by peflin, if a multiprotein complex was possible or if ALG-2 and ubiquitin bound competitively for peflin. If the latter was the case, ubiquitin recognition by peflin could preclude or dissociate the ALG-2-peflin interaction, releasing ALG-2 from inhibition and promoting vesicular formation and ER-to-Golgi transport. Alternately, as peflin could act as adapter during Histone 2AX regulation of DNA damage repair and cell cycle control, calcium signals could promote recruitment of both peflin and ubiquitin to the damage site.

6.4.2 Characterising the novel interaction between peflin and ubiquitin

The Western blot, LC-MS/MS and NMR data presented in this thesis described a direct interaction between ubiquitin and peflin for the first time. Whilst there was insufficient time to define the binding site with peflin, work was begun on characterising the binding site on ubiquitin. The effect of calcium on the peflin-ubiquitin interaction was also assessed, as peflin is a calcium-binding penta-EF hand protein, it has previously been associated with calcium-dependent modes of action, and many other ubiquitin-binding proteins require another inorganic divalent cation, zinc, for correct folding, function or ubiquitin interaction (Matthews and Sunde, 2012). The interaction appeared to be potentiated by calcium, although the data were made more complex by the unexpected observation that calcium affected ubiquitin in the absence of any other ligand.

Interaction between ubiquitin, peflin and calcium was observed by both affinity chromatography and protein NMR. The former suggested that calcium had a potentiating effect on the binding. NMR examined this more closely, demonstrating that the amplitude of residue CSP and number of affected residues increased when in the presence of calcium. Importantly, the shift was calculated relative to either ^{15}N-ubiquitin or ^{15}N-ubiquitin in the presence of calcium, as appropriate, to control for any effect of calcium on buffer pH on ubiquitin itself.

The data collected from affinity chromatography between peflin and mutant ubiquitin-Sepharose suggested that the interaction was dependent on Leu8 of ubiquitin, however this residue was only significant in NMR experiments in the absence of calcium. This may have been due to a change in presentation between immobilised ubiquitin-Sepharose and solution-based NMR, an effect of the GST tag in NMR studies, or it could be that the mutations surrounding the Leu8 had a broader impact on the binding surface beyond an isolated residue. To investigate this further, the complementary experiments should be conducted; affinity chromatography with mutant ubiquitin and GST-peflin, and NMR with cleaved peflin against

mutated ^{15}N-ubiquitin. Whilst the exact residues varied between experiments, both affinity chromatography and NMR confirmed that peflin bound ubiquitin across the canonical binding face.

Hydrophobic binding on ubiquitin is delineated by Leu8, Ile44, His68 and Val70 with an alternative hydrophobic patch around Phe4 and a hydrophilic patch around Asp58 (reviewed, Shaw *et al.*, 2011). Peflin was found to interact across the two hydrophobic patches but repeatedly showed no interaction with the hydrophilic patch. The residue that showed the strongest CSP in all conditions was His68, a member of the canonical site. This does not necessarily mean that His68 was critical for binding, however, merely that it was the most affected by the interaction. To complement the NMR, a repeat of the mutant ubiquitin affinity chromatography with peflin could be conducted with a H68A mutant. This would be a step towards dissecting the dramatic CSP seen at His68 during NMR.

Clearly further experiments are required to confirm the site of the peflin and ubiquitin interaction, particularly in relevance to the requirement for Leu8. Not many UBDs are known to have an absolute requirement for that residue; if it were confirmed as necessary for peflin and ubiquitin interaction then it could resemble the binding seen between UBM proteins and ubiquitin (Burschowsky *et al.*, 2011). The murine UBM-containing protein Pol iota interacted across a shifted site on ubiquitin surface, covering the canonical patch but moved slightly towards the C-terminal tail, where Leu8 locates. Similar to the affinity chromatography work shown here, L8A-ubiquitin nearly completely abrogated interaction with the UBM. Leu8 is in relative close proximity to His68 on the ubiquitin surface (approximately 8Å as measured by PyMOL with PDB 1UBQ (Vijay-Kumar *et al.*, 1987)), however Ile44 is located between the two residues. Ile44 showed significant involvement in the binding interact through NMR yet the I44A mutation had no effect on binding in affinity chromatography.

Whilst initial work was carried out to investigate the peflin binding patch on ubiquitin, the binding site within peflin still remains to be determined.

6.4.3 The novel ability of ubiquitin to act as a calcium sensor?

The identification of a novel ubiquitin-binding protein, peflin, with a putative novel UBD marginally contributes to balancing the discrepancy between the high numbers of E1/E2/E3 ligases, specific polyubiquitin modifications and ubiquitin-mediated processes and the comparatively few known UBDs. One individual UBD does not go far in addressing this discrepancy, however. This discovery that ubiquitin could act as a calcium-sensitive switch to alter its interaction with binding proteins is of far greater potential impact. If this switch is proven to be physiologically relevant then it could add an extra layer to ubiquitin-recognition, potentially allowing a single given UBD to alter affinity towards different ubiquitin-modifications depending on the local calcium concentration. To the author's knowledge, this is the first instance of ubiquitin recognition that is either mediated or directly regulated by calcium. This opens up an entirely new level of control over ubiquitin-mediated processes and potential crosstalk between ubiquitin and calcium signalling.

Calcium is one of the most widely used but functionally diverse secondary messengers both intra- and intercellularly in both excitable and non-excitable cells. Intracellular calcium concentrations can fluctuate rapidly over seconds or minutes, and can produce isolated spikes, train or waves of signalling through release or sequestration from the endoplasmic reticulum, mitochondria, cytosol and extracellular space (reviewed Heusinkveld and Westerink, 2011). Calcium is involved in a staggeringly diverse range of processes, including neuronal signalling, spatiotemporal control of gene expression, cytokine production in the immune system, thrombus formation, platelet activation, muscle contraction, muscle differentiation, cell migration, tumorigenesis and so on (reviewed, Berridge, 2014; Prakriya and Lewis, 2015;

Berridge, 2016). It seems reasonable to assume that every known eukaryotic biological process is regulated by either calcium or ubiquitin at some point. Areas of crossover or cooperation between the two systems are therefore highly interesting, potentially creating check points or localisation of signalling.

As well as the link between calcium binding proteins and CUL3 discussed previously, the E3 ligase NEDD4 can also bind calcium to remove NEDD4 autoinhibition and promote ubiquitination of activated CDC42-associated tyrosine kinase (ACK). This causes ACK to be degraded alongside the activated epidermal growth factor (EGF)-receptor (EGFR). As EGF has been shown to increase intracellular calcium concentrations, it is possible that this creates a negative feedback mechanism whereby EGF signalling via EGFR increases intracellular calcium, which in turn enhances NEDD4-mediated ACK and EGFR degradation and abrogation of the EGF signal. Crosstalk between ubiquitination and calcium has also been suggested to occur during mitophagy. In low calcium conditions the E3 ligase Mahogunin Ring Finger 1 (MGRN1) ubiquitinated another E3 ligase, GP78. This decreases GP78 interaction with mitochondria, presumably discouraging mitochondrial ubiquitination and mitophagy. It could be assumed then that an increase of intracellular calcium inhibits MGRN1 ubiquitination of GP78, promoting decoration of mitochondria with ubiquitin and promoting mitophagy, potentially as a cell protective mechanism (reviewed, Mukherjee *et al.*, 2017).

All these examples describe ubiquitin and calcium crossover at the E3 level, but this study is the first known instance of crossover at the level of ubiquitin-recognition. Interactome analysis highlighted that proteins that bound ubiquitin were more likely to be involved in the ‘response to inorganic substances’ than would be expected from a random protein sample. Initially this was unremarkable, but when combined with the later work on peflin and calcium this suggests an early indication or a generalised ability of ubiquitin to show differential response to the presence of calcium, as calcium is an inorganic substance.

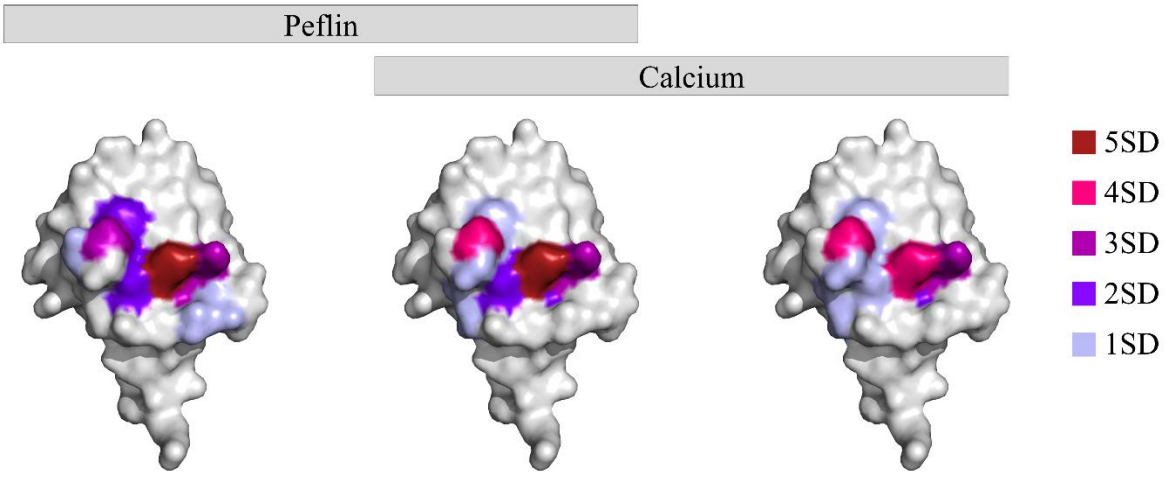
NMR spectra of ^{15}N-ubiquitin in the absence and presence of 250μM calcium, which is the physiological concentration of the ER membrane where peflin is thought to localise (Varadi and Rutter, 2002; Rayl *et al.*, 2016), indicated that ubiquitin responded to calcium even in the absence of a protein ligand. Nine residues were shown to be significantly perturbed when in the presence of calcium, suggesting that a minor structural alteration had occurred. This could represent a modulation of the binding surface through change of physicochemical properties or a micro-conformational change. It was interesting that the residues affected by physiological calcium levels all collected on the surface within close proximity to the canonical Ile44 binding patch. This may indicate that ubiquitin can act as a calcium-sensing switch, altering its binding face in response to the local microenvironment and affecting ubiquitin-recognition and downstream processes. Whilst the exact interplay between ^{15}N-ubiquitin, peflin and calcium is still unclear, three basic models were suggested;

- Competitive binding: peflin and calcium bind ubiquitin to the same site
- Independent binding: peflin and calcium bind ubiquitin at close but distinct sites
- Cooperative binding: calcium stabilises ubiquitin and peflin binding.

The first model could rationalise the very similar footprints of ^{15}N-ubiquitin CSP after exposure to GST-peflin or calcium. In this scenario, both peflin and calcium bind around the canonical patch with the strongest CSP at His68. The addition of calcium would cause ubiquitin-peflin to dissociate, presumably allowing peflin to form a homodimer as has been inferred from previous publications (Kitaura *et al.*, 2001; Kitaura *et al.*, 2002). This would leave the binding site on ubiquitin free for binding to calcium. The CSP footprint on ubiquitin would then be expected to match that seen with ^{15}N-ubiquitin + calcium in the absence of peflin: whilst the footprints were very similar, there was a lack of significant CSP at Gly47 (Fig.6.4.3.1). It could be that addition of calcium displaced some peflin but not all, creating two populations of ^{15}N-ubiquitin; that bound to peflin and that bound to calcium.

The second model suggests that peflin and calcium each bound ubiquitin independently at separate sites. In this manner ubiquitin would act as a calcium-sensitive switch whereby exposure to physiologically relevant calcium concentrations altered the ubiquitin binding patch, possibly through minor structural rearrangements, modulating the ubiquitin-peflin interaction without either directly precluding or potentiating it. This model could occur either through calcium causing minor surface alterations on ubiquitin or a minor structural change consequentially changing the binding surface.

The final model suggests that calcium is directly involved in peflin-ubiquitin binding, potentially mimicking the binding seen by several ZnF UBDs where zinc is located at the ubiquitin-binding site (for example, Kulathu *et al.* 2009). It could be that both peflin and calcium bind to ubiquitin in isolation, but in combination calcium acts to stabilise peflin binding, perhaps through direct interaction, or a minor conformational change around the canonical ubiquitin hydrophobic binding patch, or perhaps through a modulation of the patch by altering its physicochemical properties. This could explain the affinity chromatography results, which suggest that calcium is either absolutely required for or strongly potentiates binding of ubiquitin by peflin.



	CSP (SD from the mean)		
	Peflin	Peflin + Calcium	Calcium
Lys6	3	3	3
Thr7	1	-	-
Leu8	1	-	-
Leu43	1	-	-
Ile44	2	2	1
Phe45	2	1	1
Ala46	3	4	4
Gly47	-	1	1
Lys48	1		
Gln49	-	1	1
Leu67	1	2	1
His68	5	4	4
Leu69	3	2	2

Fig.6.4.3.1. The comparative GST-peflin- and calcium-interacting surfaces of ubiquitin. HSQC NMR spectra of 50µM ^{15}N-ubiquitin were captured in the presence of 400µM GST-peflin or 250µM $CaCl_2$ or both ligands, according to the top bars. The residues that showed significant CSP in the presence of ligand are highlighted according to their SD from the mean and are listed in the table below. Figure made with PyMOL using PDB 1UBQ (Vijay-Kumar *et al.*, 1987).

Of course, these three models could be further complicated by the separate action of calcium on peflin. As peflin is a known calcium-binding protein and as it is commonly found as a dimer (Kitaura *et al.*, 2001; Kitaura *et al.*, 2002) it may be that calcium interacts with both ^{15}N-ubiquitin and GST-peflin independently to modulate the binding sites of both proteins. Without further investigation it is currently impossible to tell which of these models, if any, are underpinning the CSP effects observed.

Whilst this is the first known example of calcium causing a response in ubiquitin, potentially impacting ubiquitin structure or dynamics, there has been substantial work on the interaction between ubiquitin and other divalent metal cations such as zinc or copper. Separately, there has been recent work on previously 'hidden' micro-conformational variants of ubiquitin.

Copper ions have been shown to bind across the Met1 and His68 residues, whereas zinc ions could bind across two sites; one broad site consisting of Met1, Glu16, Val17, Glu18, Ile23, Glu24, Lys27, Ala28, Gln49, Glu51, Asp52, Arg54 and Thr55 and the other centred around His68 (reviewed, Arena *et al.*, 2013).

Interaction between His68 ubiquitin and zinc or copper likely occurred through the amine of the histidine side chain, which extends out into the surrounding environment (Arena *et al.*, 2013), and calcium tends to bind negative amino acids (Asp and Glu) within EF hands (reviewed, Gifford *et al.*, 2007). These two modes of binding could possibly provide models for calcium binding to ubiquitin. These exact mechanisms are unlikely to occur during calcium-ubiquitin interaction, however. The second mode, modelled on calcium binding to EF hands, is unlikely to be replicated in the calcium-ubiquitin interaction as none of the nine residues with significant CSP in the presence of calcium were negatively charged. The first mode, modelled on zinc and copper binding is unlikely as calcium is coordinated by oxygen atoms rather than nitrogen, and the only free oxygen in His68 is packed inaccessibly close to the protein backbone. Of the other ^{15}N-ubiquitin residues that showed significant shift in the presence of calcium, only Ala46, Gly47 and Gln49 presented easily accessible oxygen atoms for calcium binding. These could represent the calcium binding sites, and the shift seen in other residues could have been consequential changes.

If this were the case, then calcium could bind one oxygen atom from ubiquitin Ala46, Gly47 or Gln49 and another from a calcium-binding protein, such as peflin. The lack of a published

peflin structure makes this difficult to model, but it could theoretically occur with any calcium-binding protein. This could suggest a mechanism for the calcium-induced potentiation of peflin binding to ubiquitin whereby calcium would act as a bridge to stabilise the existing ubiquitin-peflin interaction. This would contradict the peflin structure predictions (Fig.5.7.3) which suggested that the peptide ligand and calcium binding sites would be on opposite faces of ubiquitin. However, it must be remembered that the structure prediction was low confidence and may have incorrectly predicted the location of the binding sites either within the amino acid sequence or within the folded protein structure.

Alternatively, calcium could be acting to dimerise ^{15}N-ubiquitin. In this scenario, binding of another ubiquitin moiety could have caused the CSP seen by NMR, and would have resulted in displacement of the peflin molecule. There is precedence for divalent metal cations stabilising ubiquitin interactions in this manner, as zinc has been shown to bind two His68 residues from two ubiquitin molecules to promote ubiquitin aggregation. A double bridged form with two zinc ions binding both Glu64 and His68 was also reported (Fermani *et al.*, 2013).

1UBQPresence of 20 or 40μM Zn^{++} has been shown to prevent ubiquitin interaction with the E2 enzyme UBE2V2 and E2-25K and inhibited Lys48- and Lys63-polyubiquitination (Arena *et al.*, 2013). This ties in with the observation that zinc can bind a cysteine of the Fanconi anaemia ubiquitin-conjugating enzyme Ube2T and cause a major structural change and domain swap (Morreale *et al.*, 2017). Similar results were seen for interaction between ubiquitin, copper and E2 enzymes, although there was conflicting evidence over the success of this interaction within physiological pH ranges. This has been linked to accumulation of metal ions in the inclusion bodies of Parkinson's and Alzheimer's diseases (reviewed, Wojtunik-Kulesza *et al.*, 2019).

Given the reliance of normal biology on intracellular calcium stores and transient calcium signalling, this suggests that ubiquitin response to calcium could not only create an extra

mechanism for spatiotemporal control of ubiquitin recognition but could also contribute to pathologies underpinned by ubiquitin or calcium dysregulation.

If exposure to calcium had caused a micro-conformational change within ubiquitin, then this would build upon the recent publications on alternative ubiquitin micro-conformations. Interestingly, this also linked to the studies around how UBDs could selectively discriminate between ubiquitin and phosphoubiquitin.

Of particular relevance to the data presented here, both Dong *et al.* (2017) and Gladkova *et al.* (2017) recently identified an ubiquitin conformation that had previously been 'invisible' in standard biophysical experiments. This conformation was broadly similar to the common 'relaxed' ubiquitin conformation but the last β-strand had slipped by two positions, extending the normally well protected Ser65 loop out into the surrounding environment, and burying more of the C-terminal tail by shifting it up into the protein body. It was named Ub-CR reflecting the C-terminally retracted conformation. This shift was enabled by the alternate Leu pattern of S-T-L-H-L-V-L-R-L spanning positions 65 to 73 within the β-strand. Each of these leucines occupied what the authors called a 'Leu pocket'. In Ub-CR these leucines were shifted backwards by two amino acids to occupy the previous pockets, so that Leu73 occupied pocket 71, Leu71 occupied pocket 69, Leu69 occupied pocket 67, Leu67 occupied pocket 65, and the Ser65 was pushed out into a more solvent exposed position. (Gladkova *et al.*, 2017) This conformation was identified in both unphosphorylated and Ser65 phosphorylated forms (Dong *et al.*, 2017), although it was required for efficient interaction with and phosphorylation by PINK1 (Gladkova *et al.*, 2017).

Due to the altered pK_a of the phosphoryl group, pSer-Ub-CR was protonated and enriched at acidic pH. Under basic pH conditions, however, the deprotonated pSer-Ub was more frequently found in a conformation similar to the 'relaxed' state (Dong *et al.*, 2017; Kazansky *et al.*, 2018). This suggests that phosphoubiquitin could act as molecular pH

sensor and switch to promote minor conformation changes of ubiquitin and fine tune UBD recognition. Indeed, pH sensitivity was enhanced by double phosphorylation at both Ser65 and Thr66 by PINK1 (Ye *et al.*, 2019).

Further, the conformational change seen in Ub-CR also altered the Ile44 patch (Dong *et al.*, 2017; Kazansky *et al.*, 2018; Ye *et al.*, 2019), which drastically reduced the binding affinity of pUb-CR for the UBA of Ubiquilin-1. This suggested that the pUb-CR state may be less able to bind UBDs in general (Kazansky *et al.*, 2018). Whilst a physiological role for this has yet to be identified, cellular pH is known to change in cancerous cells, autophagy and aging and so represents a highly interesting opportunity for ubiquitin recognition research (Berezhnov *et al.*, 2016; White *et al.*, 2017). The existence of a previously undiscovered pH sensitive conformation that was physiologically relevant despite its low abundance provides precedence for the calcium-dependent modulation or micro-conformational changes presented here, and suggests that the micro-conformation could also be physiologically relevant regardless of its requirement for high calcium concentrations, which are transiently present or localised within the cell.

6.5 FUTURE WORK

Whilst the data presented here describe the creation of a useful resource for UBD annotation, the novel application of NGPD to target ubiquitin, the identification of a novel ubiquitin-binding protein with potential for a novel UBD, and the first known instance of ubiquitin interacting with calcium without any other mediator, these all remain preliminary results and further work will be required to confirm the observations seen here.

To start, whilst the ubiquitin-binding SLiM resource was comprehensive at the time of submitting this thesis (September 2019) it will need frequent updating as new domains and

structural knowledge are published. In order to simplify this and ensure reliability between updates, a standardised pipeline for UBD identification was described.

The application of NGPD against wildtype, phosphomimetic and phosphoubiquitin to identify artificial ubiquitin-binding peptides was highly promising, yet unfortunately did not pan out. The later work on ubiquitin micro-conformations and the Ub-CR conformation preferentially held by pSer65-ubiquitin (Gladkova *et al.*, 2017; Dong *et al.*, 2017) suggested that minor conformational differences in the presentation of ubiquitin between NGPD and NMR could explain the inability to identify strong interacting peptides. These may potentially be mitigated in repeat experiments by selection of a different immobilisation plate with a different method of presentation. Alternatively, use of polyubiquitin as the target protein or use of an antibody-based display system could help selection of specific binding peptides by enhancing avidity effects. The former would allow a single binding peptide to interact between multiple ubiquitin moieties or with the polyubiquitin linker region, and the latter would present a bipartite peptide library, potentially increasing affinity, survival and probability of successful binding-peptide identification. As a slightly different but no less interesting experiment, a cDNA library could be used in place of the randomised peptide library. This would present peptides based on a chosen genome (although coverage would be expected to be low) and would give a higher chance of identifying an endogenous as opposed to an artificial binding peptide.

In regards to the identification of peflin as a novel ubiquitin-binding protein, there are a wide variety of highly interesting experiments that could be conducted to further validate and characterise the peflin-ubiquitin interaction.

As a first instance, a repeat of the NMR experiments with thrombin cleaved peflin should be conducted. Initial attempts at large scale production of cleaved peflin proved inefficient but further optimisation of the purification or cleavage protocols would likely yield sufficient

material for NMR. GST-only controls had been conducted during affinity chromatography and Western blot experiments (data not shown) but were not conducted *via* NMR due to time constraints. Whilst it is unlikely that GST alone would interact with ^{15}N-ubiquitin *via* NMR, confirmation would increase the strength of the data.

The data investigating the peflin binding site on ubiquitin was inconclusive; affinity chromatography experiments suggested that Leu8 was a requirement for interaction, but this could not be validated through NMR in the presence of calcium. Additional NMR spectra with ^{15}N-L8A-ubiquitin and affinity chromatography with GST-peflin against mutant ubiquitin including H68A-ubiquitin would help to clarify this.

Whilst the absence of a known UBD was demonstrated for peflin, future experiments will be required to identify the binding site and putative novel UBD. This could initially by achieved by deletion mutants, potentially removing one EF hand or EF hand linker at a time and testing for preserved interaction with ubiquitin, although these deletions would have large scale impacts on peflin structure as a whole. Peptide arrays could prove a useful, if expensive approach. Using this method, the full length peflin sequence would be split into overlapping peptide fragments, which could be immobilised or printed on a plate surface, panned with ubiquitin and interaction confirmed by antibody detection of retained ubiquitin. Both of these approaches would narrow down the binding site but could not provide specific information. To accurately determine the boundaries of the domain and the interacting amino acids, carbene footprinting mass spectrometry could be used. This method uses irreversible and fast labelling of the target protein with reactive carbene, overcoming previous issues with poor resolution or low efficiency (Manzi *et al.*, 2016). Mass spectrometry is then used to map the location of the labelling on the protein surface, and this is compared between ligand bound and unbound states. Any or all of these suggested approaches would contribute significantly towards the

identification of the novel peflin ubiquitin-binding domain, which could then be confirmed by point mutations of the ubiquitin sequence and repeats of the experiments presented here.

Given that peflin is known to dimerise with ALG-2 and that the dimerization is regulated by both calcium and ubiquitin, an investigation into the relationship between all four interactors would help provide physiological context for the peflin-ubiquitin interaction. Not only would this determine if ubiquitin and ALG-2 could compete for peflin, but it would also significantly advance the current knowledge of peflin interaction in general, as it is a currently very poorly studied protein.

All experiments presented here made use of monoubiquitin, but given that peflin is known to dimerise, either with itself or with ALG-2, it could feasibly form a tandem UBD as described in Section 1.5. This would suggest the ability to bind polyubiquitin, which would be fairly quick to assess, for example through a repeat of the affinity chromatography experiments whereby peflin-Sepharose is used to capture polyubiquitin. This would not only determine any polyubiquitin length selectivity but could also be used to compare selectivity against different linkages. Further, repeats of these experiments in the absence and presence of calcium could assess if calcium affected peflin selectivity, as well as ubiquitin-binding and ubiquitin micro-conformations.

To determine if peflin and ubiquitin interact *in cellulo*, an initial microscopy experiment could assess co-localisation of ubiquitin and peflin in mammalian cell lines. There are several reliable commercially available antibodies for both peflin and ubiquitin that are suitable for immunocytochemistry. Endogenous peflin expression was confirmed in HEK 293T cells and was detectable by Western blot (data not shown), although levels were slightly low for immunocytochemistry so transfection would be required. If colocalisation was confirmed then

intracellular calcium release could be induced by pharmacological stimulation and confirm increased or prolonged colocalisation, or BAPTA-AM used to chelate cellular calcium.

Alternatively, further experimentation could be conducted to investigate the effect of calcium on ubiquitin micro-conformations. Initial experiments could include simple NMR titrations of ^{15}N-ubiquitin with different physiological concentrations of calcium to determine any concentration dependence of the effect. This could also begin to suggest a physiological role for the calcium-induced micro-conformational changes, as calcium concentrations are known to be highly variable yet tightly controlled across the different cell compartments.

The NMR data presented in section 5.8 identified effects of calcium on ubiquitin structure at the amino acid level. To determine if this had a broader effect on ubiquitin structure, a quick circular dichroism experiment could be conducted. By determining the overall secondary structure in the presence and absence of calcium, it could be determined if calcium caused a localised or generalised change in ubiquitin structure. This could help explain how some residues were seen by NMR to shift significantly in the presence of calcium despite not containing an easily accessible oxygen atom for calcium binding.

Finally, whilst it did not appear that calcium had any effect on GST-p62 interaction with ubiquitin, as suggested by affinity chromatography (Fig.5.6.2), additional affinity chromatography or NMR of ubiquitin with other known UBDs in the presence of calcium would begin to suggest if the calcium-potentiation was specific to peflin or if it is a widespread mechanism for ubiquitin-recognition regulation. The LC-MS/MS identification of endogenous proteins that interacted with ubiquitin-Sepharose supported the idea that calcium could affect more ubiquitin interactions beyond peflin (Section 5.9), so this could be a highly informative experiment.

The data presented here intriguingly point towards peflin as a novel ubiquitin-binding protein and ubiquitin itself as able to respond to physiological concentrations of calcium. Despite this encouraging start, there remain many exciting experiments needed to fully characterise these interactions.

6.6 CONCLUSIONS

The central hypothesis of this thesis, that as of yet undiscovered families of UBDs and aspects of specific ubiquitin-recognition exist as a form of ubiquitination 'dark matter', was examined through a combination of approaches.

Many smaller observations were made along the way, such as the current state of UBD bioinformatics and the practicality of applying NGPD to ubiquitin targets, but the majority of the work presented here focused on the identification of a novel ubiquitin-binding protein, peflin with a presumed novel form of UBD, and the suggestion of a novel reaction of ubiquitin to physiologically relevant calcium concentrations.

To the author's knowledge this was the first indication of a direct interaction between peflin and ubiquitin. Although they have previously been linked in a couple of different contexts, previous work has focused on the interaction between peflin and the related ALG-2 or the E3 ligase CUL3. As such this is the first instance of peflin and ubiquitin interaction at the level of ubiquitin-recognition rather than ubiquitination. Several physiological contexts were suggested for this but so little is known about peflin that it is difficult to speculate. Further investigation into the UBD of peflin will not only expand the small existing knowledge base around peflin but will also take a step towards reducing the knowledge mismatch between specific ubiquitination and ubiquitin recognition.

This work also presents the first indication that ubiquitin can interact with calcium directly, and that the interaction causes a minor binding or structural change within ubiquitin itself.

Precedent has been set for this, as ubiquitin is known to interact with other divalent metal cations such as copper and zinc. The consequences of such an interaction are not yet known, but it is feasible that the modulation or micro-conformational change would affect ubiquitin-recognition, possibly contributing to the fine control over ubiquitin-mediated processes. Intriguingly, this suggested that ubiquitin could act to integrate calcium signalling with ubiquitination or could act as a calcium-sensitive switch. With such a wide variety of both ubiquitin- and calcium-mediated processes, the physiological relevance of this is still highly speculative. It is still too early to suggest with any accuracy, but it could feasibly be related to signal transduction from calcium-permeable membrane channels, or to ER-to-Golgi transport, for example in cargo transport or neurotransmitter vesicle release. Both processes are central to all aspects of cell activity, however, that this integration could cause a major change in the way we understand cell signalling as a whole.

As such, the data presented here further our knowledge of ubiquitin recognition on several levels, not only addressing the hypothesis and aims set out at the start of this thesis but also opening up new areas of ubiquitin-recognition research for future investigation. It is the hope of this author that the concepts described here are received with as much enthusiasm as was involved in their investigation, and that this thesis will be of interest to other ubiquitin enthusiasts.

7 REFERENCES

Abedon, S. T., 2015. Phage therapy of pulmonary infections. *Bacteriophage*: **5 (1)**: e1020260

Ackermann, H.W., Tremblay, D. and Moineau, S., 2004. Long-term bacteriophage preservation. *WFCC Newsletter*: **38 (1)**: 35-40

Agromayor, M., Soler, N., Caballe, A., Kueck, T., Freund, S. M., Allen, M. D., Bycroft, M., Perisic, O., Ye, Y., McDonald, B., Scheel, H., Hofmann, K., Neil, S. J. D., Martin-Serrano, J. and Williams, R. L., 2012. The UBAP1 subunit of ESCRT-I interacts with ubiquitin via a SOUBA domain. *Structure*: **20 (3)**: 414-428

Alam, S. L., Sun, J., Payne, M., Welch, B. D., Blake, B. K., Davis, D. R., Meyer, H. H., Emr, S. D. and Sundquist, W. I., 2004. Ubiquitin interactions of NZF zinc fingers. *The EMBO Journal*: **23 (7)**: 1411-1421

Allen, M. D. and Bycroft, M., 2007. The solution structure of the ZnF UBP domain of USP33/VDU1. *Protein Science*: **16 (9)**: 2072-2075

Anamika, Markin, C. J., Rout, M. K., Spyracopoulos, L., 2014. Molecular basis for impaired DNA damage response function associated with the RAP80 ΔE81 defect. *The Journal of Biological Chemistry*: **289 (18)**: 12852-12862

Anindya, R., Mari, P. O., Kristensen, U., Kool, H., Giglia-Mari, G., Mullenders, L. H., Fousteri, M., Vermeulen, W., Egly, J. M. and Svejstrup, J. Q., 2010. A ubiquitin-binding domain in Cockayne syndrome B required for transcription-coupled nucleotide excision repair. *Molecular Cell*: **38 (5)**: 637-648

Arena, G., Bellia, F., Frasca, G., Grasso, G., Lanza, V., Rizzarelli, E., Tabbì, G., Zito, V. and Milardi, D., 2013. Inorganic stressors of ubiquitin. *Inorganic Chemistry*: **52 (16)**: 9567-9573

Au, J. S. Y., Puri, C., Ihrke, G., Kendrick-Jones, J. and Buss, F., 2007. Myosin VI is required for sorting of AP-1B–dependent cargo to the basolateral domain in polarized MDCK cells. *The Journal of Cell Biology:* **177 (1)**: 103-114

Baek, H., Suk, K.H., Kim, Y.H. and Cha, S., 2002. An improved helper phage system for efficient isolation of specific antibody molecules in phage display. *Nucleic Acids Research*: **30 (5)**: e18-e18

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W. and Noble, W.S., 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research*: **37 (2)**: W202-W208

Baker, R. T., Tobias, J. W. and Varshavsky, A., 1992. Ubiquitin-specific proteases of Saccharomyces cerevisiae. Cloning of UBP2 and UBP3, and functional analysis of the UBP gene family. *Journal of Biological Chemistry*: **267 (32)**: 23364-23375

Balakirev, M.Y., Tcherniuk, S.O., Jaquinod, M. and Chroboczek, J., 2003. Otubains: a new family of cysteine proteases in the ubiquitin pathway. *EMBO Reports*: **4 (5)**: 517-522

Baldwin, D. and Summer, N.S., Phage Biocontrol Research LLC, 2012. *Prevention and remediation of petroleum reservoir souring and corrosion by treatment with virulent bacteriophage*. U.S. Patent 8,168,419

Barbas, C. F., 3rd; Burton, D. R.; Scott, J. K.; Silverman, G. J. Phage Display: A Laboratory Manual; *Cold Spring Harbor Laboratory Press*: Cold Spring Harbor, NY, 2001.

Bartolacci, C., Andreani, C., Curcio, C., Occhipinti, S., Massaccesi, L., Giovarelli, M., Galeazzi, R., Iezzi, M., Tilio, M., Gambini, V., Wang, J., Marchini, C., and Amici, A., 2018. Phage-based anti-HER2 vaccination can circumvent immune tolerance against breast cancer. *Cancer Immunology Research*: **6 (12)**: 1486-1498

Batenburg, N. L., Walker, J. R., Noordermeer, S. M., Moatti, N., Durocher, D. and Zhu, X. D., 2017. ATM and CDK2 control chromatin remodeler CSB to inhibit RIF1 in DSB repair pathway choice. *Nature Communications*: **8 (1)**: 1921

Bauer, M., Flatt, JW., Seiler, D., Cardel, B., Emmerlauer, M., Boucke, K., Suomalainen, M., Hemmi, S. and Greber, UF., 2019. The E£ Ubiquitin Ligase Mind Bome 1 Controls Adenovirus Genome Release at the Nuclear Pore Complex. *Cell Reports*: **29 (12)**: 3785-3795

Baumgärtel, V., Ivanchenko, S., Dupont, A., Sergeev, M., Wiseman, P. W., Kräusslich, H. G., Bräuchle, C., Müller, B. and Lamb, D. C., 2011. Live-cell visualization of dynamics of HIV budding site interactions with an ESCRT component. *Nature Cell Biology*: **13 (4)**: 469

Bean, D. M., Heimbach, J., Ficorella, L., Micklem, G., Oliver, S. G., and Favrin, G., 2014. esyN: Network Building, Sharing and Publishing. *PLoS ONE*: **9 (9)**: e106035

Berezhnov, A.V., Soutar, M.P., Fedotova, E.I., Frolova, M.S., Plun-Favreau, H., Zinchenko, V.P. and Abramov, A.Y., 2016. Intracellular pH modulates autophagy and mitophagy. *Journal of Biological Chemistry*: **291 (16)**: 8701-8708

Bergh, O., Borsheim, K. Y., Bratbak, G., and Heldal, M., 1989. High abundance of viruses found in aquatic environments. *Nature*: **340 (6233)**: 467-468

Berridge, M.J., 2014. Calcium signalling and psychiatric disease: bipolar disorder and schizophrenia. *Cell and Tissue Research*: **357 (2)**: 477-492

Berridge, M.J., 2016. The inositol trisphosphate/calcium signaling pathway in health and disease. *Physiological Reviews*: **96 (4)**: 1261-1296

Bian, Y., Song, C., Cheng, K., Dong, M., Wang, F., Huang, J., Sun, D., Wang, L., Ye, M. and Zou, H., 2014. An enzyme assisted RP-RPLC approach for in-depth analysis of human liver phosphoproteome. *Journal of Proteomics*: **96**: 253-262

Bienko, M., Green, C. M., Crosetto, N., Rudolf, F., Zapart, G., Coull, B., Kannouche, P., Wider, G., Peter, M., Lehmann, A. R. and Hofmann, K., 2005. Ubiquitin-binding domains in Y-family polymerases regulate translesion synthesis. *Science*: **310 (5755)**: 1821-1824

Bilodeau, P. S., Winistorfer, S. C., Allaman, M. M., Surendhran, K., Kearney, W. R., Robertson, A. D. and Piper, R. C., 2004. The GAT domains of clathrin-associated GGA proteins have two ubiquitin binding motifs. *Journal of Biological Chemistry*: **279 (52)**: 54808-54816

Bloom, J., Amador, V., Bartolini, F., DeMartino, G. and Pagano, M., 2003. Proteasome-mediated degradation of p21 via N-terminal ubiquitinylation. *Cell*: **115 (1)**: 71-82

Bomar, M. G., D'Souza, S., Bienko, M., Dikic, I., Walker, G. C. and Zhou, P., 2010. Unconventional ubiquitin recognition by the ubiquitin-binding motif within the Y family DNA polymerases ι and Rev1. *Molecular Cell*: **37 (3)**: 408-417

Bomar, M.G., Pai, M.T., Tzeng, S.R., Li, S.S.C. and Zhou, P., 2007. Structure of the ubiquitin-binding zinc finger domain of human DNA Y-polymerase η. *EMBO Reports*: **8 (3)**: 247-251

Bonnet, J., Romier, C., Tora, L. and Devys, D., 2008. Zinc-finger UBPs: regulators of deubiquitylation. *Trends in biochemical sciences*, *33*(8), pp.369-375

Bremm, A., Freund, S. M. and Komander, D., 2010. Lys11-linked ubiquitin chains adopt compact conformations and are preferentially hydrolyzed by the deubiquitinase Cezanne. *Nature Structural & Molecular Biology*: **17 (8)**: 939

Brinton, L. T., Bauknight, D. K., Dasa, S. S., and Kelly, K. A., 2016. PHASTpep: Analysis Software for Discovery of Cell-Selective Peptides via Phage Display and Next-Generation Sequencing. *PLoS One*: **11 (5)**: e01555244

Brussow, H., and Hendrix, R. W., 2002. Phage genomics: small is beautiful. *Cell*: **180 (1)**: 13-16

Brussow, H., and Kitter, E., 2005. Phage ecology. In: Kutter, E., Sulakvelidze, A. (Eds.), *Bacteriophages: Biology and Application*, **1st ed**. CRC Press, Boca Raton FL, USA: 129-163

Brzovic, P.S., Lissounov, A., Christensen, D.E., Hoyt, D.W. and Klevit, R.E., 2006. A UbcH5/ubiquitin noncovalent complex is required for processive BRCA1-directed ubiquitination. *Molecular Cell:* **21 (6)**: 873-880

Bunn, R. C., Jensen, M. A. and Reed, B. C., 1999. Protein interactions with the glucose transporter binding protein GLUT1CBP that provide a link between GLUT1 and the cytoskeleton. *Molecular Biology of the Cell:* **10 (4)**: 819-832

Burrows, J.F., McGrattan, M.J. and Johnston, J.A., 2005. The DUB/USP17 deubiquitinating enzymes, a multigene family within a tandemly repeated sequence. *Genomics*: **85 (4)**: 524-529

Burschowsky, D., Rudolf, F., Rabut, G., Herrmann, T., Matthias, P. and Wider, G., 2011. Structural analysis of the conserved ubiquitin-binding motifs (UBMs) of the translesion polymerase iota in complex with ubiquitin. *Journal of Biological Chemistry*: **286 (2)**: 1364-1373

Buss, F., Arden, S. D., Lindsay, M., Luzio, J. P. and Kendrick-Jones, J., 2001. Myosin VI isoform localized to clathrin-coated vesicles with a role in clathrin-mediated endocytosis. *The EMBO Journal*: **20 (14)**: 3676-3684

Buss, F., Spudich, G. and Kendrick-Jones, J., 2004. Myosin VI: cellular functions and motor properties. *Annual Review of Cell and Developmental Biology:* **20***:* 649-676

Cabilly, S., 1999. The basic structure of filamentous phage and its use in the display of combinatorial peptide libraries. *Molecular Biotechnology*: **12 (2)**: 143-148

Cairns, B.J., Timms, A.R., Jansen, V.A., Connerton, I.F. and Payne, R.J., 2009. Quantitative models of in vitro bacteriophage–host dynamics and their application to phage therapy. *PLoS Pathogens*: **5 (1)**: 1000253

Cao, Z., Xiong, J., Takeuchi, M., Kurama, T. and Goeddel, DV., 1996. TRAF6 is a signal transducer for interleukin-1. *Nature*: **383 (6599)**: 444-446

Carmen, S. and Jermutus, L., 2002. Concepts in antibody phage display. *Briefings in Functional Genomics*: **1 (2)**: 189-203

Castagnoli, L., Mandaliti, W., Nepravishta, R., Valentini, E., Mattioni, A., Procopio, R., Iannuccelli, M., Polo, S., Paci, M., Cesareni, G. and Santonico, E., 2019. Selectivity of the

CUBAN domain in the recognition of ubiquitin and NEDD8. *The FEBS Journal*: **286 (4)**: 653-677

Chau, V., Tobias, J.W., Bachmair, A., Marriott, D., Ecker, D.J., Gonda, D. K., and Varshavasky, A., 1989. A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. *Science*: **243 (4898)**: 1576-1583

Chaugule, V.K., Burchell, L., Barber, K.R., Sidhu, A., Leslie, S.J., Shaw, G.S. and Walden, H., 2011. Autoregulation of Parkin activity through its ubiquitin-like domain. *The EMBO Journal*: **30 (14)**: 2853-2867

Chen, T., Zhou, T., He, B., Yu, H., Guo, X., Song, X, and Sha, J., 2014. mUbiSiDa: a comprehensive database for protein ubiquitination sites in mammals. *PLoS One*: **9 (1)**: e85744

Chen, Z., Hagler, J., Palombella, V.J., Melandri, F., Scherer, D., Ballard, D. and Maniatis, T., 1995. Signal-induced site-specific phosphorylation targets I kappa B alpha to the ubiquitin-proteasome pathway. *Genes and Development*: **9 (13)**: 1586-1597

Chibani-Chennouf, S., Bruttin, A., Dillmann, M. L., and Brüssow, H., 2004. Phage-host interaction: an ecological perspective. *Journal of Bacteriology*: **186 (12)**: 3677-3686

Chitale, S. and Richly, H., 2017. Nuclear organization of nucleotide excision repair is mediated by RING1B dependent H2A-ubiquitylation. *Oncotarget*: **8 (19)**: 30870

Ciani, B., Layfield, R., Cavey, J.R., Sheppard, P.W. and Searle, M.S., 2003. Structure of the ubiquitin-associated domain of p62 (SQSTM1) and implications for mutations that cause Paget's disease of bone. *Journal of Biological Chemistry*: **278 (39)**: 37409-37412

Clerici, M., Luna-Vargas, M. P., Faesen, A. C. and Sixma, T. K., 2014. The DUSP–Ubl domain of USP4 enhances its catalytic efficiency by promoting ubiquitin exchange. *Nature Communications*: **5:** 5399

Collins, B. M., Watson, P. J. and Owen, D. J., 2003. The structure of the GGA1-GAT domain reveals the molecular basis for ARF binding and membrane association of GGAs. *Developmental Cell*: **4 (3)**: 321-332

Conze, D. B., Wu, C. J., Thomas, J. A., Landstrom, A., and Ashwell, J. D., 2008. Lys63-linked polyubiquitination of IRAK-1 is required for interleukin-1 receptor- and toll-like receptormediated nf-kappab activation. *Molecular Cell Biology*: **28**: 3538-3547

Cook, J. C. and Chock, P. B., 1995. Phosphorylation of ubiquitin-activating enzyme in cultured cells. *Proceedings of the National Academy of Sciences*: **92 (8)**: 3454-3457

Cordier, F., Vinolo, E., Véron, M., Delepierre, M. and Agou, F., 2008. Solution structure of NEMO zinc finger and impact of an anhidrotic ectodermal dysplasia with immunodeficiency-related point mutation. *Journal of Molecular Biology:* **377 (5)**: 1419-1432

Cortese, R., Felici, F., Galfre, G., Luzzago, A., Monaci, P., and Nicosia, A., 1994. Epitope discovery using peptide libraries displayed on phage. *Trends in Biotechnology*: **12**: 262-267

Cortese, R., Monaci, P, Luzzago, A., Santini, C., Bartoli, F., Cartese, I., Fortungo, P., Galfre, G., Nicosia, A., and Felici, F., 1996. Selection of biologically active peptides by phage display of random peptide libraries. *Current Opinions in Biotechnology*: **7**: 616-621

Courtois, G. and Fauvarque, M.O., 2018. The many roles of ubiquitin in NF-κB signaling. *Biomedicines*: **6 (2)**: E43

Dantuma, N. P., Heinen, C. and Hoogstraten, D., 2009. The ubiquitin receptor Rad23: at the crossroads of nucleotide excision repair and proteasomal degradation. *DNA Repair*: **8 (4)**: 449-460

Dawson, T.M., and Dawson, V.L., 2010. The role of parkin in familial and sporadic Parkinson's disease. *Movement Disorders*: **25 (S1)**: S32–39

De Biévre, P. and Taylor, P. D. P. 1993. Table of the isotopic compositions of the elements. *International Journal of Mass Spectrometry and Ion Processes*: **123 (2)**: 149-166

Den Besten, W., Verma, R., Kleiger, G., Oania, R. S. and Deshaies, R. J., 2012. NEDD8 links cullin-RING ubiquitin ligase function to the p97 pathway. *Nature Structural & Molecular Biology*: **19 (5)**: 511

Deng, L., Wang, C., Spencer, E., Yang, L., Braun, A., You, J., Slaughter, C., Pickart, C. and Chen, Z.J., 2000. Activation of the IκB kinase complex by TRAF6 requires a dimeric ubiquitin-conjugating enzyme complex and a unique polyubiquitin chain. *Cell*: **103 (2)**: 351-361

Deng, Y. M., Liu, C. Q. and Dunn, N. W., 1999. Genetic organization and functional analysis of a novel phage abortive infection system, AbiL, from Lactococcus lactis. *Journal of Biotechnology*: **67 (2-3)**: 135-149

Di Rienzo, M., Antonioli, M., Fusco, C., Liu, Y., Mari, M., Orhon, I., Refolo, G., Germani, F., Corazzari, M., Romagnoli, A., Ciccosanti, F., Mandriani, B., Pellico, M.T., De La Torre, R., Ding, H., Dentice, M., Neri, M., Ferlini, A., Reggiori, F., Kulesz-Martin, M., Piacentini, M., Merla, G., and Fimia, G.M., 2019. Autophagy induction in atrophic muscle cells requires ULK1 activation by TRIM32 through unanchored K63-linked polyubiquitin chains. *Science Advances:* **5 (5)**: eaau8857

Dias-Neto, E., Nunes, D. N., Giordano R. J., Sun, J., Botz, G. H., Yang, K., Setubal, J. C., Pasqualini, R. and Arap, W., 2009. Next-generation phage display: integrating and comparing available molecular tools to enable cost-effective high-throughput analysis. *PLoS One*: 4**:** e8338

DiDonato, J., Mercurio, F., Rosette, C., Wu-Li, J., Suyang, H., Ghosh, S. and Karin, M., 1996. Mapping of the inducible IkappaB phosphorylation sites that signal its ubiquitination and degradation. *Molecular and Cellular Biology*: **16 (4)**: 1295-1304

Dieckmann, T., Withers-Ward, E. S., Jarosinski, M. A., Liu, C. F., Chen, I. S. and Feigon, J., 1998. Structure of a human DNA repair protein UBA domain that interacts with HIV-1 Vpr. *Nature Structural & Molecular Biology*: **5 (12)**: 1042

Dinarello, CA., 2009. Immunological and inflammatory functions of the interleukin-1 family. *Annual Reviews of Immunology*: **27**: 519-550

Ding, D., Bergmaier, P., Sachs, P., Klangwart, M., Rückert, T., Bartels, N., Demmers, J., Dekker, M., Poot, R.A., and Mermoud, J.E., 2018. The CUE1 domain of the SNF2-like chromatin remodeler SMARCAD1 mediates its association with KRAB-associated protein 1 (KAP1) and KAP1 target genes. *Journal of Biological Chemistry:* **293 (8)**: 2711-2724

Dong, X., Gong, Z., Lu, Y.B., Liu, K., Qin, L.Y., Ran, M.L., Zhang, C-L., Liu, Z., Zhang, W-P. and Tang, C., 2017. Ubiquitin S65 phosphorylation engenders a pH-sensitive conformational switch. *Proceedings of the National Academy of Sciences*: **114 (26)**: 6770-6775

Dueñas, M. and Borrebaeck, C.A., 1995. Novel helper phage design: intergenic region affects the assembly of bacteriophages and the size of antibody libraries. *FEMS Microbiology Letters*: **125 (2-3)**: 317-321

Ea, C.K., Deng, L., Xia, Z.P., Pineda, G. and Chen, Z.J., 2006. Activation of IKK by TNFα requires site-specific ubiquitination of RIP1 and polyubiquitin binding by NEMO. *Molecular Cell*: **22 (2)**: 245-257

Edelmann, M. J., Iphöfer, A., Akutsu, M., Altun, M., Di Gleria, K., Kramer, H. B., Fiebiger, E., Dhe-Paganon, S. and Kessler, B. M., 2009. Structural basis and specificity of human otubain 1-mediated deubiquitination. *Biochemical Journal*: **418 (2)**: 379-390

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*: **32 (5)**: 1792-1797

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., Qureshi, M., Richardson, L. J., Salazar, G. A., Smart, A., Sonnhammer, E. L. L., Hirsh, L., Paladin, L., Piovesan, D., Tosatto, S. C. E. and Finn, R. D., 2019. The Pfam protein families database in 2019. *Nucleic Acids Research*: **47 (D1)**: D427-D432

Emmerich, CH., Ordureau, A., Strickson, S., Arthur, JSC., Pedrioli, PGA., Komander, D. and Cohen, P., 2013. Activation of the canonical IKK complex by K63/M1-linked hybrid ubiquitin chains. *PNAS*: **110 (38)**: 15247-15452

Enshell-Seijffers, D., Smelyanski, L., Vardinon, N., Yust, I. and Gershoni, J. M., 2001. Dissection of the humoral immune response towards an immunodominant epitope of HIV: a model for the analysis of antibody diversity in HIV+ individuals. *FASEB Journal*: **15**: 2112-2120

Erba, E. B. and Zenobi, R., 2011. *Annual Reports Physical Chemistry "C"*: **107**: 199-228

Ernst, A.., Avvakumov, G., Tong, J., Fan, Y., Zhao, Y., Alberts, P., Persaud, A., Walker, J. R., Neculai, D., Vorobyov, A., Garg, P., Beatty, L., Chan, P-K., Juang, Y-C., Landry, M-C., Yeh, C., Zeqiraj, E., Karamboulas, K., Allai-Hassani, A., Vedadi, M., Tyers, M., Moffat, J., Sicheri, F., Pelletier, L., Durocher, D., Raught, B., Rotin, D., Yang, J., Moran, M. F., Dhe-Paganon, S. and Sidhu, S. S., 2013. A Strategy for Modulation of Enzymes in the Ubiquitin System. *Science*: **339 (6119)**: 590-595

Fellhouse, F. A. and Pal, G., 2005. Methods for the Construction of Phage-Displayed Libraries. In: Sidhu, S. S. and Geyer, C. R., eds., 2005. *Phage Display in Biotechnology and Drug Discovery*. Taylor and Francis Group, Boca Raton, Florida. **Vol. 3**: 111-142

Fenner, B. J., Scannell, M. and Prehn, J. H., 2009. Identification of polyubiquitin binding proteins involved in NF-κB signaling using protein arrays. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*: **1794 (7)**: 1010-1016

Fermani, S., Falini, G., Calvaresi, M., Bottoni, A., Calò, V., Mangini, V., Arnesano, F. and Natile, G., 2013. Conformational Selection of Ubiquitin Quaternary Structures Driven by Zinc Ions. *Chemistry–A European Journal*: **19 (46)**: 15480-15484

Ferrage, F., Dutta, K., Nistal-Villán, E., Patel, J. R., Sánchez-Aparicio, M. T., De Ioannes, P., Buku, A., Aseguinolaza, G. G., García-Sastre, A. and Aggarwal, A. K., 2012. Structure and dynamics of the second CARD of human RIG-I provide mechanistic insights into regulation of RIG-I activation. *Structure*: **20 (12)**: 2048-2061

Fisher, R.D., Wang, B., Alam, S.L., Higginson, D.S., Robinson, H., Sundquist, W.I. and Hill, C.P., 2003. Structure and ubiquitin binding of the ubiquitin-interacting motif. *Journal of Biological Chemistry*: **278 (31)**: 28976-28984

Fletcher, A.J., Vaysburd, M., Maslen, S., Zeng, J., Skehel, J.M., Towers, G.J. and James, L.C., 2018. Trivalent RING assembly on retroviral capsids activates TRIM5 ubiquitination and innate immune signaling. *Cell Host and Microbe*: **24 (6)**: 761-775

Folgori, A., Tafi, R., Meola, A., Felici, F., Galfre, G., Cortese, R., Monaci, O., and Nicosia, A., 1994. A general strategy to identify mimotopes of pathological antigens using only random peptide libraries and human sera. *EMBO Journal*: **13**: 2236-2243

Freeman, A.A., Mandilaras, K., Missirlis, F. and Sanyal, S., 2013. An emerging role for Cullin-3 mediated ubiquitination in sleep and circadian rhythm: insights from Drosophila. *Fly*: **7 (1)**: 39-43

Fu, Q. S., Zhou, C. J., Gao, H. C., Jiang, Y. J., Zhou, Z. R., Hong, J., Yao, W. M., Song, A. X., Lin, D. H. and Hu, H. Y., 2009. Structural basis for ubiquitin recognition by a novel domain from Human Phospholipase A2Activtating Protein. *The Journal of Biological Chemistry*: **284**: 19043-19052

Gallo, L. H., Meyer, A. N., Motamedchaboki, K., Nelson, K. N., Haas, M. and Donoghue, D. J., 2014. Novel Lys63-linked ubiquitination of IKKβ induces STAT3 signaling. *Cell Cycle*: **13 (24)**: 3964-3976

Garner, T. P., Strachan, J., Shedden, E. C., Long, J. E., Cavey, J. R., Shaw, B., Layfield, R. and Searle, M. S., 2011. Independent Interactions of Ubiquitin-Binding Domains in a Ubiquitin-Mediated Ternary Complex. *Biochemistry*: **50**: 9076-9087

Georgieva, Y. and Konthur, Z., 2011. Design and Screening of M13 Phage Display cDNA Libraries. *Molecules*: **16**: 1667-1681

Gerlach, B., Cordier, S.M., Schmukle, A.C., Emmerich, C.H., Rieser, E., Haas, T.L., Webb, A.I., Rickard, J.A., Anderton, H., Wong, W.W.L. and Nachbur, U., 2011. Linear ubiquitination prevents inflammation and regulates immune signalling. *Nature*: **471 (7340)**: 591-596

Germaschewski, V., and Murray, K., 1996. Identification of polyclonal serum specificities with phage-display libraries. *Journal of Virology Methods*: **58** 21-32

Gibbons, JM., Marno, KM., Pike, R., Lee, WJ., Jones, CE., Ogunkolade, BW., Pardieu, C., Bryan, A., Fu, RM., Warnes, G., Rowley, PA., Sloan, RD. and McKnight, A., 2020. HIV-1 Accessory Protein Vpr Interacts with REAF/RPRD2 To Mitigate Its Antiviral Activity. *Journal of Virology*: **94 (4)**: e01591-e01619

Gifford, J.L., Walsh, M.P. and Vogel, H.J., 2007. Structures and metal-ion-binding properties of the Ca2+-binding helix–loop–helix EF-hand motifs. *Biochemical Journal*: **405 (2)**: 199-221

Gilberto, S. and Peter, M., 2017. Dynamic ubiquitin signaling in cell cycle regulation. *Journal of Cell Biology*: **216 (8)**: 2259-2271

Gladkova, C., Schubert, A.F., Wagstaff, J.L., Pruneda, J.N., Freund, S.M. and Komander, D., 2017. An invisible ubiquitin conformation is required for efficient phosphorylation by PINK1. *The EMBO Journal*: **36 (24)**: 3555-3572

Goldstein, G., Scheid, M., Hammerling, U., Schlesinger, D. H., Niall, H. D. and Boyse, E. A., 1975. Isolation of a polypeptide that has lymphocyte-differentiating properties and is probably

represented universally in living cells. *Proceedings of the National Academy of Sciences*: **72 (1)**: 11-15

Goode, A., Butler, K., Long, J., Cavey, J., Scott, D., Shaw, B., Sollenberger, J., Gell, C., Johansen, T., Oldham, N.J., Searle, M.S., and Layfield, R., 2016. Defective recognition of LC3B by mutant SQSTM1/p62 implicates impairment of autophagy as a pathogenic mechanism in ALS-FTLD. *Autophagy*: **12 (7)**: 1094-1104

Gorelik, M., Manczyk, N., Pavlenco, A., Kurinov, I., Sidhu, S.S. and Sicheri, F., 2018. A Structure-Based Strategy for Engineering Selective Ubiquitin Variant Inhibitors of Skp1-Cul1-F-Box Ubiquitin Ligases. *Structure*: **26 (9)**: 1226-1236

Gracheva, E., Chitale, S., Wilhelm, T., Rapp, A., Byrne, J., Stadler, J., Medina, R., Cardoso, M. C. and Richly, H., 2016. ZRF1 mediates remodeling of E3 ligases at DNA lesion sites during nucleotide excision repair. *Journal of Cell Biology:* **213 (2)**: 185-200

Greenwood, J., Willis, A. E., and Perham, R. N., 1991. Multiple display of forgein peptides on a filamentous bacteriophage: peptides from Plasmodium falciparum circumsporozoite protein as antigens. *Journal of Molecular Biology*: **220 (4)**: 821-827

Grubisha, O., Kaminska, M., Duquerroy, S., Fontan, E., Cordier, F., Haouz, A., Raynal, B., Chiaravalli, J., Delepierre, M., Israël, A. and Véron, M., 2010. DARPin-assisted crystallography of the CC2-LZ domain of NEMO reveals a coupling between dimerization and ubiquitin binding. *Journal of Molecular Biology*: **395 (1)**: 89-104

Guo Y., Yang F., and Tang X., 2017. An Overview of Protein Secretion in Yeast and Animal Cells. In: Jiang L. (eds) *Plant Protein Secretion. Methods in Molecular Biology:* **vol 1662**: 1-17. Humana Press, New York, NY

Guo, J., Gao, X., Liu, B., Li, Y., Liu, W., Lu, J., Liu, C., Xue, R. and Li, X., 2020. Mouse adaption of the H9N2 avian influenza virus causes the downregulation of genes related to innate immune response and ubiquitin-mediated proteolysis in mice. *Medical Microbiology and Immunology*: Epub ahead of print. doi: 10.1007/s00430-020-00656-4

Guo, Z., Wang, X., Li, H. and Gao, Y., 2013. Screening E3 Substrates Using a Live Phage Display Library. *PLoS One*: **8 (10)**: e76622

Haas, TL., Emmerich, CH., Gerlach, B., Schmukle, AC., Cordier, SM., Rieser, E., Feltham, R., Vince, J., Warnken, U., Wenger, T., Koschny, R., Komander, D., Silke, J. and Walczak, H., 2009. Recruitment of the linear ubiquitin chain assembly complex stabilises the TNF-R1 signaling complex and is required for the TNF-mediated gene induction. *Molecular Cell*: **36 (5)**: 831-844

Halidi, K.N.S., Fontan, E., Boucharlat, A., Davignon, L., Charpentier, M., Boullé, M., Weil, R., Israël, A., Laplantine, E. and Agou, F., 2019. Two NEMO-like Ubiquitin-Binding Domains in CEP55 Differently Regulate Cytokinesis. *iScience*: **20**: 292-309

Hansen, SWK., Ohtani, K., Roy, N. and Wakamiya, N., 2016. The Collectins CL-L1, CL-K1 and CL-P1, and Their Roles in Complement and Innate Immunity. *Immunobiology*: **221 (10)**: 1058-1067

Haq, I. U., Chaudhry, W. N., Akhtar, M. N., Andleeb, S. and Qadri, I., 2012. Bacteriophages and their implications on future biotechnology: a review. *Virology Journal*: **9**: 9

Harada, L. K., Silva, E. C., Campos, W. F., Del Fiol, F. S., Vila, M., Dąbrowska, K., Krylov, V. N., and Balcão., V. M., 2018. Biotechnological applications of bacteriophages: State of the art. *Microbiological Research:* **212-213**: 38-58

Harper, S., Besong, T. M., Emsley, J., Scott, D. J. and Dreveny, I., 2011. Structure of the USP15 N-terminal domains: a β-hairpin mediates close association between the DUSP and UBL domains. *Biochemistry*: **50 (37)**: 7995-8004

Harper, S., Gratton, H. E., Cornaciu, I., Oberer, M., Scott, D. J., Emsley, J. and Dreveny, I., 2014. Structure and catalytic regulatory function of ubiquitin specific protease 11 N-terminal and ubiquitin-like domains. *Biochemistry*: **53 (18)**: 2966-2978

He, B., Chen, H., Huang, J., 2019. PhD7Faster 2.0: predicting clones propagating faster from the Ph.D.-7 phage display library by coupling PseAAC and tripeptide composition. *PeerJ*: **7**: e7131

He, F., Wollscheid, H. P., Nowicka, U., Biancospino, M., Valentini, E., Ehlinger, A., Acconcia, F., Magistrati, E., Polo, S. and Walters, K. J., 2016. Myosin VI contains a compact structural motif that binds to ubiquitin chains. *Cell Reports*: **14 (11)**: 2683-2694

Heo, J. M., Ordureau, A., Paulo, J. A., Rinehart, J. and Harper, J. W., 2015. The PINK1-PARKIN mitochondrial ubiquitylation pathway drives a program of OPTN/NDP52 recruitment and TBK1 activation to promote mitophagy. *Molecular Cell*: **60 (1)**: 7-20

Herhaus, L., van den Bedem, H., Tang, S., Maslennikov, I., Wakatsuki, S., Dikic, I. and Rahighi, S., 2019. Molecular recognition of M1-linked ubiquitin chains by native and phosphorylated UBAN domains. *Journal of Molecular Biology:* **17 (9)**: 3146-3156

Heride, C., Urbé, S. and Clague, M. J., 2014. Ubiquitin code assembly and disassembly. *Current Biology*: **24 (6)**: R215-R220

Heusinkveld, H.J. and Westerink, R.H., 2011. microCaveats and limitations of plate reader-based high-throughput kinetic measurements of intracellular calcium levels. *Toxicology and Applied Pharmacology:* **255 (1)**: 1-8

Hicke, L. and Riezman, H., 1996. Ubiquitination of a yeast plasma membrane receptor signals its ligand-stimulated endocytosis. *Cell*: **84 (2)**: .277-287

Hicke, L., Schubert, H. L., and Hill, C. P., 2005. Ubiquitin-binding domains. *Nature Reviews Molecular Cell Biology*: **6**: 610–621

Hirano, S., Kawasaki, M., Ura, H., Kato, R., Taiborg, C., Stenmark, H. and Wakatsuki, S., 2006. Double-sided ubiquitin binding of Hrs-UIM in endosomal protein sorting. *Nature Structural & Molecular Biology*: **13**: 272-277

Hoege, C., Pfander, B., Moldovan, G. L., Pyrowolakis, G. and Jentsch, S., 2002. RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. *Nature*: **419 (6903)**: 135

Hofmann, K., 1996. The UBA domain: a sequence motif present in multiple enzyme classes of the ubiquitination pathway. *Trends in Biochemical Sciences*: **21 (5)**: 172-173

Hofmann, K., and Falquet, L., 2001. A ubiquitin-interacting motif conserved in components of the proteasomal and lysosomal protein degradation systems. *Trends Biochemical Sciences*: **26**: 347–350

Hofmann, K., Bucher, P. and Tschopp, J., 1997. The CARD domain: a new apoptotic signalling motif. *Trends in Biochemical Sciences*: **22 (5)**: 155-156

Hologne, M., Cantrelle, F. X., Riviere, G., Guillière, F., Trivelli, X., and Walker, O., 2016. NMR Reveals the Interplay among the AMSH SH3 Binding Motif, STAM2, and Lys63-Linked Diubiquitin. *Journal of Molecular Biology*: **428 (22)**: 4544-4558

Holstein, M.A., Parimal, S., McCallum, S.A. and Cramer, S.M., 2013. Effects of urea on selectivity and protein–ligand interactions in multimodal cation exchange chromatography. *Langmuir*: **29 (1)**: 158-167

Honda, K., Takaoka, A., and Taniguchi, T., 2006. Type I interferon gene induction by the interferon regulatory factor family of transcription factors. *Immunity*: **25**: 349-360

Honda, K., Yanai, H., Mizutani, T., Negishi, H., Shimada, N., Suzuki, N., Ohba, Y., Takaoka, A., Yeh, W. C., and Taniguchi, T., 2004. Role of a transductional-transcriptional processor complex involving Myd88 and IRF-7 in toll-like receptor signaling. *PNAS*: **101**: 15416-15421

Hu, X., Paul, A., and Wang, B., 2012. Rap80 protein recruitment to DNA double-strand breaks requires binding to both small ubiquitin-like modifier (SUMO) and ubiquitin conjugates. *Journal of Biological Chemistry*: **287**: 25510–25519

Huang, C.H., Su, M.G., Kao, H.J., Jhong, J.H., Weng, S.L. and Lee, T.Y., 2016, December. UbiSite: incorporating two-layered machine learning method with substrate motifs to predict ubiquitin-conjugation site on lysines. *BMC Systems Biology*: **10 (1)**: 6

Huang, F., Zeng, X., Kim, W., Balasubramani, M., Fortian, A., Gygi, S. P., Yates, N. A. and Sorkin, A., 2013. Lysine 63-linked polyubiquitination is required for EGF receptor degradation. *PNAS USA*: **110 (39)**: 15722-15727

Huber, K. E. and Waldor, M. K., 2002. Filamentous phage integration requires the host recombinases Xer C and Xer D. *Nature*: **417 (6889)**: 656-659

Huntley, R. P., Sawford, T., Mutowo-Meullenet, P., Shypitsyna, A., Bonilla, C., Martin, M. J. and O'Donovan, C., 2014. The GOA database: gene ontology annotation updates for 2015. *Nucleic Acids Research*: **43 (D1)**: D1057-D1063

Hurley, J. H., Lee, S., and Prag, G., 2006. Ubiquitin-binding domains. *The Biochemical Journal*: **399**:361–372

Iannolo, G., Minenkova, O., Petruzzelli, R. and Cesareni, G., 1995. Modifying filamentous phage capsid: limits in the size of the major capsid protein. *Journal of Molecular Biology*: **248 (4)**: 835-844

Ikeda, F., Deribe, Y.L., Skånland, S.S., Stieglitz, B., Grabbe, C., Franz-Wachtel, M., Van Wijk, S.J., Goswami, P., Nagy, V., Terzic, J. and Tokunaga, F., 2011. SHARPIN forms a linear ubiquitin ligase complex regulating NF-κB activity and apoptosis. *Nature*: **471 (7340)**: 637-641

Inoue, T., Shoji, W. and Obinata, M., 2000. MIDA1 is a sequence specific DNA binding protein with novel DNA binding properties. *Genes to Cells*: **5 (9)**: 699-709

Janeway, CA. Jr., 1989. Approaching the asymptote? Evolution and revolution in immunology. *Cold Spring Harbor Symposia on Quantitative Biology*: **54**: 1-13

Jepson, C.D. and March, J.B., 2004. Bacteriophage lambda is a highly stable DNA vaccine delivery vehicle. *Vaccine*: **22 (19)**: 2413-2419

Jiang, X., Kinch, L. N., Brautigam, C. A., Chen, X., Du, F., Grishin, N. V. and Chen, Z. J., 2012. Ubiquitin-induced oligomerization of the RNA sensors RIG-I and MDA5 activates antiviral innate immune response. *Immunity*: **36 (6)**: 959-973

Jin, L., Pahuja, K.B., Wickliffe, K.E., Gorur, A., Baumgärtel, C., Schekman, R. and Rape, M., 2012. Ubiquitin-dependent regulation of COPII coat size and function. *Nature*: **482 (7386)**: 495-500

Jouvenet, N., Zhadina, M., Bieniasz, P. D. and Simon, S. M., 2011. Dynamics of ESCRT protein recruitment during retroviral assembly. *Nature Cell Biology*: **13 (4)**: 394

Kabanova, A.P., Shneider, M.M., Korzhenkov, A.A., Bugaeva, E.N., Miroshnikov, K.K., Zdorovenko, E.L., Kulikov, E.E., Toschakov, S.V., Ignatov, A.N., Knirel, Y.A. and Miroshnikov, K.A., 2019. Host specificity of the Dickeya bacteriophage PP35 is directed by a tail spike interaction with bacterial O-antigen, enabling the infection of alternative non-pathogenic bacterial host. *Frontiers in Microbiology*: **9**: 3288

Kahle, P.J., and Haass, C., 2004. How does parkin ligate ubiquitin to Parkinson's disease? *EMBO Reports*: **5**: 681–685

Kanayama, A., Seth, R. B., Sun, L., Ea, C. K., Hong, M., Shaito, A., Chiu, Y. H., Deng, L. and Chen, Z. J., 2004. TAB2 and TAB3 activate the NF-κB pathway through binding to polyubiquitin chains. *Molecular Cell*: **15 (4)**: 535-548

Kane, L. A., Lazarou, M., Fogel, A. I., Li, Y., Yamano, K., Sarraf, S. A., Banerjee, S. and Youle, R. J., 2014. PINK1 phosphorylates ubiquitin to activate Parkin E3 ubiquitin ligase activity. *Journal of Cell Biology*: **205 (2)**: 143-153

Kanehisa, M., Goto, S., Kawashima, S. and Nakaya, A., 2002. The KEGG databases at GenomeNet. *Nucleic Acids Research*: **30 (1)**: 42-46

Kang, R.S., Daniels, C.M., Francis, S.A., Shih, S.C., Salerno, W.J., Hicke, L. and Radhakrishnan, I., 2003. Solution structure of a CUE-ubiquitin complex reveals a conserved mode of ubiquitin binding. *Cell*: **113 (5)**: 621-630

Kasamatsu, J., 2013. Evolution of innate and adaptive immune systems in jawless vertebrates. *Microbiology and Immunology*: **57 (1)**: 1-12

Katzmann, D. J., Babst, M. and Emr, S. D., 2001. Ubiquitin-dependent sorting into the multivesicular body pathway requires the function of a conserved endosomal protein sorting complex, ESCRT-I. *Cell*: **106 (2)**: 145-155

Kawasaki, H., Mizutome, H. and Kretsinger, R.H., 2019. Interaction sites of PEF proteins for recognition of their targets. *International Journal of Biological Macromolecules*: **133**: 1035-1041

Kazansky, Y., Lai, M.Y., Singh, R.K. and Fushman, D., 2018. Impact of different ionization states of phosphorylated Serine-65 on ubiquitin structure and interactions. *Scientific Reports*: **8 (1)**: 2651

Kazlauskaite, A., Martínez-Torres, R.J., Wilkie, S., Kumar, A., Peltier, J., Gonzalez, A., Johnson, C., Zhang, J., Hope, A.G., Peggie, M., Trost, M., van Aalten, D.M., Alessi, D> R., Prescott, A. R., Knebel, A., Walden, H. and Muqit, M.M., 2015. Binding to serine 65-

phosphorylated ubiquitin primes Parkin for optimal PINK1-dependent phosphorylation and activation. *EMBO Reports*: **16 (8)**: 939-954

Keusekotten, K., Elliott, P. R., Glockner, L., Fiil, B. K., Damgaard, R. B., Kulathu, Y., Wauer, T., Hospenthal, M. K., Gyrd-Hansen, M., Krappmann, D. and Hofmann, K., 2013. OTULIN antagonizes LUBAC signaling by specifically hydrolyzing Met1-linked polyubiquitin. *Cell*: **153 (6)**: 1312-1326

Kikuma, K., Li, X., Perry, S., Li, Q., Goel, P., Chen, C., Kim, D., Stavropoulos, N. and Dickman, D., 2019. Cul3 and insomniac are required for rapid ubiquitination of postsynaptic targets and retrograde homeostatic signaling. *Nature Communications:* **10 (1)**: 2998

Kim, H., Chen, J. and Yu, X., 2007. Ubiquitin-binding protein RAP80 mediates BRCA1-dependent DNA damage response. *Science*: **316**: 1202-1205

Kirisako, T., Kamei, K., Murata, S., Kato, M., Fukumoto, H., Kanie, M., Sano, S., Tokunaga, F., Tanaka, K. and Iwai, K., 2006. A ubiquitin ligase complex assembles linear polyubiquitin chains. *The EMBO Journal*: **25 (20)**: 4877-4887

Kirkin, V., Lamark, T., Sou, Y. S., Bjørkøy, G., Nunn, J. L., Bruun, J. A., Shvets, E., Mcewan, D. G., Clausen, T. H., Wild, P., Bilusic, I., Theurillat, J. P., Øvervatn, A., Ishii, T., Elazar, Z., Komatsu, M., Dikic, I. and Johansen, T., 2009a. A role for NBR1 in autophagosomal degradation of ubiquitinated substrates. *Molecular Cell*: **33**: 505-516

Kitada, T., Asakawa, S., Hattori, N., Matsumine, H., Yamamura, Y., Minoshima, S., Yokochi, M., Mizuno, Y. and Shimizu, N., 1998. Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. *Nature*: **392 (6676)**: 605-608

Kitaura, Y., Matsumoto, S., Saton, H., Hitomi, K. and Maki, M., 2001. Peflin and ALG-2, members of the penta-EF-hand protein family form a heterodimer that dissociates in a Ca2+-dependent manner. *Journal of Biological Chemistry*: **276 (17)**: 14053-14058

Kitaura, Y., Satoh, H., Takahashi, H., Shibata, H. and Maki, M., 2002. Both ALG-2 and peflin, penta-EF-hand (PEF) proteins, are stabilized by dimerization through their fifth EF-hand regions. *Archives of Biochemistry and Biophysics*: **399 (1)**: 12-18

Kitaura, Y., Watanabe, M., Satoh, H., Kawai, T., Hitomi, K. and Maki, M., 1999. Peflin, a novel member of the five-EF-hand-protein family, is similar to the apoptosis-linked gene 2 (ALG-2) protein but possesses nonapeptide repeats in the N-terminal hydrophobic region. *Biochemical and biophysical Research Communications:* **263 (1)**: 68-75

Klecker, M. and Dissmeyer, N., 2016. Peptide arrays for binding studies of E3 ubiquitin ligases. In *Plant Proteostasis*: **85-94**. Humana Press, New York, NY

Koide, A., Wojcik, J., Gilbreth, R.N., Reichel, A., Piehler, J. and Koide, S., 2009. Accelerating phage-display library selection by reversible and site-specific biotinylation. *Protein Engineering, Design & Selection*: **22 (11)**: 685-690

Komander, D., Reyes-Turcu, F., Licchesi, J. D., Odenwaelder, P., Wilkinson, K. D. and Barford, D., 2009. Molecular discrimination of structurally equivalent Lys 63-linked and linear polyubiquitin chains. *EMBO Reports*: **10 (5)**: 466-473

Kondapalli, C., Kazlauskaite, A., Zhang, N., Woodroof, H. I., Campbell, D. G., Gourlay, R., Burchell, L., Walden, H., Macartney, T. J., Deak, M. and Knebel, A., 2012. PINK1 is activated by mitochondrial membrane potential depolarization and stimulates Parkin E3 ligase activity by phosphorylating Serine 65. *Open Biology*: **2 (5)**: 120080

Koyano, F., Okatsu, K., Kosako, H., Tamura, Y., Go, E., Kimura, M., Kimura, Y., Tsuchiya, H., Yoshihara, H., Hirokawa, T., Endo, T., Fon, EA., Trempe, JF., Saeki, Y., Tanaka, K. and Matsuda, N., 2014. Ubiquitin is phosphorylated by PINK1 to activate parkin. *Nature*: **510 (7503)**: 162.

Kristensen, P. and Winter, G., 1998. Proteolytic selection for protein folding using filamentous bacteriophages. *Folding and Design:* **3 (5)**: 321-328

Kulathu, Y., Akutsu, M., Bremm, A., Hofmann, K. and Komander, D., 2009. Two-sided ubiquitin binding explains specificity of the TAB2 NZF domain. *Nature Structural and Molecular Biology*: **16**: 1328-1330

Kumar, A., Chaugule, V.K., Condos, T.E.C., Barber, K.R., Johnson, C., Toth, R., Sundarmoorthy, R., Knebel, A., Shaw, G.S., and Walden, H., 2017. Parkin-phosphoubiquitin complex reveals a cryptic ubiquitin binding site required for RBR ligase activity. *Nature Structural and Molecular Biology*: **24 (5)**: 475-483

Kumar, S., Kao, W.H. and Howley, P.M., 1997. Physical interaction between specific E2 and Hect E3 enzymes determines functional cooperativity. *Journal of Biological Chemistry*: **272 (21)**: 13548-13554

Lammer, D., Mathias, N., Laplaza, J.M., Jiang, W., Liu, Y., Callis, J., Goebl, M. and Estelle, M., 1998. Modification of yeast Cdc53p by the ubiquitin-related protein rub1p affects function of the SCFCdc4 complex. *Genes and Development*: **12 (7)**: 914-926

Laplantine, E., Fontan, E., Chiaravalli, J., Lopez, T., Lakisic, G., Véron, M., Agou, F. and Israël, A., 2009. NEMO specifically recognizes K63-linked poly-ubiquitin chains through a new bipartite ubiquitin-binding domain. *The EMBO Journal*: **28 (19)**: 2885–2895

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R. and Thompson, J.D., 2007. Clustal W and Clustal X version 2.0. *Bioinformatics*: **23 (21)**: 2947-2948

Lazarou, M., Sliter, D. A., Kane, L. A., Sarraf, S. A., Wang, C., Burman, J. L., Sideris, D. P., Fogel, A. I. and Youle, R. J., 2015. The ubiquitin kinase PINK1 recruits autophagy receptors to induce mitophagy. *Nature*: **524 (7565)**: 309

Lechtenberg, B.C., Rajput, A., Sanishvili, R., Dobaczewska, M.K., Ware, C.F., Mace, P.D. and Riedl, S.J., 2016. Structure of a HOIP/E2~ ubiquitin complex reveals RBR E3 ligase mechanism and regulation. *Nature*: **529 (7587)**: 546-550

Lee, H. J., Na, K., Kwon, M. S., Kim, H., Kim, K. S. and Paik, Y. K., 2009. Quantitative analysis of phosphopeptides in search of the disease biomarker from the hepatocellular carcinoma specimen. *Proteomics*: **9 (12)**: 3395-3408

Lee, S., Tsai, Y. C., Mattera, R., Smith, W. J., Kostelansky, M. S., Weissman, A. M., Bonifacino, J. S., and Hurley, J. H., 2006. Structural basis of ubiquitin recognition and autoubiquitination by Rabex-5. *Nature Structural and Molecular Biology*: **13**: 264-271

Lee, S., Tumolo, J.M., Ehlinger, A.C., Jernigan, K.K., Qualls-Histed, S.J., Hsu, P.C., McDonald, W.H., Chazin, W.J. and MacGurn, J.A., 2017. Ubiquitin turnover and endocytic trafficking in yeast are regulated by Ser57 phosphorylation of ubiquitin. *Elife*: **6**: e29176

Letunic, I., and Bork, P., 2018. 20 years of the SMART protein domain annotation resource. *Nucleic Acids Research*: **46 (D1)**: D493-D496

Li, D. and Roberts, R., 2001. Human Genome and Diseases: WD-repeat proteins: structure characteristics, biological function, and their involvement in human diseases. *Cellular and Molecular Life Sciences*: **58 (14)**: 2085-2097

Li, F., Xu, D., Wang, Y., Zhou, Z., Liu, J., Hu, S., Gong, Y., Yuan, J. and Pan, L., 2018. Structural insights into the ubiquitin recognition by OPTN (optineurin) and its regulation by TBK1-mediated phosphorylation. *Autophagy*: **14 (1)**: 66-79

Li, J., Chai, Q.Y. and Liu, C.H., 2016. The ubiquitin system: a critical regulator of innate immunity and pathogen–host interactions. *Cellular & Molecular Immunology*: **13 (5)**: 560-576

Liakopoulos, D., Doenges, G., Matuschewski, K. and Jentsch, S., 1998. A novel protein modification pathway related to the ubiquitin system. *The EMBO Journal*: **17 (8)**: 2208-2214

Lide, D. R., ed., 1991. *Handbook of Chemistry and Physics 72nd Edition*. Boca Raton, FL: CRC Press

Lim, K.L., Dawson, V.L., and Dawson, T.M., 2006. Parkin-mediated lysine 63-linked polyubiquitination: a link to protein inclusions formation in Parkinson's and other conformational diseases? *Neurobiology of Aging*: **27**: 524–529

Lim, M., Newman, J.A., Williams, H.L., Masino, L., Aitkenhead, H., Gravard, A.E., Gileadi, O., and Svejstrup, J.Q., 2019. A Ubiquitin-Binding Domain that Binds a Structural Fold Distinct from that of Ubiquitin. *Structure*: ***27 (8)***: 1316-1325

Lin, H., Jiang, M., Liu, L., Yang, Z., Ma, Z., Liu, S., Ma, Y., Zhang, L. and Cao, X., 2019. The long noncoding RNA Lnczc3h7a promotes a TRIM25-mediated RIG-I antiviral innate immune response. *Nature Immunology*: **20 (7)**: 812-823

Lin, L., Hong, W., Ji, X., Han, J., Huang, L. and Wei, Y., 2010. Isolation and characterization of an extremely long tail Thermus bacteriophage from Tengchong hot springs in China. *Journal of Basic Microbiology*: **50 (5)**: 452-456

Liu, Z., Chen, P., Gao, H., Gu, Y., Yang, J., Peng, H., Xu, X., Wang, H., Yang, M., Liu, X. and Fan, L., 2014. Ubiquitylation of autophagy receptor Optineurin by HACE1 activates selective autophagy for tumor suppression. *Cancer Cell*: **26 (1)**: 106-120

Llopis, J., McCaffery, J.M., Miyawaki, A., Farquhar, M.G. and Tsien, R.Y., 1998. Measurement of cytosolic, mitochondrial, and Golgi pH in single living cells with green fluorescent proteins. *Proceedings of the National Academy of Sciences*: **95 (12)**: 6803-6808

Lo, Y. C., Lin, S. C., Rospigliosi, C. C., Conze, D. B., Wu, C. J., Ashwell, J. D., Eliezer, D. and Wu, H., 2009. Structural basis for recognition of diubiquitins by NEMO. *Molecular Cell:* **33 (5)**: 602-615

Lohi, O. and Lehto, V. P., 1998. VHS domain marks a group of proteins involved in endocytosis and vesicular trafficking. *FEBS Letters*: **440 (3)**: 255-257

Lohi, O., Poussu, A., Mao, Y., Quiocho, F. and Lehto, V. P., 2002. VHS domain–a longshoreman of vesicle lines. *FEBS Letters*: **513 (1)**: 19-23

Lomaga, MA., Yeh, WC., Sarosi, I., Duncan, GS., Furlonger, C., Ho, A., Morony, S., Capparelli, C., Van, G., Kaufman, S., van der Heiden, A., Itie, A., Wakeham, A., Khoo, W., Sasaki, T., Cao, Z., Penninger, JM., Paige, CJ., Lacey, DL., Dunstan, CR., Boyle, WJ., Goeddel, DV., Mak, TW., 1999. TRAF6 deficiency results in osteopetrosis and defective interleukin-1, CD40 and LPS signalling. *Genes and Development*: **13 (8)**: 1015-1024

Long, J., Gallagher, T. R., Cavey, J. R., Sheppard, P. W., Ralston, S. H., Layfield, R. and Searle, M. S., 2008. Ubiquitin recognition by the ubiquitin-associated domain of p62 involves a novel conformational switch. *Journal of Biological Chemistry*: **283 (9)**: 5427-5440

Lopez, J., John, S. W., Tenev, T., Rautureau, G. J., Hinds, M. G., Francalanci, F., Wilson, R., Broemer, M., Santoro, M. M., Day, C. L. and Meier, P., 2011. CARD-mediated autoinhibition of cIAP1's E3 ligase activity suppresses cell proliferation and migration. *Molecular Cell*: **42 (5)**: 569-583

Lopez-Pelaez, M., Lamont, D.J., Peggie, M., Shpiro, N., Gray, N.S. and Cohen, P., 2014. Protein kinase IKKβ-catalyzed phosphorylation of IRF5 at Ser462 induces its dimerization and nuclear translocation in myeloid cells. *PNAS*: **111 (49)**: 17432-17437

Løset, G.Å., Roos, N., Bogen, B. and Sandlie, I., 2011. Expanding the versatility of phage display II: improved affinity selection of folded domains on protein VII and IX of the filamentous phage. *PLoS One*: **6 (2)**: e17433

Maki, M., Kitaura, Y., Satoh, H., Ohkouchi, S. and Shibata, H., 2002. Structures, functions and molecular evolution of the penta-EF-hand Ca2+-binding proteins. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics:* **1600 (1-2)**: 51-60

Maki, M., Suzuki, H. and Shibata, H., 2011. Structure and function of ALG-2, a penta-EF-hand calcium-dependent adaptor protein. *Science China Life Sciences*: **54 (8)**: 770-779

Manczyk, N., Veggiani, G., Gish, G. D., Yates, B. P., Ernst, A., Sidhu, S. S. and Sicheri, F., 2019. Dimerization of a ubiquitin variant leads to high affinity interactions with a ubiquitin interacting motif. *Protein Science*: **28 (5)**: 848-856

Manczyk, N., Yates, B. P., Veggiani, G., Ernst, A., Sicheri, F., and Sidhu, S. S., 2017. Structural and functional characterization of a ubiquitin variant engineered for tight and specific binding to an alpha-helical ubiquitin interacting motif. *Protein Science*: **26 (5)**: 1060-1069

Mansour, M.A., 2018. Ubiquitination: Friend and foe in cancer. *The International Journal of Biochemistry & Cell Biology*: **101**: 80-93

Manzi, L., Barrow, A.S., Scott, D., Layfield, R., Wright, T.G., Moses, J.E. and Oldham, N.J., 2016. Carbene footprinting accurately maps binding sites in protein–ligand and protein–protein interactions. *Nature Communications*: **7**: 13288

Mao, Y., Nickitenko, A., Duan, X., Lloyd, T.E., Wu, M.N., Bellen, H. and Quiocho, F.A., 2000. Crystal structure of the VHS and FYVE tandem domains of Hrs, a protein involved in membrane trafficking and signal transduction. *Cell*: **100 (4)**: 447-456

Martinelli, C. and Reichhart, JM., 2005. Evolution and integration of innate immune systems from fruit flies to man: lessons and questions. *Journal of Endotoxin Research*: **11 (4)**: 243-248

Martinez, J., Almendinger, J., Oberst, A., Ness, R., Dillion, CP., Fitzgerald, P., Hengartner, MO. And Green, DR., 2011. Microtubule-associated protein 1 light chain 3 alpha (LC3)-associated phagocytosis is required for the efficient clearance of dead cells. *PNAS*: **108 (42)**: 17396-17401

Martin-Serrano, J. and Neil, S. J., 2011. Host factors involved in retroviral budding and release. *Nature Reviews Microbiology:* **9 (7)**: 519

Matamp, N. and Bhat, S.G., 2019. Genome characterization of novel lytic Myoviridae bacteriophage ϕVP-1 enhances its applicability against MDR-biofilm-forming Vibrio parahaemolyticus. *Archives of Virology*: Epub ahead of print.

Matochko, W. L., Li, S. C., Tang, S. K. Y., and Derda, R., 2014. Prospective identification of parasitic sequences in phage display screens. *Nucleic Acids Research*: **42 (3)**: 1784-1798

Matsumoto, M., Hatakeyama, S., Oyamada, K., Oda, Y., Nishimura, T. and Nakayama, K. I., 2005. Large-scale analysis of the human ubiquitin-related proteome. *Proteomics*: **5 (16)**: 4145-4151

Mattern, M., Sutherland, J., Kadimisetty, K., Barrio, R. and Rodriguez, M.S., 2019. Using Ubiquitin Binders to Decipher the Ubiquitin Code. *Trends in Biochemical Sciences:* **44 (4)***:* 599-615

Matthews, J.M. and Sunde, M., 2002. Zinc Fingers--Folds for Many Occasions. *IUBMB Life:* **54 (6)**: 351-355

Maurer, T. and Wertz, I. E., 2016. Length matters: MINDY is a new deubiquitinase family that preferentially cleaves long polyubiquitin chains. *Molecular Cell*: **63 (1)**: 4-6

Mayer, R. J., Landon, M. and Layfield, R., 1998. Ubiquitin superfolds: intrinsic and attachable regulators of cellular activities?. *Folding and Design*: **3 (5)**: R97-R99

McCallin, S., Alam, S. S., barrette, C., Sultana, S., Berger, B., Huq, S., Krause, L., Bibiloni, R., Schmitt, B., Reuteler, G., and Brüssow, H., 2013. Safety analysis of a Russian phage cocktail: from MetaGenomic analysis to oral application in healthy human subjects. *Virology*: **443**: 187-196

McGourty, C.A., Akopian, D., Walsh, C., Gorur, A., Werner, A., Schekman, R., Bautista, D. and Rape, M., 2016. Regulation of the CUL3 ubiquitin ligase by a calcium-dependent co-adaptor. *Cell*: **167 (2)**: 525-538

McHugh, A., Fernandes, K., Chinner, N., Ibrahim, A.F., Garg, A.K., Boag, G., Hepburn, L.A., Proby, C.M., Leigh, I.M. and Saville, M.K., 2019. The Identification of Potential Therapeutic Targets for Cutaneous Squamous Cell Carcinoma. *Journal of Investigative Dermatology*: Epub ahead of print. DOI: 10.1016/j.jid.2019.09.024

McVoy, M. M., Tenorio, E. and Kauvar, L., 2012. A Native Human Monoclonal Antibody Targeting HCMV gB (AD-2 Site I). *International Journal of Molecular Sciences*: **19 (12)**: E3982

Mevissen, T. E. T., Hospenthal, M. K., Geurink, P. P., Elliot, P. R., Akutsu, M., Arnaudo, N., Ekkebus, R., Kulathu, Y., Wauer, T., El Oualid, F., Freund, S. M. V., Ovaa, H. and Komander, D., 2013. OTU Deubiquitinases Reveal Mechanisms of Linkage Specificity and Enable Ubiquitin Chain Restriction Analysis. *Cell*: **154**: 169-184

Mi, H., Muruganujan, A., Ebert, D., Huang, X., Thomas, P. D., 2019. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Research*: **47 (D1)**: D419-D426

Michel, M. A., Elliott, P. R., Swatek, K. N., Simicek, M., Pruneda, J. N., Wagstaff, J. L., Freund, S. M. and Komander, D., 2015. Assembly and specific recognition of k29-and k33-linked polyubiquitin. *Molecular Cell*: **58 (1)**: 95-109

Miller, J., McLachlan, A. D. and Klug, A., 1985. Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. *The EMBO Journal*: **4 (6)**: 1609-1614

Milo, R., Jorgensen, P., Moran, U., Weber, G. and Springer, M., 2010. BioNumbers – the database of key numbers in molecular cell biology. *Nucleic Acids Research*: **38**: D750-D753

Mintz, P. J., Kim, J., Do, K. A., Wang, X., Zinner, R. G., Cristofanilli, M., Arap, M. A., Hong, W. K., Troncoso, P., Logothetis, C. J., Pasqualini, R., and Arap, W., 2003. Fingerprinting the circulating repertoire of antibodies from cancer patients. *Nature Biotechnology*: **21 (1)**: 57-63

Mizuno, E., Kawahata, K., Kato, M., Kitamura, N. and Komada, M., 2003. STAM proteins bind ubiquitinated proteins on the early endosome via the VHS domain and ubiquitin-interacting motif. *Molecular Biology of the Cell:* **14 (9)**: 3675-3689

Monera, O. D., Sereda, T. J., Zhou, N. E., Kay, C. M. and Hodges, R. S., 1995. Relationship of sidechain hydrophobicity and alpha-helical propensity on the stability of the single-stranded amphipathic alpha-helix. *Journal of Peptide Science*: **1 (5)**: 319-329

Morag, O., Sgourakis, N.G., Baker, D. and Goldbourt, A., 2015. The NMR–Rosetta capsid model of M13 bacteriophage reveals a quadrupled hydrophobic packing epitope. *Proceedings of the National Academy of Sciences*: **112 (4)**: 971-976

Morita, E., and Sundquist, W. I., 2004. Retrovirus budding. *Annual Review of Cell and Developmental Biology*: **20**: 395-425

Morita, E., Sandrin, V., McCullough, J., Katsuyama, A., Hamilton, I.B. and Sundquist, W.I., 2011. ESCRT-III protein requirements for HIV-1 budding. *Cell Host & Microbe*: **9 (3)**: 235-242

Moritsugu, K., Nishi, H., Inariyama, K., Kobayashi, M. and Kidera, A., 2018. Dynamic recognition and linkage specificity in K63 di-ubiquitin and TAB2 NZF domain complex. *Scientific Reports*: **8 (1)**: 16478

Morreale, F.E., Testa, A., Chaugule, V.K., Bortoluzzi, A., Ciulli, A. and Walden, H., 2017. Mind the metal: a fragment library-derived zinc impurity binds the E2 ubiquitin-conjugating enzyme Ube2T and induces structural rearrangements. *Journal of Medicinal Chemistry*: **60 (19)**: 8183-8191

Morriswood, B., Ryzhakov, G., Puri, C., Arden, S. D., Roberts, R., Dendrou, C., Kendrick-Jones, J. and Buss, F., 2007. T6BP and NDP52 are myosin VI binding partners with potential roles in cytokine signalling and cell adhesion. *Journal of Cell Science:* **120 (15)**: 2574-2585

Mukherjee, R., Das, A., Chakrabarti, S. and Chakrabarti, O., 2017. Calcium dependent regulation of protein ubiquitination – Interplay between E3 ligases and calcium binding proteins. *Biochemica et Biophysica Acta*: **1864 (7)**: 1227-1235

Mukhopadhyay, D. and Riezman, H., 2007. Proteasome-independent functions of ubiquitin in endocytosis and signaling. *Science*: **315 (5809)**: 201-205

Muñoz, E. and Deem, M. W., 2008. Amino acid alphabet size in protein evolution experiments: better to search a small library thoroughly or a large library sparsely? *Protein Engineering, Design and Selection*: **21 (5)**: 311-317).

Musacchio, A., Noble, M., Pauptit, R., Wierenga, R. and Saraste, M., 1992. Crystal structure of a Src-homology 3 (SH3) domain. *Nature:* **359 (6398)**: 851-855

Naito, A., Azuma, S., Tanaka, S., Miyazaki, T., Takaki, S., Takatsu, K., Nakao, K., Nakamura, K., Katsuki, M., Yamamoto, T., and Inoue, J., 1999. Severe osteopetrosis, defective interleukin-1 signaling and lymph node organogenesis in TRAF6-defective mice. *Genes to Cells*: **4 (6)**: 353-362

Namdee, K., Khongkow, M., Boonrungsiman, S., Nittayasut, N., Asavarut, P., Temisak, S., Saengkrit, N., Puttipipatkhachorn, S., Hajitou, A., Ruxrungtham, K. and Yata, T., 2018. Thermoresponsive Bacteriophage Nanocarrier as a Gene Delivery Vector Targeted to the Gastrointestinal Tract. *Molecular Therapy-Nucleic Acids*: **12**: 33-44

Naqid, I. A., Owen, J. P., Maddison, B. C., Spiliotopoulos, A., Emes, R. D., Warry, A., Tchórzewska, M. A., Martelli, F., Gosling, R. J., Davies, R. H., La Ragione, R. M. and Gough, K. C., 2016. Mapping polyclonal antibody responses to bacterial infection using next generation phage display. *Scientific Reports*: **6**: 24232

Nepravishta, R., Ferrentino, F., Mandaliti, W., Mattioni, A., Weber, J., Polo, S., Castagnoli, L., Cesareni, G., Paci, M. and Santonico, E., 2019. CoCUN, a Novel Ubiquitin Binding Domain Identified in N4BP1. *Biomolecules*: **9 (7)**: 284

Nijman, S.M., Luna-Vargas, M.P., Velds, A., Brummelkamp, T.R., Dirac, A.M., Sixma, T.K. and Bernards, R., 2005. A genomic and functional inventory of deubiquitinating enzymes. *Cell*: **123 (5)**: 773-786

Nilsson, N., Malmborg, A. C., and Borrebaeck, C. A., 2000. The phage infection process: a functional role for the distal linker region of bacteriophage protein 3. *Journal of Virology*: **74 (9)**: 4229-4235

Nomura, F., Kawai, T., Nakanishi, K. and Akira, S., 2000. NF-κB activation through IKK-i-dependent I-TRAF/TANK phosphorylation. *Genes to Cells*: **5 (3)**: 191-202

Nov, Y., 2012. When second best is good enough: another probabilistic look at saturation mutagenesis. *Applied and Environmental Microbiology*: **78 (1)**: 258-262

Nuber, U. and Scheffner, M., 1999. Identification of determinants in E2 ubiquitin-conjugating enzymes required for hect E3 ubiquitin-protein ligase interaction. *Journal of Biological Chemistry:* **274 (11)**: 7576-7582

Oberst, A., Malatesta, M., Aqeilan, R. I., Rossi, M., Salomoni, P., Murillas, R., Sharma, P., Kuehn, M. R., Oren, M., Croce, C. M. and Bernassola, F., 2007. The Nedd4-binding partner 1 (N4BP1) protein is an inhibitor of the E3 ligase Itch. *Proceedings of the National Academy of Sciences*: **104 (27)**: 11280-11285

Ohno, A., Jee, J., Fujiwara, K., Tenno, T., Goda, N., Tochio, H., Kobayashi, H., Hiroaki, H. and Shirakawa, M., 2005. Structure of the UBA domain of Dsk2p in complex with ubiquitin: molecular determinants for ubiquitin recognition. *Structure*: **13 (4)**: 521-532

Ohtake, F., Saeki, Y., Sakamoto, K., Ohtake, K., Nishikawa, H., Tsuchiya, H., Ohta, T., Tanaka, K. and Kanno, J., 2015. Ubiquitin acetylation inhibits polyubiquitin chain elongation. *EMBO Reports*: **16 (2)**: 192-201

Ordureau, A., Heo, J. M., Duda, D. M., Paulo, J. A., Olszewski, J. L., Yanishevski, D., Rinehart, J., Schulman, B. A. and Harper, J. W., 2015. Defining roles of PARKIN and ubiquitin phosphorylation by PINK1 in mitochondrial quality control using a ubiquitin replacement strategy. *Proceedings of the National Academy of Sciences*: **112 (21)**: 6637-6642

Ordureau, A., Sarraf, S. A., Duda, D. M., Heo, J. M., Jedrychowski, M. P., Sviderskiy, V. O., Olszewski, J. L., Koerber, J. T., Xie, T., Beausoleil, S. A. and Wells, J. A., 2014. Quantitative proteomics reveal a feedforward mechanism for mitochondrial PARKIN translocation and ubiquitin chain synthesis. *Molecular Cell*: **56 (3)**: 360-375

Oughtred, R., Stark, C., Breikreutz, B. J., Rust, J., Boucher, L., Chang, C., Kolas, N., O'Donnell, L., Leung, G., McAdam, R., Zhang, F., Dolma, S., Willems, A., Coulombe-Huntington, J., Chatr-Aryamontri A., Dolinkski, K., and Tyers, M., 2019. The BioGRID interaction database: 2019 update. *Nucleic Acids Research*: **47 (D1)**: D529-D541

Oved, S., Mosesson, Y., Zwang, Y., Santonico, E., Shtiegman, K., Marmor, M. D., Kochupurakkal, B. S., Katz, M., Lavi, S., Cesareni, G. and Yarden, Y., 2006. Conjugation to Nedd8 instigates ubiquitylation and down-regulation of activated receptor tyrosine kinases. *Journal of Biological Chemistry*: **281 (31)**: 21640-21651

Pai, M. T., Tzeng, S. R., Kovacs, J. J., Keaton, M. A., Li, S. S. C., Yao, T. P. and Zhou, P., 2007. Solution structure of the Ubp-M BUZ domain, a highly specific protein module that recognizes the C-terminal tail of free ubiquitin. *Journal of Molecular Biology*: **370 (2)**: 290-302

Papa, F. R. and Hochstrasser, M., 1993. The yeast DOA4 gene encodes a deubiquitinating enzyme related to a product of the human tre-2 oncogene. *Nature*: **366 (6453)**: 313-319

Park, H. H., 2019. Caspase recruitment domains for protein interactions in cellular signaling. *International Journal of Molecular Medicine:* **43 (3)**: 1119-1127

Park, H. H., Lo, Y. C., Lin, S. C., Wang, L., Yang, J. K. and Wu, H., 2007b. The death domain superfamily in intracellular signaling of apoptosis and inflammation. *Annual Review of Immunology*: **25**: 561-586

Park, H. H., Logette, E., Raunser, S., Cuenin, S., Walz, T., Tschopp, J. and Wu, H., 2007a. Death domain assembly mechanism revealed by crystal structure of the oligomeric PIDDosome core complex. *Cell*: **128 (3)**: 533-546

Pashkova, N., Gakhar, L., Winistorfer, S. C., Yu, L., Ramaswamy, S., and Piper, R. C., 2010. WD40 Repeat Propellers Define a Ubiquitin-Binding Domain that Regulates Turnover of F Box Proteins. Molecular Cell: 40 (3): 433-443

Paul, A. and Wang, B., 2017. RNF8-and Ube2S-dependent ubiquitin lysine 11-linkage modification in response to DNA damage. *Molecular Cell*: **66 (4)**: 458-472

Penengo, L., Mapelli, M., Murachelli, A. G., Confalonieri, S., Magri, L., Musacchio, A., Di Fiore, P. P., Polo, S. and Schneider, T. R., 2006. Crystal structure of the ubiquitin binding domains of rabex-5 reveals two modes of interaction with ubiquitin. *Cell*: **124 (6)**: 1183-1195

Peng, J., Schwartz, D., Elias, J. E., Thoreen, C. C., Cheng, D., Marsischky, G., Roelofs, J., Finley, D. and Gygi, S. P., 2003. A proteomics approach to understanding protein ubiquitination. *Nature Biotechnology*: **21 (8)**: 921

Pesiri, V., Di Muzio, E., Polticelli, F., and Acconcia, F., 2016. Selective binding of estrogen receptor α to ubiquitin chains. *IUBMB Life*: **68 (7)**: 569-577

Pesiri, V., La Rosa, P., Stano, P. and Acconcia, F., 2013. Identification of an estrogen receptor α non-covalent ubiquitin-binding surface: role in 17β-estradiol-induced transcriptional activity. *Journal of Cell Sci*ence: **126 (12)**: 2577-2582

Peters, E. A., Schatz, P. J., Johnson, S. S. and Dower, W. J., 1994. Membrane insertion defects caused by positive charges in the early mature region of protein pIII of filamentous phage fd can be corrected by prlA suppressors. *Journal of Bacteriology*: **176**: 4296-4305

Pham, H.T., Nguyen, T.T., Nguyen, L.P., Han, S.S., Lim, Y.S. and Hwang, S.B., 2019. Hepatitis C Virus Downregulates Ubiquitin-Conjugating Enzyme E2S Expression To Prevent Proteasomal Degradation of NS5A, Leading to Host Cells More Sensitive to DNA Damage. *Journal of Virology*: **93 (2)**: e01240-18

Pickart, C.M. and Rose, I.A., 1985. Ubiquitin carboxyl-terminal hydrolase acts on ubiquitin carboxyl-terminal amides. *Journal of Biological Chemistry*: **260 (13)**: 7903-7910

Pie, J., Kim, B.H. and Grishin, N.V., 2008. PROMALS3D: a tool for multiple sequence and structure alignment. *Nucleic Acids Research:* **36 (7):** 2295-300

Pinato, S., Gatti, M., Scandiuzzi, C., Confalonieri, S. and Penengo, L., 2011. UMI, a novel RNF168 ubiquitin binding domain involved in the DNA damage signaling pathway. *Molecular and Cellular Biology*: **31 (1)**: 118-126

Pinheiro, L.A., Pereira, C., Barreal, M.E., Gallego, P.P., Balcão, V.M. and Almeida, A., 2019. Use of phage ϕ6 to inactivate Pseudomonas syringae pv. actinidiae in kiwifruit plants: in vitro and ex vivo experiments. *Applied Microbiology and Biotechnology*: Epub ahead of print.

Pirnay, J. P., Blasdel, B. G., Bretaudeau, L., Buckling, A., Chanishvili, N., Clar, J. R., Corte-Real, S., Debarbieux, L., Dublanchet, A., De, V. D., Gabard, J., Garcia, M., Goderdzishvili, M., Górski, A., Hardcastle, J., Huys, I., Kutter, E., Lavigne, R., Merbishvili, M., Olchawa, E., Parikka, K. J., Patey, O., Pouilot, F., Resch, G., Rohde, C., Scheres, J., Skurnik, M., Vaneechoutte, M., Van, P. L., Verbeken, G., Zizi, M., and Van den Eede, G., 2015. Quality and safety requirements for sustainable phage therapy products. *Pharmaceutical Research*: **32 (7)**: 2173-2179

Ponting, C.P., 2000. Proteins of the endoplasmic-reticulum-associated degradation pathway: domain detection and function prediction. *Biochemical Journal*: **351 (2)**: 527-535

Pornillos, O., Alam, S. L., Rich, R. L., Myszka, D. G., Davis, D. R., and Sundquist, W. L., 2002. Structure and functional interactions of the Tsg101 UEV domain. *EMBO Journal:* **21 (10)**: 2397-2406

Prag, G., Lee, S., Mattera, R., Arighi, C. N., Beach, B. M., Bonifacino, J. S. and Hurley, J. H., 2005. Structural mechanism for ubiquitinated-cargo recognition by the Golgi-localized, γ-ear-containing, ADP-ribosylation-factor-binding proteins. *Proceedings of the National Academy of Sciences*: **102 (7)**: 2334-2339

Prag, G., Misra, S., Jones, E.A., Ghirlando, R., Davies, B.A., Horazdovsky, B.F. and Hurley, J.H., 2003. Mechanism of ubiquitin recognition by the CUE domain of Vps9p. *Cell:* **113 (5)**: 609-620

Prakriya, M. and Lewis, R.S., 2015. Store-operated calcium channels. *Physiological Reviews*: **95 (4)**: 1383-1436

PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC

Qian, J., Yang, J., Zhang, X., Zhang, B., Wang, J., Zhou, M., Tang, K., Li, W., Zeng, Z., Zhao, X., Shen, S., and Li., G., 2001. Isolation and characterization of a novel cDNA, UBAP1, derived from the tumor suppressor locus in human chromosome 9p21–22. *Journal of Cancer Research and Clinical Oncology:* **127 (10)**: 613-618

Qin, H., Srinivasula, S. M., Wu, G., Fernandes-Alnemri, T., Alnemri, E. S. and Shi, Y., 1999. Structural basis of procaspase-9 recruitment by the apoptotic protease-activating factor 1. *Nature*: **399 (6736)**: 549-557

Rahbarnia, L., Farajnia, S., Babaei, H., Majidi, J., Veisi, K., Ahmadzadeh, V. and Akbari, B., 2017. Evolution of Phage Display Technology: From Discovery to Application. *Journal of Drug Targeting*: **25 (3):** 216-224

Rahighi, S., Ikeda, F., Kawaskai, M., Akutsu, M., Suzuki, N., Kato, R., Kensche, T., Uejima, T., Bloor, S., Komander, D., Randow, F., Wakatsuki, D. and Dikic, I., 2009. Specific recognition of linear ubiquitin chains by NEMO is important for NF-kappaB activation. *Cell*: **136**: 1098-1109

Rahim, A., Coutelle, C. and Harbottle, R., 2003. High-throughput Pyrosequencing of a phage display library for identification of enriched target-specific peptides. *Biotechniques*: **35**: 317-324

Railoa, L., Lussier-Price, M., Gagnon, D., Lafrance-Vanasse, J., Mascle, X., Arseneault, G., Legault, P., Archambault, J. and Omichinski, J. G., 2013. Structural Characterization of a Noncovalent Complex between Ubiquitin and the Transactivation Domain of the Erythroid-Specific Factor EKLF. *Structure*: **21 (11)**: 2014-2024

Rakonjac, J., Feng, J.N. and Model, P., 1999. Filamentous phage are released from the bacterial membrane by a two-step mechanism involving a short C-terminal fragment of pIII. *Journal of Molecular Biology*: **289 (5)**: 1253-1265

Randles, L. and Walters, K.J., 2012. Ubiquitin and its binding domains. *Frontiers in Bioscience (Landmark edition)*: **17**: 2140-2157

Rape, M., 2018. Ubiquitylation at the crossroads of development and disease. *Nature reviews Molecular Cell Biology*: **19 (1)**: 59-70

Ray, D. S., 1969. Replication of bacteriophage M13: II The role of replicative forms in a single-strand synthesis. *Journal of Molecular Biology*: **43 (3)**: 631-643

Rayl, M., Truitt, M., Held, A., Sargeant, J., Thorsen, K. and Hay, J. C., 2016. Penta-EF-Hand Protein Peflin Is a Negative Regulator of ER-To-Golgi Transport. *PLoS One*: **11 (6)**: e0157227

Reche, P., 2018. *SeqLogos* [online]. Available at <http://imed.med.ucm.es/Tools/seqlogo.html> [Accessed 27th July 2019)

Rego, M.A., Kolling, F.W., Vuono, E.A., Mauro, M. and Howlett, N.G., 2012. Regulation of the Fanconi anemia pathway by a CUE ubiquitin-binding domain in the FANCD2 protein. *Blood*: **120 (10)**: 2109-2117

Rehman, S. A. A., Kristariyanto, Y. A., Choi, S. Y., Nkosi, P. J., Weidlich, S., Labib, K., Hofmann, K. and Kulathu, Y., 2016. MINDY-1 is a member of an evolutionarily conserved and structurally distinct new family of deubiquitinating enzymes. *Molecular Cell*: **63 (1)**: 146-155

Ren, J., Chen, X. and Chen, Z.J., 2014. IKKβ is an IRF5 kinase that instigates inflammation. *PNAS*: **111 (49)**: 17438-17443

Ren, X. and Hurley, J. H., 2010. VHS domains of ESCRT-0 cooperate in high-avidity binding to polyubiquitinated cargo. *The EMBO Journal*: **29 (6)**: 1045-1054

Resto, V. A., Caballero, O. L., Buta, M. R., Westra, W. H., Wu, L., Westendorf, J. M., Jen, J., Hieter, P. and Sidransky, D., 2000. A putative oncogenic role for MPP11 in head and neck squamous cell cancer. *Cancer Research*: **60 (19)**: 5529-5535

Reyes-Turcu, F. E., Horton, J. R., Mullally, J. E., Heroux, A., Cheng, X. and Wilkinson, K. D., 2006. The ubiquitin binding domain ZnF UBP recognizes the C-terminal diglycine motif of unanchored ubiquitin. *Cell*: **124 (6)**: 1197-1208

Rice,P. Longden,I. and Bleasby,A, 2000. *EMBOSS: The European Molecular Biology Open Software Suite*. *Trends in Genetics*: 16 **(6)**: *276-277*

Richly, H. and Di Croce, L., 2011. The flip side of the coin: role of ZRF1 and histone H2A ubiquitination in transcriptional activation. *Cell Cycle*: **10 (5)**: 745-750

Richly, H., Rocha-Viegas, L., Ribeiro, J. D., Demajo, S., Gundem, G., Lopez-Bigas, N., Nakagawa, T., Rospert, S., Ito, T. and Di Croce, L., 2010. Transcriptional activation of polycomb-repressed genes by ZRF1. *Nature*: **468 (7327)**: 1124-1128

Richter, B., Sliter, D. A., Herhaus, L., Stolz, A., Wang, C., Beli, P., Zaffagnini, G., Wild, P., Martens, S., Wagner, S. A., Youle, R. J. and Dikic, I., 2016. Phosphorylation of OPTN by TBK1 enhances its binding to Ub chains and promotes selective autophagy of damaged mitochondria. *Proceedings of the National Academy of Sciences*: **113 (15)**: 4039–4044

Riechmann, L. and Hollinger, P., 1997. The C-terminal domain of TolA is the coreceptor for filamentous phage infection of E. coli. *Cell***: 90 (2)**: 351-360

Riedinger, C., Boehringer, J., Trempe, J.F., Lowe, E.D., Brown, N.R., Gehring, K., Noble, M.E., Gordon, C. and Endicott, J.A., 2010. Structure of Rpn10 and its interactions with polyubiquitin chains and the proteasome subunit Rpn12. *Journal of Biological Chemistry*: **285 (44)**: 33992-34003

Rittinger, K. and Ikeda, F., 2017. Linear ubiquitin chains: enzymes, mechanisms and biology. *Open Biology*: **7 (4)**: 170026

Rizzo, A. A., Salerno, P. E., Bezsonova, I. and Korzhnev, D. M., 2014. NMR structure of the human Rad18 zinc finger in complex with ubiquitin defines a class of UBZ domains in proteins linked to the DNA damage response. *Biochemistry*: **53 (37)**: 5895-5906

Rodi, D. J., Soares, A. S., and Makowski, L., 2002. Quantitative assessment of peptide sequence diversity in M13 combinatorial peptide phage display libraries. *Journal of Molecular Biology*: **322 (5)**: 1039-1052

Rollinson, S., Rizzu, P., Sikkink, S., Baker, M., Halliwell, N., Snowden, J., Traynor, B. J., Ruano, D., Cairns, N., Rohrer, J. D., Mead, S., Collinge, J., Rossor, M., Akay, E., Guerreiro, R., Rademakers, R., Morrison, K. E., Pastor, P., Alonso, E., Martinez-Lage, P., Graff-Radford, N., Neary, D., Heutink, P., Mann, D. M., Van Swieten, J., and Pickering-Brown, S.M., 2009. Ubiquitin associated protein 1 is a risk factor for frontotemporal lobar degeneration. *Neurobiology of Aging*: **30 (4)**: 656-665

Rossmann, M. G., Morais, M. C., Leiman, P. G., and Zhang, W., 2005. Combining X-ray crystallography and electron microscopy. *Structure*: **13**: 35-362

Rothwarf, D.M., Zandi, E., Natoli, G. and Karin, M., 1998. IKK-γ is an essential regulatory subunit of the IκB kinase complex. *Nature*: **395 (6699)**: 297-300

Russel, M. and Model, P., 2006. Filamentous Phage. In: Calendar RC, ed., 2006. *The Bacteriophages 2nd Edition*. Oxford University Press, New York: 146-160

Russel, M., Lowman, H. B. and Clackson, T., 2004. Introduction to phage biology and phage display. In: Clackson, T. and Lowman, H. B., eds., 2004. *Phage Display; A Practical Approach*. Oxford University Press, New York: 1-26

Sahlender, D. A., Roberts, R. C., Arden, S. D., Spudich, G., Taylor, M. J., Luzio, J. P., Kendrick-Jones, J. and Buss, F., 2005. Optineurin links myosin VI to the Golgi complex and is involved in Golgi organization and exocytosis. *The Journal of Cell Biology:* **169 (2)**: 285-295

Santonico, E., Nepravishta, R., Mandaliti, W., Castagnoli, L., Cesareni, G. and Paci, M., 2019. CUBAN, a Case Study of Selective Binding: Structural Details of the Discrimination between Ubiquitin and NEDD8. *International Journal of Molecular Sciences*: **20 (5)**: E1185

Santos, S. B., Carvalho, C., Azeredo, J., and Ferreira, E. C., 2014. Population dynamics of a Salmonella lytic phage and its host: implications of the host bacterial growth rate in modelling. *PLoS One*: **9 (7)**: e102507

Sato, Y., Fujita, H., Yoshikawa, A., Yamashita, M., Yamagata, A., Kaiser, SE., Iwai, K., and Fukai, S., 2011. Specific recognition of linear ubiquitin chains by the Npl4 zinc finger (NZF) domain of HOIL-1 subunit of the linear ubiquitin chain assembly complex. *PNAS*: **108 (51)**: 20520-20525

Sato, Y., Yoshikawa, A., Yamashita, M., Yamagata, A. and Fukai, S., 2009. Structural basis for specific recognition of Lys-63-linked polyubiquitin chains by NZF domains of TAB2 and TAB3. *The EMBO Journal*: **28**: 3903-3909

Scala, G., Chen, X., Liu, W., Telles, J. N., Cohen, O. J., Vaccarezza, M., Igarashi, T., and Fauci, A. S., 1999. Selection of HIVE-specific immunogenic epitopes by screening random peptide libraries with HIV-1-positive sera. *Journal of Immunology*: **162**: 6155-6161

Schauber, C., Chen, L., Tongaonkar, P., Vega, I., Lambertson, D., Potts, W. and Madura, K., 1998. Rad23 links DNA repair to the ubiquitin/proteasome pathway. *Nature*: **391 (6668)**: 715-718

Schimmack, G., Schorpp, K., Kutzner, K., Gehring, T., Brenke, J.K., Hadian, K. and Krappmann, D., 2017. YOD1/TRAF6 association balances p62-dependent IL-1 signaling to NF-κB. *Elife*: **6**: e22416

Schmidtke, G., Kalveram, B., Weber, E., Bochtler, P., Lukasiak, S., Hipp, M.S. and Groettrup, M., 2006. The UBA domains of NUB1L are required for binding but not for accelerated degradation of the ubiquitin-like modifier FAT10. *Journal of Biological Chemistry*: **281 (29)**: 20045-20054

Schreiner, P., Chen, X., Husnjak, K., Randles, L., Zhang, N., Elsasser, S., Finley, D., Dikic, I., Walters, K. J. and Groll, M., 2008. Ubiquitin docking at the proteasome through a novel pleckstrin-homology domain interaction. *Nature*: **453 (7194)**: 548-552

Scott, D., Garner, T.P., Long, J., Strachan, J., Mistry, S.C., Bottrill, A.R., Tooth, D.J., Searle, M.S., Oldham, N.J. and Layfield, R., 2016. Mass spectrometry insights into a tandem ubiquitin-binding domain hybrid engineered for the selective recognition of unanchored polyubiquitin. *Proteomics*: **16 (14)**: 1961-1969

Scott, J. K. and Smith, G. P., 1990. Searching for peptide ligands with an epitope library. *Science*: **249**: 386-390

Sedki, M., Chen, X., Chen, C., Ge, X. and Mulchandani, A., 2020. Non-lytic M13 phage-based highly sensitive impedimetric cytosensor for detection of coliforms. *Biosensors and Bioelectronics:* **148**: 111794

Shabek, N., Herman-Bachinsky, Y., Buchsbaum, S., Lewinson, O., Haj-Yahya, M., Hejjaoui, M., Lashuel, H. A., Sommer, T., Brik, A. and Ciechanover, A., 2012. The Size of the Proteasomal Substrate Determines Whether Its Degradation Will Be Mediated by Mono- or Polyubiquitylation. *Molecular Cell*: **48**: 87-97

Sharma, B., Joshi, D., Yadav, P.K., Gupta, A.K. and Bhatt, T.K., 2016. Role of ubiquitin-mediated degradation system in plant biology. *Frontiers in Plant Science*: **7**: 806

Sharma, K., D'Souza, R. C., Tyanova, S., Schaab, C., Wiśniewski, J. R., Cox, J. and Mann, M., 2014. Ultradeep human phosphoproteome reveals a distinct regulatory nature of Tyr and Ser/Thr-based signaling. *Cell Reports*: **8 (5)**: 1583-1594

Sharma, P., Murillas, R., Zhang, H. and Kuehn, M. R., 2010. N4BP1 is a newly identified nucleolar protein that undergoes SUMO-regulated polyubiquitylation and proteasomal turnover at promyelocytic leukemia nuclear bodies. *Journal of Cell Science*: **123 (8)**: 1227-1234

Sharp, P. M., and Li, W.-H., 1987. Molecular evolution of ubiquitin genes. *Trends in Ecology & Evolution:* **2 (11):** 328-332

Shave, S., Mann, S., Koszela, J., Kerr, A., and Auer, M., 2018. PuLSE: Quality control and quantification of peptide sequences explored by phage display libraries. *PLoS One*: **13 (2)**: e0193332

Shaw, B.F., Schneider, G.F., Arthanari, H., Narovlyansky, M., Moustakas, D., Durazo, A., Wagner, G. and Whitesides, G.M., 2011. Complexes of native ubiquitin and dodecyl sulfate illustrate the nature of hydrophobic and electrostatic interactions in the binding of proteins and surfactants. *Journal of the American Chemical Society*: **133 (44)**: 17681-17695

Shiba, Y., Katoh, Y., Shiba, T., Yoshino, K., Takatsu, H., Kobayashi, H., Shin, H.W., Wakatsuki, S. and Nakayama, K., 2004. GAT (GGA and Tom1) domain responsible for ubiquitin binding and ubiquitination. *Journal of Biological Chemistry:* **279 (8)**: 7105-7111

Shih, S. C., Prag, G., Francis, S. A., Sutanto, M. A., Hurley, J. H. and Hicke, L., 2003. A ubiquitin-binding motif required for intramolecular monoubiquitylation, the CUE domain. *The EMBO Journal*: **22 (6)**: 1273-1281

Shim, J.H., Xiao, C., Paschal, A.E., Bailey, S.T., Rao, P., Hayden, M.S., Lee, K.Y., Bussey, C., Steckel, M., Tanaka, N. and Yamada, G., 2005. TAK1, but not TAB1 or TAB2, plays an essential role in multiple signaling pathways in vivo. *Genes and Development*: **19 (22)**: 2668-2681

Shiromizu, T., Adachi, J., Watanabe, S., Murakami, T., Kuga, T., Muraoka, S. and Tomonaga, T., 2013. Identification of missing proteins in the neXtProt database and unregistered phosphopeptides in the PhosphoSitePlus database as part of the Chromosome-centric Human Proteome Project. *Journal of Proteome Research*: **12 (6)**: 2414-2421

Sidhu, S. S., Lowman, H. B., Cunningham, B. C. and Wells, J. A., 2000. Phage display for selection of novel binding peptides. In: Thorner, J., Emr, S. D. and Abelson, J. N., eds., 2000. *Methods in Enzymology*. Academic Press, Cambridge, USA. **Vol 328**: 333-363

Sieber, V., Plückthun, A. and Schmid, F.X., 1998. Selecting proteins with improved stability by a phage-based method. *Nature Biotechnology*: **16 (10)**: 955-960

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J. and Thompson, J.D., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*: **7 (1)**: 539

Sigrist, C. J. A., de Castro, E., Cerutti, L., Cuche, B. A., Hulo, N., Bridge, A., Bougueleret, L., and Xenarios, I., 2012. *New and continuing developments at PROSITE*. *Nucleic Acids Research*: **Database issue**: D344-D347

Silva, G. M., Finley, D. and Vogel, C., 2015. K63 polyubiquitination is a new modulator of the oxidative stress response. *Nature Structural & Molecular Biology*: **22 (2)**: 116-12

Sims, J. J. and Cohen, R. E., 2009. Linkage-Specific Avidity Defines the Lysine 63-Linked Polyubiquitin Binding Preference of Rap80. *Molecular Cell*: **33**: 775-783

Sims, J.J., Haririnia, A., Dickinson, B.C., Fushman, D. and Cohen, R.E., 2009. Avid interactions underlie the Lys63-linked polyubiquitin binding specificities observed for UBA domains. *Nature Structural & Molecular Biology*: **16 (8)**: 883-889

Singh, R. K., Zerath, S., Kleifeld, O., Scheffner, M., Glickman, M. H. and Fushman, D., 2012. Recognition and cleavage of related to ubiquitin 1 (Rub1) and Rub1-ubiquitin chains by components of the ubiquitin-proteasome system. *Molecular & Cellular Proteomics*: **11 (12)**: 1595-1611

Singh, S., Hinkley, T., Nugen, S.R. and Talbert, J.N., 2019. Colorimetric detection of Escherichia coli using engineered bacteriophage and an affinity reporter system. *Analytical and Bioanalytical Chemistry*: **411 (27)**: 7273-7279

Skaug, B., Jiang, X. and Chen, Z. J., 2009. The role of ubiquitin in NF-κB regulatory pathways. *Annual Review of Biochemistry*: **78**: 769-796

Sloper-Mould, KE., Jemc, JC., Pickart, CM. and Hicke, L., 2001. Distinct functional surface regions on ubiquitin. *Journal of Biological Chemistry*: **276 (32)**: 30483-30489

Smeal, S.W., Schmitt, M.A., Pereira, R.R., Prasad, A. and Fisk, J.D., 2017. Simulation of the M13 life cycle I: Assembly of a genetically-structured deterministic chemical kinetic simulation. *Virology*: **500**: 259-274

Smith, G. P., 1985. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science*: **228**: 1315-1317

Sobhian, B., Shao, G., Lilli, D. R., Culhane, A. C., Moreau, L. A., Xia, B., Livingston, D. M. and Greenberg, R. A., 2007. RAP80 targets BRCA1 to specific ubiquitin structures at DNA damage sites. *Science*: **316**: 1198-1202

Sokratous, K., Roach, L. V., Channing, D., Strachan, J., Long, J., Searle, M. S., Layfield, R. and Oldham, N. J., 2012. Probing affinity and ubiquitin linkage selectivity of ubiquitin-binding domains using mass spectrometry. *Journal of the American Chemical Society*: **134**: 6416-6424

Sowa, M. E., Bennett, E. J., Gygi, S. P. and Harper, J. W., 2009. Defining the human deubiquitinating enzyme interaction landscape. *Cell*: **138 (2)**: 389-403

Spel, L., Nieuwenhuis, J., Haarsma, R., Stickel, E., Bleijerveld, O. B., Altelaar, M., Boelens, J. J., Brummelkamp, T. R., Nierkens, S. and Boes, M., 2018. Nedd4-binding protein 1 and TNFAIP3-interacting protein 1 control MHC-1 display in neuroblastoma. *Cancer Research*: **78 (23)**: 6621-6631

Spiliotopoulos, A., Ferreras, L. B., Densham, R. M., Caulton, S. G., Maddison, B. C., Morris, J. R., Dixon, J. E., Gough, K. C. and Dreveny, I., 2019. Discovery of peptide ligands targeting a specific ubiquitin-like domain–binding site in the deubiquitinase USP11. *Journal of Biological Chemistry*: **294 (2)**: 424-436

Spudich, G., Chibalina, M. V., Au, J. S. Y., Arden, S. D., Buss, F. and Kendrick-Jones, J., 2007. Myosin VI targeting to clathrin-coated structures and dimerization is mediated by binding to Disabled-2 and PtdIns (4, 5) P 2. *Nature Cell Biology*: **9 (2)**: 176

Srimathi, T., Robbins, S. L., Dubas, R. L., Hasegawa, M., Inohara, N. and Park, Y. C., 2008. Monomer/dimer transition of the caspase-recruitment domain of human Nod1. *Biochemistry*: **47 (5)**: 1319-1325

Stamenova, S.D., French, M.E., He, Y., Francis, S.A., Kramer, Z.B. and Hicke, L., 2007. Ubiquitin binds to and regulates a subset of SH3 domains. *Molecular Cell*: **25 (2)**: 273-284

Stefani, F., Zhang, L., Taylor, S., Donovan, J., Rollinson, S., Doyotte, A., Brownhill, K., Bennion, J., Pickering-Brown, S. and Woodman, P., 2011. UBAP1 is a component of an endosome-specific ESCRT-I complex that is essential for MVB sorting. *Current Biology*: **21 (14)**: 1245-1250

Steiner, D., Forrer, P., Stumpp, M. T. and Pluckthun, A., 2006. Signal sequences directing cotranslational translocation expand the range of proteins amendable to phage display. *Nature Biotechnology*: **24**: 823-831

Stolz, A., Putyrski, M., Kutle, I., Huber, J., Wang, C., Major, V., Sidhu, S. S., Youle, R. J., Rogov, V. V., Dötsch, V. and Ernst, A., 2017. Fluorescence-based ATG8 sensors monitor localization and function of LC3/GABARAP proteins. *The EMBO Journal*: **36 (4)**: 549-564

Stothard, P., 2000. The Sequence Manipulation Suite: JavaScript programs for analysing and formatting protein and DNA sequences. *Biotechniques*: **28**: 1102-1104

Strous, G.J., Van Kerkhof, P., Govers, R., Ciechanover, A. and Schwartz, A.L., 1996. The ubiquitin conjugation system is required for ligand-induced endocytosis and degradation of the growth hormone receptor. *The EMBO Journal*: **15 (15)**: 3806-3812

Suire, C. N., Nainar, S., Fazio, M., Kreutzer, A. G., Paymozd-Yazdi, T., Topper, C. L., Thompson, C. R., and Leissring, M. A., 2018. Peptidic inhibitors of insulin-degrading enzyme with potential for dermatological applications discovered via phage display. *PLoS One*: **13**: e0193101

Swaney, D. L., Rodríguez-Mias, R. A. and Villén, J., 2015. Phosphorylation of ubiquitin at Ser65 affects its polymerization, targets, and proteome-wide turnover. *EMBO Reports*: **16 (9)**: 1131-1144

Swanson, K. A., Kang, R. S., Stamenova, D. S., Hicke, L. and Radhakrishnan, I., 2003. Solution structure of Vps27 UIM-ubiquitin complex important for endosomal sorting and receptor downregulation. *The EMBO Journal*: **22**: 4597-4606

Takahashi, T. S., Sato, Y., Yamagata, A., Goto-Ito, S., Saijo, M. and Fukai, S., 2019. Structural basis of ubiquitin recognition by the winged-helix domain of Cockayne syndrome group B protein. *Nucleic Acids Research*: **47 (7)**: 3784-3794

Takahashi, T.S., Hirade, Y., Toma, A., Sato, Y., Yamagata, A., Goto-Ito, S., Tomita, A., Nakada, S. and Fukai, S., 2018. Structural insights into two distinct binding modules for Lys63-linked polyubiquitin chains in RNF168. *Nature Communications*: **9 (1)**: 170

Takaoka, A., Yanai, H., Kondo, S., Duncan, G., Negishi, H., Mizutani, T., Kano, S.I., Honda, K., Ohba, Y., Mak, T.W. and Taniguchi, T., 2005. Integral role of IRF-5 in the gene induction programme activated by Toll-like receptors. *Nature*: **434 (7030)**: 243-249

Takatsu, H., Katoh, Y., Shiba, Y. and Nakayama, K., 2001. Golgi-localizing, γ-adaptin ear homology domain, ADP-ribosylation factor-binding (GGA) proteins interact with acidic dileucine sequences within the cytoplasmic domains of sorting receptors through their Vps27p/Hrs/STAM (VHS) domains. *Journal of Biological Chemistry*: **276 (30)**: 28541-28545

Teo, H., Gill, D.J., Sun, J., Perisic, O., Veprintsev, D. B., Vallis, Y., Emr, S. D. and Williams, R. L., 2006. ESCRT-I core and ESCRT-II GLUE domain structures reveal role for GLUE in linking to ESCRT-I and membranes. *Cell*: **125 (1)**: 99-111

Terzaghi, B. E., Terzaghi, E., and Coombs, D., 1979. The role of the collar/whisker complex in bacteriophage T4D tail fibre attachment. *Journal of Molecular Biology*: **127**: 1-14

Teyra, J., Huang, H., Jain, S., Guan, X., Dong, A., Liu, Y., Tempel, W., Min, J., Tong, Y., Kim, P. M., Bader, G. D., and Sidhu, S. S., 2017. Comprehensive Analysis of the Human SH3 Domain Family Reveals a Wide Variety of Non-canonical Specificities. *Structure*: **25 (10)**: 1598-1610

Thomas, P. D., Kejariwal, A., Guo, N., Mi, H., Campbell, M., Muruganujan, A. and Lazareva-Ulitsky, B., 2006. Application for protein sequence-function evolution data: mRNA/protein expression analysis and coding SNP scoring tools. *Nucleic Acids Research*: **34 (2)**: W645-W650

Thrower, J. S., Hoffman, L., Rechsteiner, M. and Pickart, C. M., 2000. Recognition of the polyubiquitin proteolytic signal. *The EMBO Journal*: **19 (1)**: 94-102

Tokunaga, F., Nakagawa, T., Nakahara, M., Saeki, Y., Taniguchi, M., Sakata, S.I., Tanaka, K., Nakano, H. and Iwai, K., 2011. SHARPIN is a component of the NF-κB-activating linear ubiquitin chain assembly complex. *Nature*: **471 (7340)**: 633-636

Toma, A., Takahashi, T.S., Sato, Y., Yamagata, A., Goto-Ito, S., Nakada, S., Fukuto, A., Horikoshi, Y., Tashiro, S. and Fukai, S., 2015. Structural basis for ubiquitin recognition by ubiquitin-binding zinc finger of FAAP20. *PloS One*: **10 (3)**: e0120887

Tomatis, V. M., Papadopulos, A., Malintan, N. T., Martin, S., Wallis, T., Gormal, R. S., Kendrick-Jones, J., Buss, F. and Meunier, F. A., 2013. Myosin VI small insert isoform maintains exocytosis by tethering secretory granules to the cortical actin. *Journal of Cell Biology*: **200 (3)**: 301-320

Tran, H. J., Allen, M. D., Löwe, J. and Bycroft, M., 2003. Structure of the Jab1/MPN domain and its implications for proteasome function. *Biochemistry*: **42 (39)**: 11460-11465

Troelstra, C., van Gool, A., de Wit, J., Vermeulen, W., Bootsma, D. and Hoeijmakers, J. H., 1992. ERCC6, a member of a subfamily of putative helicases, is involved in Cockayne's syndrome and preferential repair of active genes. *Cell:* **71 (6)**: 939-953

Trojet, S.N., Caumont-Sarcos, A., Perrody, E., Comeau, A.M. and Krisch, H.M., 2011. The gp38 adhesins of the T4 superfamily: a complex modular determinant of the phage's host specificity. *Genome Biology and Evolution*: **3**: 674-686

Tsai, C. F., Wang, Y. T., Yen, H. Y., Tsou, C. C., Ku, W. C., Lin, P. Y., Chen, H. Y., Nesvizhskii, A. I., Ishihama, Y. and Chen, Y. J., 2015. Large-scale determination of absolute phosphorylation stoichiometries in human cells by motif-targeting quantitative proteomics. *Nature Communications*: **6**: 6622

Tumbarello, D. A., Kendrick-Jones, J. and Buss, F., 2013. Myosin VI and its cargo adaptors–linking endocytosis and autophagy. *Journal of Cell Science:* **126 (12)**: 2561-2570

Tzvetkov, N. and Breuer, P., 2007. Josephin domain-containing proteins from a variety of species are active de-ubiquitination enzymes. *Biological Chemistry*: **388 (9)**: 973-978

UniProt Consortium, 2018. UniProt: the universal protein knowledgebase. *Nucleic Acids Research:* **46 (5)**: 2699

Uzunova, K., Göttsche, K., Miteva, M., Weisshaar, S. R., Glanemann, C., Schnellhardt, M., Niessen, M., Scheel, H., Hofmann, K., Johnson, E. S. and Praefcke, G. J., 2007. Ubiquitin-dependent proteolytic control of SUMO conjugates. *Journal of Biological Chemistry*: **282 (47)**: 34167-34175

Varadan, R., Assfalg, M., Haririnia, A., Raasi, S., Pickart, C. and Fushman, D., 2004. Solution conformation of Lys63-linked di-ubiquitin chain provides clues to functional diversity of polyubiquitin signaling. *Journal of Biological Chemistry*: **279 (8)**: 7055-7063

Varadan, R., Assfalg, M., Raasi, S., Pickart, C. and Fushman, D., 2005. Structural determinants for selective recognition of a Lys48-linked polyubiquitin chain by a UBA domain. *Molecular Cell:* **18 (6)**: 687-698

Varadan, R., Walker, O., Pickart, C. and Fushman, D., 2002. Structural properties of polyubiquitin chains in solution. *Journal of Molecular Biology*: **324 (4)**: 637-647

Varadi, A. and Rutter, G.A., 2002. Dynamic imaging of endoplasmic reticulum Ca2+ concentration in insulin-secreting MIN6 Cells using recombinant targeted cameleons: roles of sarco (endo) plasmic reticulum Ca2+-ATPase (SERCA)-2 and ryanodine receptors. *Diabetes*: **51 (S1)**: S190-S201

Ver Heul, A. M., Fowler, C. A., Ramaswamy, S. and Piper, R. C., 2013. Ubiquitin regulates caspase recruitment domain-mediated signaling by nucleotide-binding oligomerization domain-containing proteins NOD1 and NOD2. *Journal of Biological Chemistry*: **288 (10)**: 6890-6902

Ver Heul, A. M., Gakhar, L., Piper, R. C. and Subramanian, R., 2014. Crystal structure of a complex of NOD1 CARD and ubiquitin. *PLoS One*: **9 (8)**: e104017

Vijay-Kumar, S., Bugg, C.E. and Cook, W.J., 1987. Structure of ubiquitin refined at 1.8 Å resolution. *Journal of Molecular Biology*: **194 (3)**: 531-544

Virdee, S., Ye, Y., Nguyen, D. P., Komander, D. and Chin, J. W., 2010. Engineered diubiquitin synthesis reveals Lys29-isopeptide specificity of an OTU deubiquitinase. *Nature Chemical Biology*: **6 (10)**: 750

Wagner, S., Carpentier, I., Rogov, V., Kreike, M., Ikeda, F., Löhr, F., Wu, C. J., Ashwell, J. D., Dötsch, V., Dikic, I. and Beyaert, R., 2008. Ubiquitin binding mediates the NF-κB inhibitory potential of ABIN proteins. *Oncogene*: **27 (26)**: 3739-3745

Walsh, C., 2006. *Posttranslational modification of proteins: Expanding nature's inventory*. Englewood, Colorado: Roberts and Co. Publishers: xxi: 490 p. p.

Wan, M., Wang, X., Huang, C., Xu, D., Wang, Z., Zhou, Y. and Zhu, Y., 2019. A bacterial effector deubiquitinase specifically hydrolyses linear ubiquitin chains to inhibit host inflammatory signalling. *Nature Microbiology*: **4 (8)**: 1282-1293

Wang, B. and Elledge, S. J., 2007. Ubc13/Rnf8 ubiquitin ligases control foci formation of the Rap80/Abraxas/Brca1/Brcc36 complex in response to DNA damage. *Proceedings of the National Academy of Sciences*: **104 (52)**: 20759-20763

Wang, B., Alam, S.L., Meyer, H.H., Payne, M., Stemmler, T.L., Davis, D.R. and Sundquist, W.I., 2003. Structure and ubiquitin interactions of the conserved zinc finger domain of Npl4. *Journal of Biological Chemistry*: **278 (22)**: 20225-20234

Wang, B., Matsuoka, S., Ballif, B. A., Zhang, D., Smogorzewska, A., Gygi, S. P. and Elledge, S. J., 2007. Abraxas and RAP80 form a BRCA1 protein complex required for the DNA damage response. *Science*: **316 (5828)**: 1194-1198

Wang, C., Deng, L., Hong, M., Akkaraju, G. R., Inoue, J. I. and Chen, Z. J., 2001. TAK1 is a ubiquitin-dependent kinase of MKK and IKK. *Nature*: **412 (6844)**: 346

Wang, J., Peng, Q., Lin, Q., Childress, C., Carey, D. and Yang, W., 2010. Calcium activates Nedd4 E3 ubiquitin ligases by releasing the C2 domain-mediated auto-inhibition. *Journal of Biological Chemistry*: **285 (16)**: 12279-12288

Wang, T., Yin, L., Cooper, E. M., Lai, M. Y., Dickey, S., Pickart, C. M., Fushman, D., Wilkinson, K. D., Cohen, R. E. and Wolberger, C., 2009. Evidence for bidentate substrate binding as the basis for the K48 linkage specificity of otubain 1. *Journal of Molecular Biology*: **386 (4)**: 1011-1023

Wang, Y., Nartiss, Y., Steipe, B., McQuibban, G.A. and Kim, P.K., 2012. ROS-induced mitochondrial depolarization initiates PARK2/PARKIN-dependent mitochondrial degradation by autophagy. *Autophagy:* **8 (10)**: 1462-1476.

Watanabe, N., Wachi, S. and Fujita, T., 2003. Identification and Characterization of BCL-3-binding Protein implications for transcription and DNA repair or recombination. *Journal of Biological Chemistry*: **278 (28)**: 26102-26110

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. and Barton, G. K., 2009. Jalview Version 2 – a multiple sequence alignment editor ad analysis workbench. *Bioinformatics*: **25 (9)**: 1189-1191

Wauer, T., Swatek, K. N., Wagstaff, J. L., Gladkova, C., Pruneda, J. N., Michel, M. A., Gersch, M., Johnson, C. M., Freund, S. M. and Komander, D., 2015. Ubiquitin Ser65 phosphorylation affects ubiquitin structure, chain assembly and hydrolysis. *The EMBO Journal*: **34 (3)**: 307-325

Webster, R. E., 1996. Biology of the filamentous bacteriophage. In: Kay, B., Winter, J. and McCafferty, J., eds., 1996. *Phage Display of Peptides and Proteins, 1st edition*. Academic Press, Cambridge, USA: 1-20

Weissenhorn, W. and Göttlinger, H., 2011. Essential ingredients for HIV-1 budding. *Cell Host & Microbe*: **9 (3)**: 172-174

Werner, A., Iwasaki, S., McGourty, C.A., Medina-Ruiz, S., Teerikorpi, N., Fedrigo, I., Ingolia, N.T. and Rape, M., 2015. Cell-fate determination by ubiquitin-dependent regulation of translation. *Nature*: **525 (7570)**: 523-7

Wertz, I. E., O'rourke, K. M., Zhou, H., Eby, M., Aravind, L., Seshagiri, S., Wu, P., Wiesmann, C., Baker, R., Boone, D. L. and Ma, A., 2004. De-ubiquitination and ubiquitin ligase domains of A20 downregulate NF-κB signalling. *Nature*: **430 (7000)**: 694-699

White, K.A., Grillo-Hill, B.K. and Barber, D.L., 2017. Cancer cell behaviors mediated by dysregulated pH dynamics at a glance. *Journal of Cell Science:* **130 (4)**: 663-669

Wiemuth, D., van de Sandt, L., Herr, R. and Gründer, S., 2012. Transient receptor potential N (TRPN1) from Xenopus interacts with the penta-EF-hand protein peflin. *FEBS Letters*: **586 (24)**: 4276-4281

Williamson, M.P., 2013. Using chemical shift perturbation to characterise ligand binding. *Progress in Nuclear Magnetic Resonance Spectroscopy*: **73**: 1-16

Winston, J.T., Strack, P., Beer-Romero, P., Chu, C.Y., Elledge, S.J. and Harper, J.W., 1999. The SCFβ-TRCP–ubiquitin ligase complex associates specifically with phosphorylated destruction motifs in IκBα and β-catenin and stimulates IκBα ubiquitination in vitro. *Genes and Development:* **13 (3)**: 270-283

Withers-Ward, E. S., Mueller, T. D., Chen, I. S. and Feigon, J., 2000. Biochemical and structural analysis of the interaction between the UBA (2) domain of the DNA repair protein HHR23A and HIV-1 Vpr. *Biochemistry*: **39 (46)**: 14103-14112

Wojtunik-Kulesza, K., Oniszczuk, A. and Waksmundzka-Hajnos, M., 2019. An attempt to elucidate the role of iron and zinc ions in development of Alzheimer's and Parkinson's diseases. *Biomedicine & Pharmacotherapy:* **111**: 1277-1289

Wright, J. D., Mace, P. D. and Day, C. L., 2016. Noncovalent ubiquitin interactions regulate the catalytic activity of ubiquitin writers. *Trends in Biochemical Sciences*: **41 (11)**: 924-937

Wright, T., Rea, S.L., Goode, A., Bennett, A.J., Ratajczak, T., Long, J.E., Searle, M.S., Goldring, C.E., Park, B.K., Copple, I.M. and Layfield, R., 2013. The S349T mutation of SQSTM1 links Keap1/Nrf2 signalling to Paget's disease of bone. *Bone*: **52 (2)**: 699-706

Wu, C.J., Conze, D.B., Li, T., Srinivasula, S.M. and Ashwell, J.D., 2006. Sensing of Lys 63-linked polyubiquitination by NEMO is a key event in NF-κB activation. *Nature Cell Biology*: **8 (4)**: 398-406

Wurzer, B., Zaffagnini, G., Fracchiolla, D., Turco, E., Abert, C., Romanov, J. and Martens, S., 2015. Oligomerization of p62 allows for selection of ubiquitinated cargo and isolation membrane during selective autophagy. *eLife*: **4**: e08941

Xia, Z. P., Sun, L., Chen, X., Pineda, G., Jiang, X., Adhikari, A., Zeng, W. and Chen, Z. J., 2009. Direct activation of protein kinases by unanchored polyubiquitin chains. *Nature*: **461 (7260)**: 114

Xu, C., Peng, Y., Zhang, Q., Xu, X-P., Kong, X-M., and Shi, W-F., 2018. USP4 positively regulates RLR-induced NF-κB activation by targeting TRAF6 and K48-linked deubiquitination and inhibits enterovirus 71 replication. *Scientific Reports*: **8**: 13418

Yamaoka, S., Courtois, G., Bessia, C., Whiteside, S. T., Weil, R., Agou, F., Kirk, HE., Kay, RJ. and Israel, A., 1998. Complementation cloning of NEMO, a component of the IkappaB kinase complex essential for NF-kappaB activation. *Cell*: **93**: 1231-1240

Yan, J., Kim, Y-S., Yang, X-P., Li, L-P., Liao, G., Xia, F., and Jetten, A. M., 2007. The ubiquitin- interacting motif containing protein RAP80 interacts with BRCA1 and functions in DNA damage repair response. Cancer Research: 67: 6647–6656

Yang, J. and Zhang, Y., 2015. I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Research*: **43**: W174-W181

Yang, X., Zou, P., Yao, J., Yun, D., Bao, H., Du, R., Long, J. and Chen, X., 2010. Proteomic dissection of cell type-specific H2AX-interacting protein complex associated with hepatocellular carcinoma. *Journal of Proteome Research*: **9 (3)**: 1402-1415

Yang, Y., Ding, Y., Zhou, C., Wen, Y., and Zhang, N., 2019a. Structural and function studies of USP20 ZnF_UBP domain by NMR. *Protein Science*: doi:10.1002/pro.3675 (epub ahead of print)

Yang, Z.Q., Tao, X.Y., Zhang, H., Rao, S.Q., Gao, L., Pan, Z.M. and Jiao, X.A., 2019b. Isolation and characterization of virulent phages infecting Shewanella baltica and Shewanella putrefaciens, and their application for biopreservation of chilled channel catfish (Ictalurus punctatus). *International Journal of Food Microbiology*: **292**: 107-117

Ye, S.X., Gong, Z., Yang, J., An, Y.X., Liu, Z., Zhao, Q., Lescop, E., Dong, X. and Tang, C., 2019. Ubiquitin is double-phosphorylated by PINK1 for enhanced pH-sensitivity of conformational switch. *Protein & Cell*: 1-6

Ye, Y., Akutsu, M., Reyes-Turcu, F., Enchev, R. I., Wilkinson, K. D. and Komander, D., 2011. Polyubiquitin binding and cross-reactivity in the USP domain deubiquitinase USP21. *EMBO Reports*: **12 (4)**: 350-357

Yuan, T., Mohan, D., Laserson, U., Ruczinski, I., Baer, A. N. and Larman, B., 2018. Improved Analysis of Phage ImmunoPrecipitation Sequencing (PhIP-Seq) Data Using a Z-score Algorithm. *bioRxiv*: DOI: 10.1101/285916

Zarnegar, B., Westin, Annika, Evangelidou, S., and Hallgren, J., 2018. Innate Immunity Induces the Accumulation of Lung Mast Cells During Influenza Infection. *Frontiers in Immunology*: **9**: 2288

Zeng, W., Sun, L., Jiang, X., Chen, X., Hou, F., Adhikari, A., Xu, M. and Chen, Z. J., 2010. Reconstitution of the RIG-I pathway reveals a signaling role of unanchored polyubiquitin chains in innate immunity. *Cell*: **141 (2)**: 315-330

Zhang, D., Chen, T., Ziv, I., Rosenzweig, R., Matiuhin, Y., Bronner, V., Glickman, M. H. and Fushman, D., 2009. Together, Rpn10 and Dsk2 can serve as a polyubiquitin chain-length sensor. *Molecular Cell*: **36 (6)**: 1018-1033

Zhang, D., Raasi, S. and Fushman, D., 2008. Affinity makes the difference: nonselective interaction of the UBA domain of Ubiquilin-1 with monomeric ubiquitin and polyubiquitin chains. *Journal of Molecular Biology*: **377 (1)**: 162-180

Zhang, J., Clark, K., Lawrence, T., Peggie, M.W. and Cohen, P., 2014. An unexpected twist to the activation of IKKβ: TAK1 primes IKKβ for activation by autophosphorylation. *Biochemical Journal*: **461 (3)**: 531-537

Zhang, W., Wu, K. P., Sartori, M. A., Kamadurai, H. B., Ordureau, A., Jiang, C., Mercredi, P. Y., Murchie, R., Hu, J., Persaud, A., Mukherjee, M., Li, N., Doye, A., Walker, J. R., Sheng, Y., Hao, Z., Li, Y., Brown, K. R., Lemichez, E.., Chen, J., Tong, Y., Harper, J. W., Moffat, J., Rotin, D., Schulman, B. A. and Sidhu, S. S., 2016. System-wide modulation of HECT E3 ligases with selective ubiquitin variant probes. *Molecular Cell*: **62 (1)**: 121-136

Zhang, Y. H., Zhou, C. J., Zhou, Z. R., Song, A. X. and Hu, H. Y., 2011. Domain analysis reveals that a deubiquitinating enzyme USP13 performs non-activating catalysis for Lys63-linked polyubiquitin. *PLoS One*: **6 (12)**: e29362

Zhang, Z., Wang, Y., Li, C., Shi, Z., Hao, Q., Wang, W., Song, X., Zhao, Y., Jiao, S. and Zhou, Z., 2015. The transitional endoplasmic reticulum ATPase p97 regulates the alternative nuclear factor NF-κB signaling via partial degradation of the NF-κB subunit p100. *Journal of Biological Chemistry*: **290 (32)**: 19558-19568

Zheng, N. and Shabek, N., 2017. Ubiquitin ligases: structure, function, and regulation. *Annual Review of Biochemistry*: **86**: 129-157

Zhou, H., Di Palma, S., Preisinger, C., Peng, M., Polat, A. N., Heck, A. J. and Mohammed, S., 2012. Toward a comprehensive characterization of a human cancer cell phosphoproteome. *Journal of Proteome Research*: **12 (1)**: 260-271

Zhou, J., Xu, Y., Lin, S., Guo, Y., Deng, W., Zhang, Y., Guo, A. and Xue, Y., 2017. iUUCD 2.0: an integrated database of regulators for ubiquitin and ubiquitin-like conjugation. *Nucleic Acids Research*: **46 (D1)**: D447-D453

Zhou, R., Kabra, R., Olson, D.R., Piper, R.C. and Snyder, P.M., 2010. Hrs controls sorting of the epithelial Na+ channel between endosomal degradation and recycling pathways. *Journal of Biological Chemistry*: **285 (40)**: 30523-30530

Zong, N., Ping, P., Lau, E., Choi, H. J., Ng, D. C., Meyer, D., Fang, C., Li, H., Wang, D., Zelaya, I. M. and Yates III, J. R., 2014. Lysine ubiquitination and acetylation of human cardiac 20S proteasomes. *PROTEOMICS–Clinical Applications*: **8 (7-8)**: 590-594

Zu, X., Liu, Y., Wang, S., Jin, R., Zhou, Z., Liu, H., Gong, R., Xiao, G., and Wang, W., 2014. Peptide inhibitor of Japanese encephalitis virus infection targeting envelope protein domain III. *Antiviral Research*: **104**: 7-14

8 APPENDICES

A.1 Introduction

A1.1 Table of known human UBDs

Table 1.2.1. The 29 types of human UBDs, their key structures and examples of UBD-containing proteins.

Family	Name	Structure	Example Protein
UBA	Ubiquitin associated	Three-helix bundle	p62/SQSTM1 (Ciani *et al.*, 2003)
UIM	Ubiquitin-interacting motif	Single α-helix (Fisher *et al.*, 2003)	Epsin (Hofmann and Falquet, 2001)
CUE	Coupling of ubiquitin conjugation to endoplasmic reticulum degradation	Three-helix bundle (Prag *et al.*, 2003)	FANCD2 (Rego *et al.*, 2012)
VHS	VPS-27, Hrs and STAM	Super helix of eight α-helices (Mao *et al.*, 2000)	STAM (Mizuno *et al.*, 2003)
GAT	GGA and TOM	Three-helix bundle	GGA3 (Shiba *et al.*, 2004)
MIU	Motif interacting with ubiquitin	Single α-helix	Rabex-5 (Penengo *et al.*, 2006)
DUIM	Double-sided ubiquitin-interacting motif	Single α-helix binding two ubiquitin molecules	Hrs (Hirano *et al.*, 2006)
UBAN	Ubiquitin binding in ABIN and NEMO	Parallel coiled-coil dimer (Grubisha *et al.*, 2010)	NEMO (Wagner *et al.*, 2008)
UMI	UIM- and MIU-related UBD	α-helical (Takahashi *et al.*, 2018)	RNF168 (Pinato *et al.*, 2011)
UBM	Ubiquitin-binding motif	Helix turn helix (Burschowsky *et al.*, 2011)	Polymerase iota (Bomar *et al.*, 2010)
NZF	Npl4 zinc finger	Four antiparallel β-strands	Npl4 (Wang *et al.*, 2003)
UBP	Ubiquitin-specific processing protease	Zinc finger, globular fold with a deep cleft and pocket	USP5 (Reyes-Turcu *et al.*, 2006)
UBZ	Ubiquitin-binding zinc finger	Zinc finger, ββα-fold	Pol η (Bomar *et al.*, 2007)
ZnF_A20	A20 type zinc finger	Zinc finger	Rabex-5 (Lee *et al.*, 2006)
UBC	Ubiquitin conjugating enzyme	β-sheet	UBE2D3 (Brzovic *et al.*, 2006)
UEV	UBC E2 variant	αβ-sequence	TSG101 (Pornillos *et al.*, 2002)
Beta-Prp/ WD40	Beta-Propeller/ WD40	four antiparallel beta strands (Li *et al.*, 2001)	Fbw7 (Pashkova et atl., 2010)

SH3	Src homology 3	β-barrel fold (Noble *et al.*, 1992)	Amphiphysin (Stamenova *et al.*, 2007)
CARD	Caspase recruitment domain	Six helix bundle (Park *et al.*, 2007b)	NOD2 (Ver Huel *et al.*, 2013)
*PRU	Pleckstrin homology domain	Two antiparallel β-sheets	Rpn13 (Schreiner *et al.*, 2008)
*Jab1/MPN	Jab1, Mpr1, Pad1 N-terminal	Two β-sheets and two α-helices	Prp8p (Tran *et al.*, 2003)
*PFU	PLAA family ubiquitin binding	Four β-strands and two α-helices	PLAA (Fu *et al.*, 2009)
*GLUE	Split-pleckstrin homology domain	Two β-sheets in a barrel and a C-terminal α-helix	Vps36 (Teo *et al.*, 2006)
*CUBAN	Cullin-binding domain associating with NEDD8	Three α-helix bundle	KHNYN (Castagnoli *et al.*, 2019)
*CoCUN	Cousin of CUBAN	Three α-helix bundle	N4BP1 (Nepravishta *et al.*, 2019)
*SOUBA	Solenoid of overlapping UBAs	Seven α-helices in a solenoid	UBAP1 (Agromayor *et al.*, 2012)
*ZRF1-UBD	Zuotin-related factor 1 UBD	Unknown	ZRF1 (Richly *et al.*, 2010)
*UBW	Ubiquitin-binding winged-helix domain	Helix-turn-helix	CSB (Anindya *et al.*, 2010)
*MyUb	Myosin VI Ubiquitin-binding domain	Helix-turn-helix	Myosin VI (He *et al.*, 2016)

Unique domains are marked with an asterisk.

A.2 Methods

A2.1 Buffers and solutions

All stocks and solutions were made up to volume with water. Unless stated otherwise distilled water was used for all solutions.

Buffer	**Composition**
Minimal media	42mM Na_2HPO_4, 22mM KH_2PO_4, 8.5mM NaCl, 2.5mM $MgSO_4$, 100mM $CaCl_2.2H_2O$, pH 7.4. Autoclaved, then when cool, sterile filtered in 2% (v/v) Minimal Media Additives
Minimal media additives	935mM NH_4Cl, 1M glucose, 2mM biotin, 2mM thiamine, 5% (v/v) Trace Elements Solution, 5% (v/v) MEM vitamins (Sigma-Aldrich, #M6895)
Trace Elements Solution	590mM $ZnCl_2$, 2.5mM $FeCl_3$, 150µM $CuCl_2$, 160µM $MnCl_2 \cdot 4H_2O$, 52µM $Na_2B_4O_7.10H_2O$, 16µM $(NH_4)_6Mo_7O_{24}.4H_2O$

5x Acetic acid/ EDTA (AE) buffer	250mM acetic acid, 10mM EDTA (pH 8.0), pH 4.0. Prepared with deionised water. Sterile filtered and degassed.
AE High Salt buffer	5% (v/v) 5x AE buffer, 40% (v/v) 5M NaCl. Prepared with deionised water. Sterile filtered and degassed.
Gel Filtration buffer	60% (v/v) KP buffer, 4% (v/v) 5M NaCl. Prepared with deionised water. Sterile filtered and degassed.
KP buffer	12.3% (v/v) 1M K_2HPO_4, 7.7% 1M KH_2PO_4
Tris Acetate EDTA (TAE) buffer	40mM Tris-acetate, 1mM EDTA, pH 8.0
Glutathione Elution buffer	50mM Tris, 100mM reduced glutathione, pH 8.0. Made immediately before use.
Thrombin Cleavage buffer	20mM Tris, 150mM NaCl, 3.33mM $CaCl_2$, pH 8.4
CNBr Coupling buffer	100mM $NaCHO_3$, 500mM NaCl, pH 8.3
CNBr Tris buffer	100mM Tris, 500mM NaCl, pH 8.0
CNBr Acetate buffer	100mM NaOAc, 500mM NaCl, pH 4.0
Gel Loading buffer (GLB)	0.15M Tris, 8M urea, 2.5% (w/v) SDS, 20% (v/v) glycerol, 10% (w/v) 2-mercaptoethanol, 3% (w/v) DTT, 0.1% (w/v) bromophenol blue, pH 6.8
SDS-PAGE Buffer A	1.1M Tris, 0.1% (w/v) SDS, 30% (w/v) glycerol, pH 8.8
SDS-PAGE Buffer B	1.1M Tris, 0.1% (w/v) SDS, pH 8.8
SDS-PAGE Stacking buffer	0.14M Tris, 0.1% (w/v) SDS, pH 6.8
Electrode buffer	25mM Tris, 190mM glycine, 0.1% (w/v) SDS
Western Transfer buffer	25mM Tris, 192mM glycine, 20% (v/v) methanol
Coomassie Blue stain	50% (v/v) methanol, 20% (v/v) glacial acetic acid, 0.12% (w/v) Coomassie Brilliant Blue R-250
Coomassie Destain	10% (v/v) methanol, 10% (v/v) glacial acetic acid
Tris Buffer Saline (TBS)	20mM Tris, 150mM NaCl, pH 7.5
Ponceau-S Stain	0.1% (w/v) Ponceau S, 10% (v/v) glacial acetic acid

Capture Buffer	50mM NaCl, 10mM Tris, 10% (v/v) glycerol, 1% (v/v) Triton X-100, pH 6.5. Kept at 4°C for a maximum of one week.
Minimal Capture Buffer	50mM NaCl, 10mM Tris, pH 6.5. Kept at 4°C for a maximum of one week.
5x M9 Salt Solution	240mM Na_2HPO_4, 110mM H_2KO_4P, 90mM NH_4Cl, 42mM NaCl. Autoclaved and cooled before use.
M9 Minimal Media Agar	15g/L agar, 25% (v/v), 5x M9 salt solution. Autoclaved and cooled to 50°C before addition of 6.6% (w/v) glucose, 2.1mM $MgCl_2$ (autoclaved), 1.5mM thiamine hydrochloride and antibiotics as required
LB Broth	10g/L tryptone, 5g/L yeast extract, 10g/L NaCl. Autoclaved
LB Agar	10g/L tryptone, 5g/L yeast extract, 10g/L NaCl, 15g/L agar, 2.5g/L select agar. Autoclaved and cooled to 50°C before addition of antibiotics as required
2YT	16g/L tryptone, 10g/L yeast extract, 5g/L NaCl. Autoclaved and cooled to 50°C

A2.2 Reagents and chemicals

	Manufacturer	**Product code**
Phosphate buffered saline (PBS)	Sigma	P4417
Glutathione-Sepharose 4B	GE Healthcare	17-0756-01
p-nitrophenyl phosphate (pNPP)	Thermo Fisher Scientific	37620
phospho-Ser65-ubiquitin	Boston Biochem	U-102
Luria Bertani Broth (LB)	Sigma	L3022
MEM Vitamins	Sigma	M6895
HiTrap SP HP sepharose cation exchange column	GE Healthcare	17115201
HiLoad 26/600 Superdex 75 pg column	GE Healthcare	GE28-9893-34
HiTrap 5mL Desalting column	GE Healthcare	GE29-0486-84
pOTB7-PEF1 plasmid	Harvard PlasmID	HsCD00325639
T4 10x buffer	New England Biolabs	B0202S
T4 DNA ligase	New England Biolabs	M0202S
Calf intestinal alkaline phosphatase	New England Biolabs	M0290S
Xho1 restriction enzyme	New England Biolabs	R0146S

EcoR1 restriction enzyme	New England Biolabs	R3101S
CutSmart buffer	New England Biolabs	B7204S
QIAprep Spin Miniprep kit	QIAGEN	27104
QIAquick Gel Extraction kit	QIAGEN	28704
Pierce™ BCA Protein Assay Kit	Thermo Scientific	23225
Cyanogen Bromide-activated-Sepharose 4B	GE Healthcare	17043001
Phosphatase inhibitor cocktail 2	Sigma-Aldrich	P5726
Protease inhibitor cocktail	Sigma-Aldrich	P8340
SYBR Safe	Thermo Fisher	S33102
High DNA molecular weight marker	New England Biolabs	B7025
DNA gel loading dye	New England Biolabs	B7024S
BLUeye prestained molecular weight marker	Geneflow	S6-0024
PQ Universal developer	Ilford	1757314
Hypam fixer	Ilford	1758285
Nucleospin Gel and PCR clean up kit	Machery-Nagel	740609
Q5 buffer Reaction Buffer pack	New England Biolabs	B9027
dNTP mix 10mM	New England Biolabs	N0447
Q5 DNA polymerase	New England Biolabs	M0491S
Low DNA molecular weight marker	New England Biolabs	N0557S
dsDNA Broad Range Assay kit	Thermo Fisher Scientific	Q32850
AMPure XP system	Agencourt	A63880
EasySep Magnet	StemCell Technologies	18000

A2.3 Barcode-incorporating primers for second round PCR of panning samples

The red text indicates barcode sequence, and the highlighted section indicates the linker. All primers were referred to in house as A-key lnk1 primers. Barcode 42 was not used as it caused bleed through from other barcodes.

Primer	Sequence
BC1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTAAGGTAACGTAATCCTTGTGGTATCG
BC2	CCATCTCATCCCTGCGTGTCTCCGACTCAGTAAGGAGAACGTAATCCTTGTGGTATCG
BC3	CCATCTCATCCCTGCGTGTCTCCGACTCAGAAGAGGATTCGTAATCCTTGTGGTATCG
BC4	CCATCTCATCCCTGCGTGTCTCCGACTCAGTACCAAGATCGTAATCCTTGTGGTATCG
BC5	CCATCTCATCCCTGCGTGTCTCCGACTCAGCAGAAGGAACGTAATCCTTGTGGTATCG
BC6	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGCAAGTTCGTAATCCTTGTGGTATCG
BC7	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCGTGATTCGTAATCCTTGTGGTATCG
BC8	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCCGATAACGTAATCCTTGTGGTATCG
BC9	CCATCTCATCCCTGCGTGTCTCCGACTCAGTGAGCGGAACGTAATCCTTGTGGTATCG
BC10	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGACCGAACGTAATCCTTGTGGTATCG
BC11	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCCTCGAATCGTAATCCTTGTGGTATCG
BC12	CCATCTCATCCCTGCGTGTCTCCGACTCAGTAGGTGGTTCGTAATCCTTGTGGTATCG
BC13	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTAACGGACGTAATCCTTGTGGTATCG
BC14	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTGGAGTGTCGTAATCCTTGTGGTATCG
BC15	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTAGAGGTCGTAATCCTTGTGGTATCG
BC16	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTGGATGACGTAATCCTTGTGGTATCG
BC17	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTATTCGTCGTAATCCTTGTGGTATCG
BC18	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGGCAATTGCGTAATCCTTGTGGTATCG
BC19	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTAGTCGGACGTAATCCTTGTGGTATCG
BC20	CCATCTCATCCCTGCGTGTCTCCGACTCAGCAGATCCATCGTAATCCTTGTGGTATCG
BC21	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCGCAATTACGTAATCCTTGTGGTATCG
BC22	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCGAGACGCGTAATCCTTGTGGTATCG
BC23	CCATCTCATCCCTGCGTGTCTCCGACTCAGTGCCACGAACGTAATCCTTGTGGTATCG
BC24	CCATCTCATCCCTGCGTGTCTCCGACTCAGAACCTCATTCGTAATCCTTGTGGTATCG
BC25	CCATCTCATCCCTGCGTGTCTCCGACTCAGCCTGAGATACGTAATCCTTGTGGTATCG
BC26	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTACAACCTCGTAATCCTTGTGGTATCG
BC27	CCATCTCATCCCTGCGTGTCTCCGACTCAGAACCATCCGCGTAATCCTTGTGGTATCG
BC28	CCATCTCATCCCTGCGTGTCTCCGACTCAGATCCGGAATCGTAATCCTTGTGGTATCG
BC29	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCGACCACTCGTAATCCTTGTGGTATCG
BC30	CCATCTCATCCCTGCGTGTCTCCGACTCAGCGAGGTTATCGTAATCCTTGTGGTATCG
BC31	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCCAAGCTGCGTAATCCTTGTGGTATCG
BC32	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTTACACACGTAATCCTTGTGGTATCG
BC33	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCTCATTGAACGTAATCCTTGTGGTATCG
BC34	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCGCATCGTTCGTAATCCTTGTGGTATCG
BC35	CCATCTCATCCCTGCGTGTCTCCGACTCAGTAAGCCATTGTCGTAATCCTTGTGGTATCG
BC36	CCATCTCATCCCTGCGTGTCTCCGACTCAGAAGGAATCGTCGTAATCCTTGTGGTATCG
BC37	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTTGAGAATGTCGTAATCCTTGTGGTATCG
BC38	CCATCTCATCCCTGCGTGTCTCCGACTCAGTGGAGGACGGACGTAATCCTTGTGGTATCG
BC39	CCATCTCATCCCTGCGTGTCTCCGACTCAGTAACAATCGGCGTAATCCTTGTGGTATCG
BC40	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGACATAATCGTAATCCTTGTGGTATCG
BC41	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCCACTTCGCGTAATCCTTGTGGTATCG
BC43	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTTGACACCGCGTAATCCTTGTGGTATCG
BC44	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTGGAGGCCAGCGTAATCCTTGTGGTATCG
BC45	CCATCTCATCCCTGCGTGTCTCCGACTCAGTGGAGCTTCCTCGTAATCCTTGTGGTATCG
BC46	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCAGTCCGAACGTAATCCTTGTGGTATCG
BC47	CCATCTCATCCCTGCGTGTCTCCGACTCAGTAAGGCAACCACGTAATCCTTGTGGTATCG
BC48	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCTAAGAGACGTAATCCTTGTGGTATCG
BC49	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCCTAACATAACGTAATCCTTGTGGTATCG
BC50	CCATCTCATCCCTGCGTGTCTCCGACTCAGCGGACAATGGCGTAATCCTTGTGGTATCG
BC51	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTGAGCCTATTCGTAATCCTTGTGGTATCG
BC52	CCATCTCATCCCTGCGTGTCTCCGACTCAGCCGCATGGAACGTAATCCTTGTGGTATCG

BC53	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGGCAATCCTCGTAATCCTTGTGGTATCG
BC54	CCATCTCATCCCTGCGTGTCTCCGACTCAGCCGGAGAATCGCGTAATCCTTGTGGTATCG
BC55	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCCACCTCCTCGTAATCCTTGTGGTATCG
BC56	CCATCTCATCCCTGCGTGTCTCCGACTCAGCAGCATTAATTCGTAATCCTTGTGGTATCG
BC57	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTGGCAACGGCGTAATCCTTGTGGTATCG
BC58	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCCTAGAACACGTAATCCTTGTGGTATCG
BC59	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCCTTGATGTTCGTAATCCTTGTGGTATCG
BC60	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTAGCTCTTCGTAATCCTTGTGGTATCG
BC61	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCACTCGGATCGTAATCCTTGTGGTATCG
BC62	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCCTGCTTCACGTAATCCTTGTGGTATCG

A.3 Defining consensus sequences for human UBDs – a bioinformatics approach

A3.1 Reanalysis of established UBD family consensus motifs

In an effort to update and consolidate consensus sequences and motifs for the 19 families of human UBDs, lists of human UBD domains were analysed through PROMALS3D and Jalview and motifs determined by manual annotation. Whilst this was a reliable and reproducible method, for some families the motif was already published, well established and still applicable. These data were not presented in the main chapter but were instead collected in the Appendices for completeness.

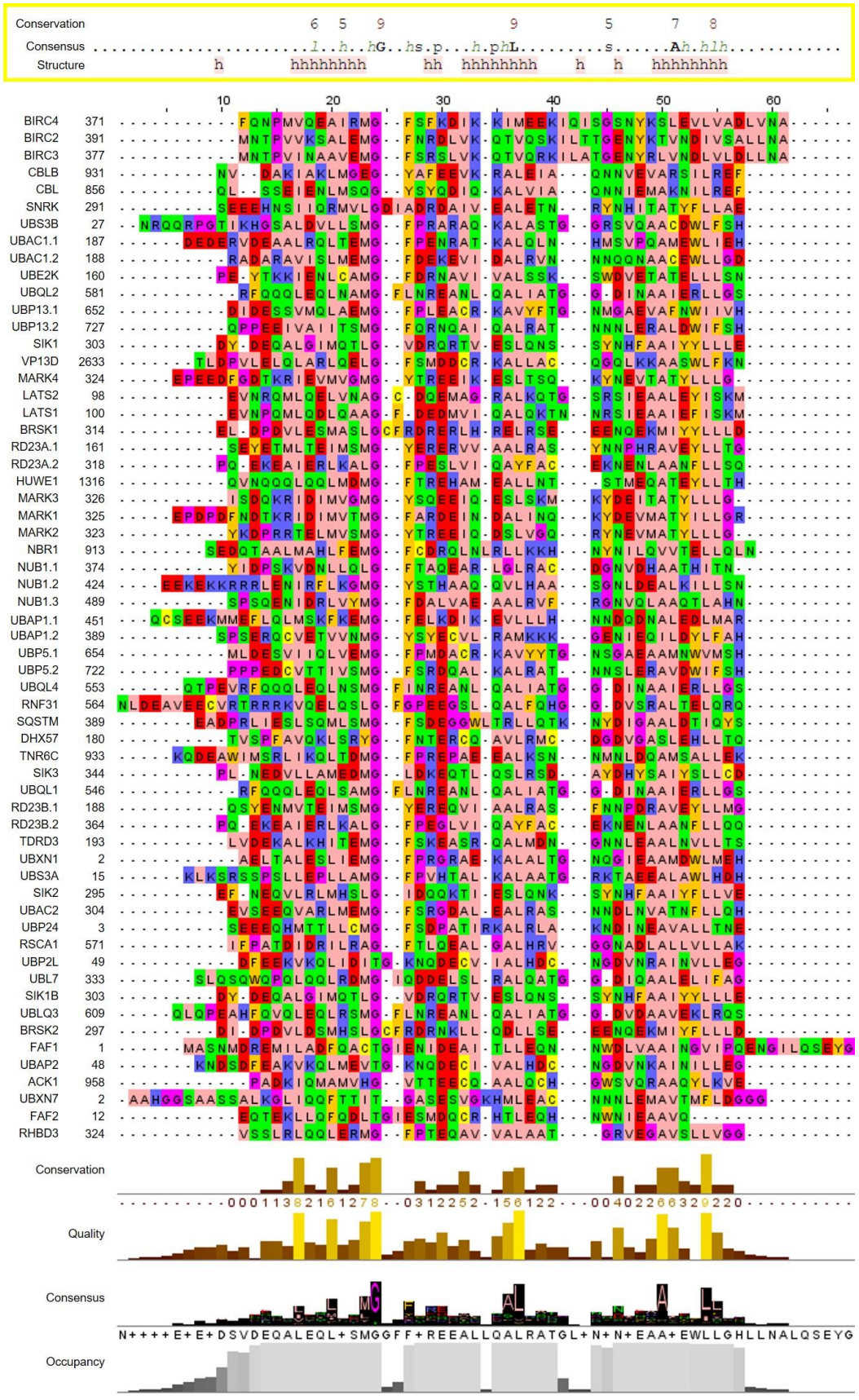

Conservation
Consensus
Structure
Quality
Consensus
Occupancy

Fig.A3.1.1. Structural, physicochemical and consensus alignments of human UBA domains. 57 human UBA domains were determined by cross database comparison, then aligned with PROMALS3D (upper, yellow panel) and Clustal W and X within Jalview (lower panel). Zappo physicochemical properties were indicated by colour within Jalview and alignment conservation, quality, consensus and occupancy calculated automatically. The position of the domain within each protein was indicated on the left by the number of the first amino acid and the relative position within the alignment at the top.

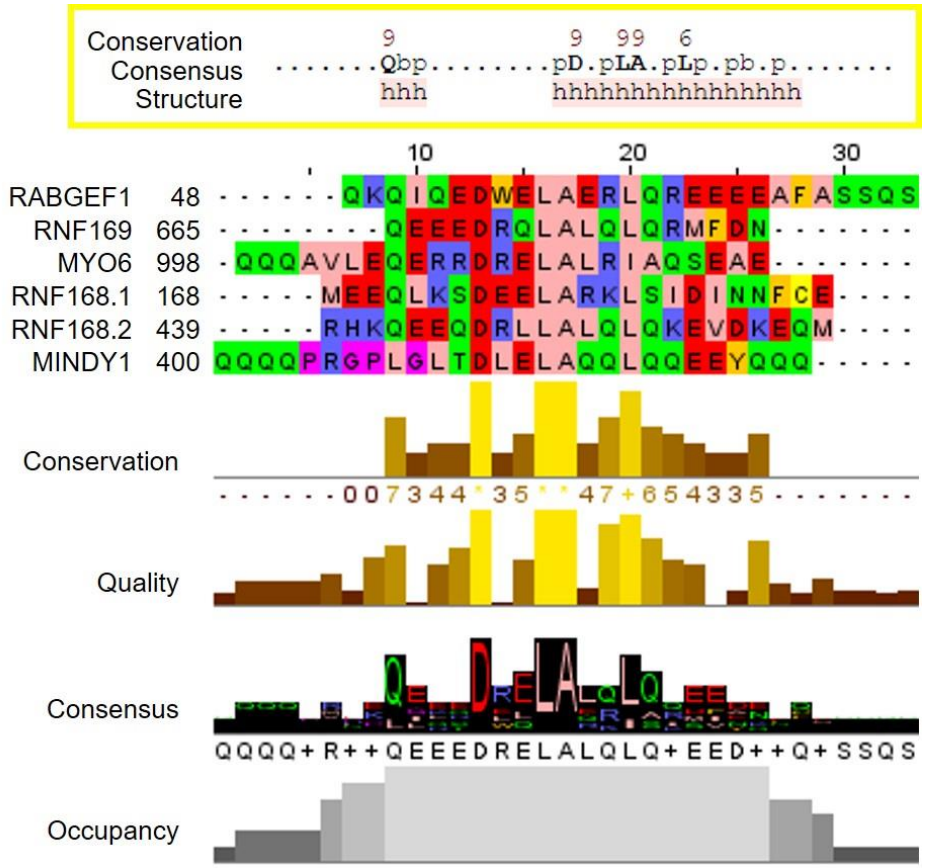


Fig.A3.1.2. Structural, physicochemical and consensus alignments of human MIU domains. Six human MIU domains were determined by cross database comparison, then aligned with PROMALS3D (upper, yellow panel) and Clustal W and X within Jalview (lower panel). Zappo physicochemical properties were indicated by colour within Jalview and alignment conservation, quality, consensus and occupancy calculated automatically. The position of the domain within each protein was indicated on the left by the number of the first amino acid and the relative position within the alignment at the top.

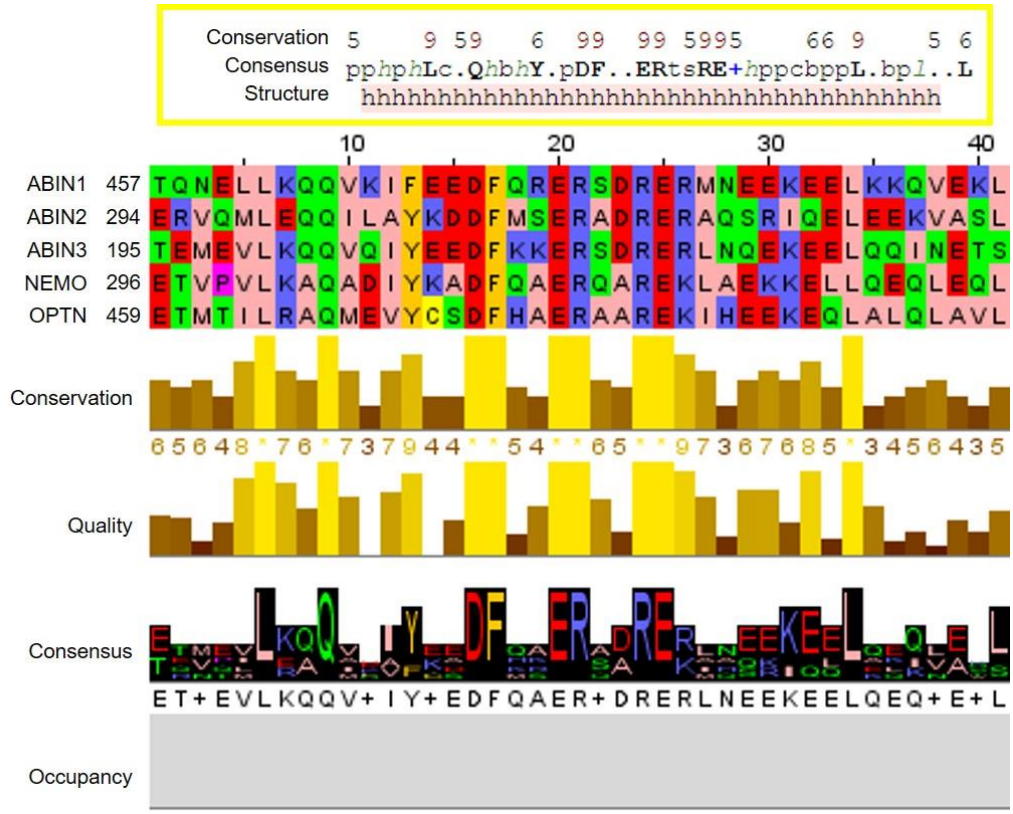


Fig.A3.1.3. Structural, physicochemical and consensus alignments of human UBAN domains. Five human UBAN domains were determined by cross database comparison, then aligned with PROMALS3D (upper, yellow panel) and Clustal W and X within Jalview (lower panel). Zappo physicochemical properties were indicated by colour within Jalview and alignment conservation, quality, consensus and occupancy calculated automatically. The position of the domain within each protein was indicated on the left by the number of the first amino acid and the relative position within the alignment at the top.

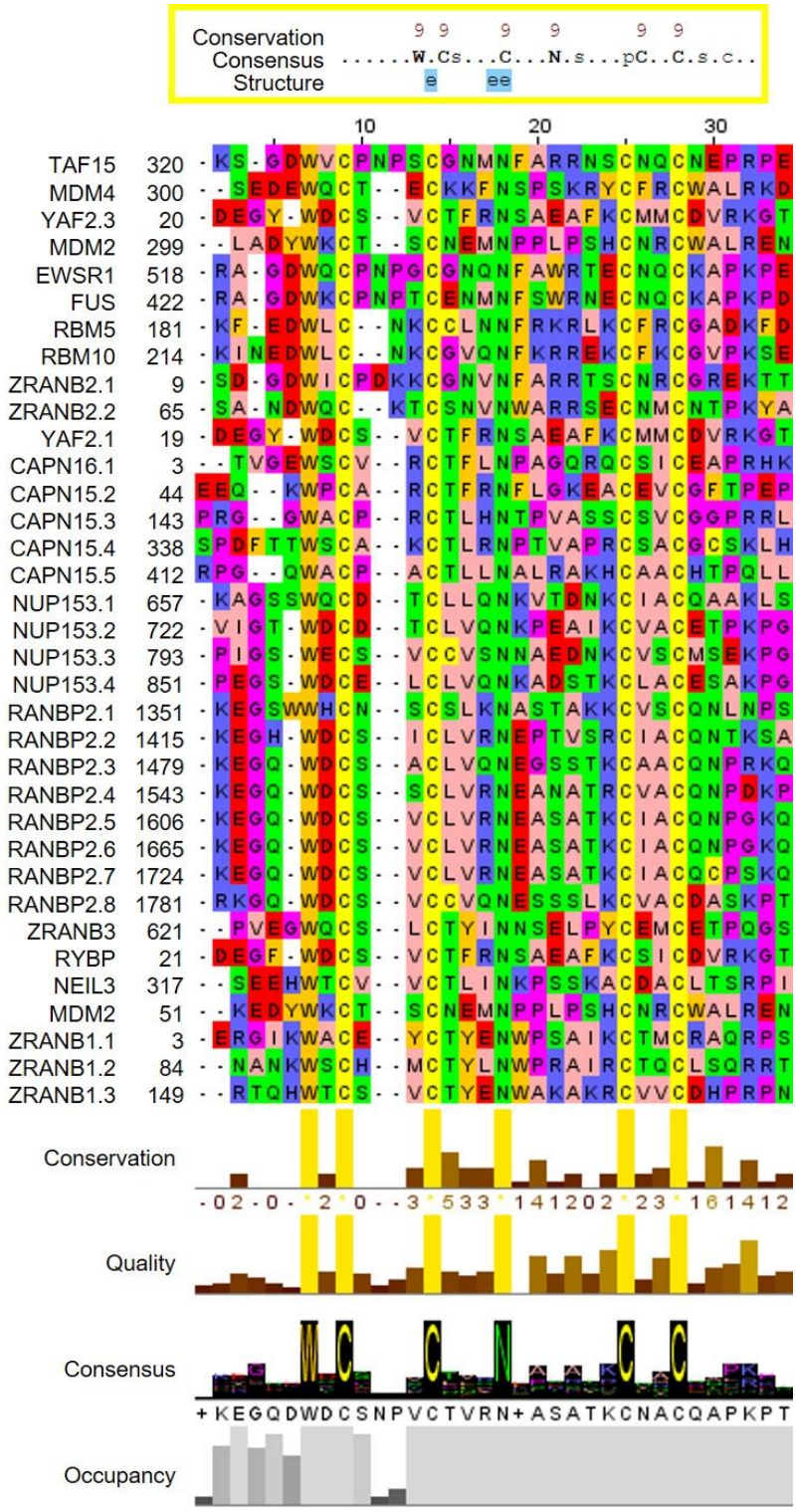


Fig.A3.1.4. Structural, physicochemical and consensus alignments of human NZF domains. 35 human NZF domains were determined by cross database comparison, then aligned with PROMALS3D (upper, yellow panel) and Clustal W and X within Jalview (lower panel). Zappo physicochemical properties were indicated by colour within Jalview and alignment conservation, quality, consensus and occupancy calculated automatically. The position of the domain within each protein was indicated on the left by the number of the first amino acid and the relative position within the alignment at the top.

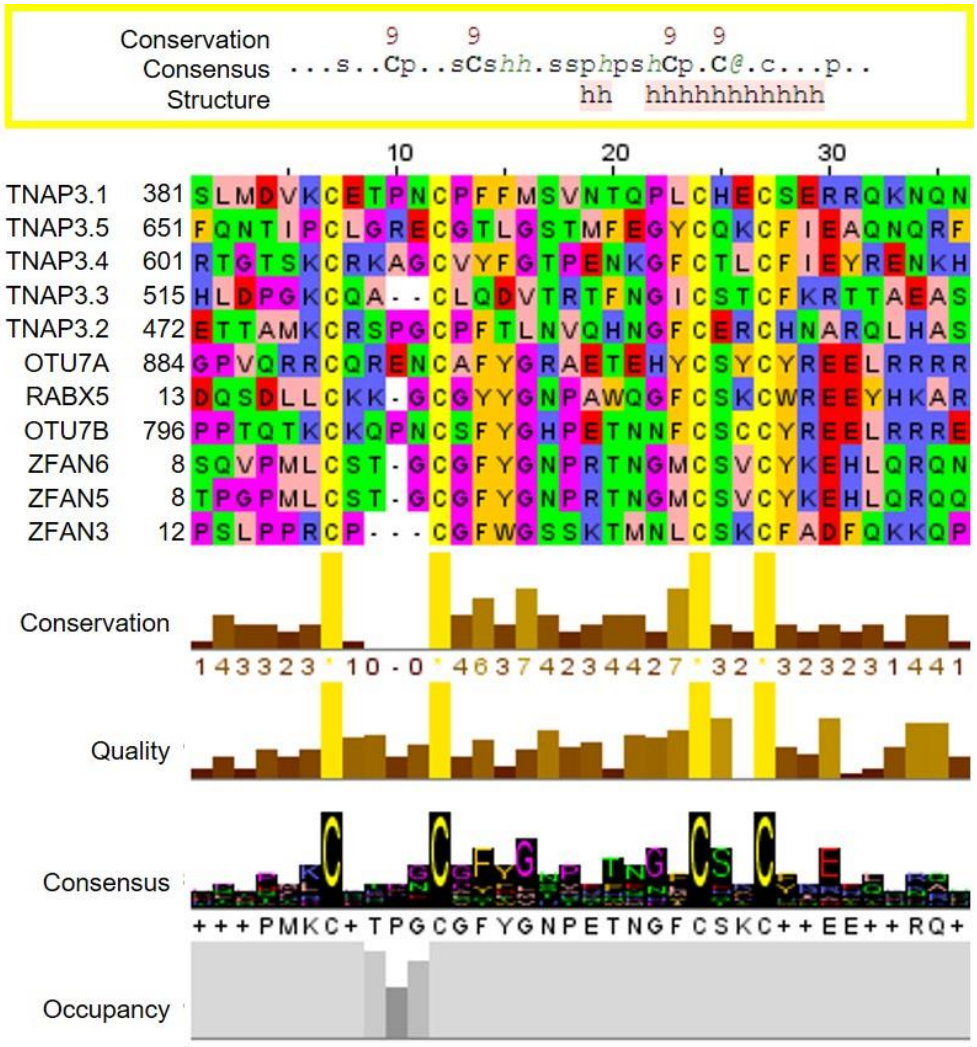


Fig.A3.1.5. Structural, physicochemical and consensus alignments of human ZnF_A20 domains. 11 human ZnF_A20 domains were determined by cross database comparison, then aligned with PROMALS3D (upper, yellow panel) and Clustal W and X within Jalview (lower panel). Zappo physicochemical properties were indicated by colour within Jalview and alignment conservation, quality, consensus and occupancy calculated automatically. The position of the domain within each protein was indicated on the left by the number of the first amino acid and the relative position within the alignment at the top.

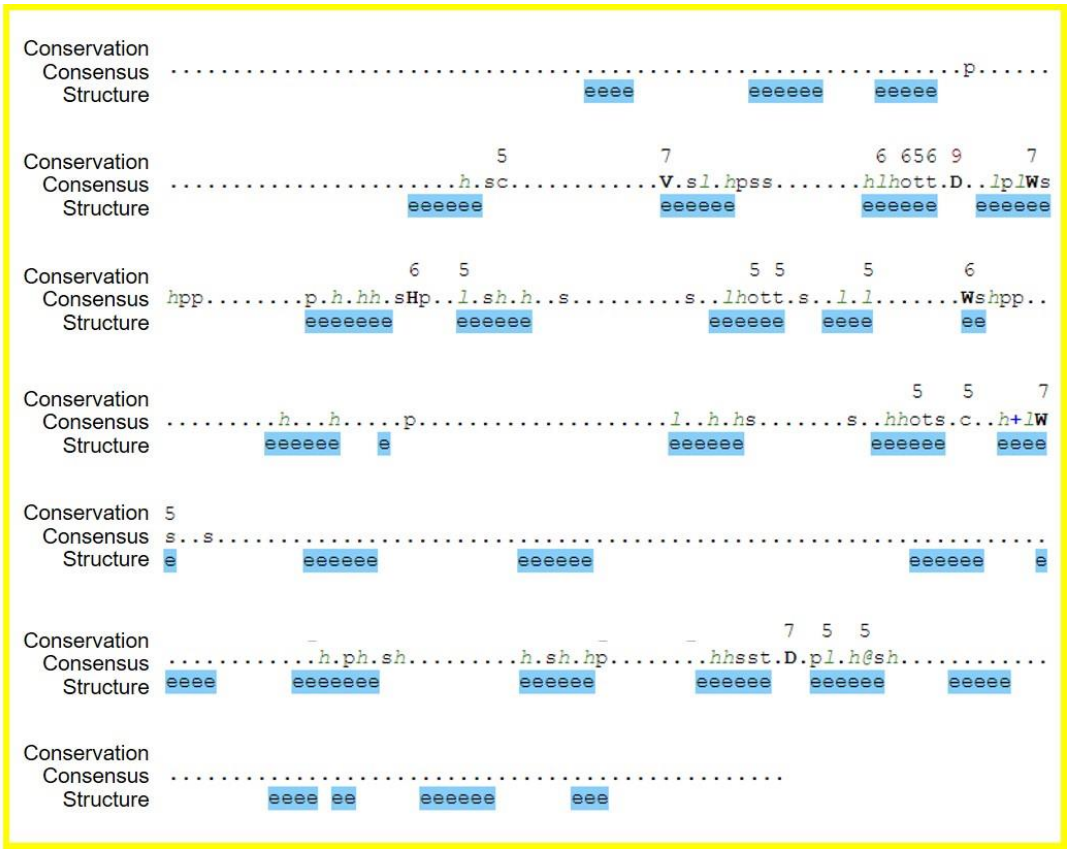


Fig.A3.1.6. Structural, physicochemical and consensus alignments of human beta-Prp domains. 36 human beta-Prp domains were determined by cross database comparison, then aligned with PROMALS3D.

A.4 Next-Generation Phage Display targeting monoubiquitin towards generating novel ubiquitin-binding candidate peptides

A4.1 Submitted NGS samples and yield

Barcode	Set/ Round	Sample	pfu/ml	x10^	Bin	dsDNA (ng)	Reads	Sequences
1	S1/R1	WT	1.94	5	6	15.80	60,525	22,115
2	S1/R1	S57E	2.41	3	3	7.90	25,106	3,266
3	S1/R1	S65E	9.70	3	4	10.53	28,995	4,654
4	S1/R1	Marvel	1.95	4	5	13.16	45,178	9,706
5	S1/R2	WT	9.30	3	4	10.53	37,907	4,302
6	S1/R2	WT Marvel	3.98	5	7	18.43	97,467	30,778
7	S1/R2	S57E	2.02	4	5	13.16	51,168	3,979
8	S1/R2	S57E-Sub	7.30	4	6	15.80	69,886	5,126
9	S1/R2	S57E Marvel	5.70	3	4	10.53	40,763	3,124
10	S1/R2	S65E	4.70	3	4	10.53	28,978	2,633
11	S1/R2	S65E-Sub	9.20	3	4	10.53	39,712	1,067
12	S1/R2	S65E Marvel	1.56	4	5	13.16	44,998	1,301
13	S1/R3	WT	1.90	2	2	5.27	31,716	2,136
14	S1/R3	WT Marvel	4.37	5	7	18.43	71,481	7,879
15	S1/R3	S57E	8.00	1	2	5.27	42,075	1,077
16	S1/R3	S57E Marvel	3.50	4	5	13.16	106,130	7,322
17	S1/R3	S57E-Sub	6.70	2	3	7.90	37,961	1,838
18	S1/R3	S57E-Sub Marvel	3.00	1	1	2.63	5,886	169
19	S1/R3	S65E	8.00	1	2	5.27	26,376	702
20	S1/R3	S65E Marvel	1.39	5	6	15.80	49,643	8,861
21	S1/R3	S65E-Sub	1.20	2	2	5.27	23,587	711
22	S1/R3	S65E-Sub Marvel	2.00	1	1	2.63	10,379	379
23	S2/R1	WT	3.00	3	4	10.53	44,073	4,714
24	S2/R1	S65E	7.30	3	4	10.53	35,582	6,100
25	S2/R1	pS65	3.70	3	4	10.53	54,351	5,564
26	S2/R1	Marvel	1.61	5	6	15.80	37,761	14,066
27	S2/R2	WT	2.78	4	5	13.16	45,642	4,441
28	S2/R2	WT	2.27	5	6	15.80	60,475	5,608
29	S2/R2	WT Marvel	1.89	4	5	13.16	37,902	3,982
30	S2/R2	WT Marvel	4.30	4	6	15.80	45,509	4,501
31	S2/R2	S65E	2.47	4	5	13.16	58,173	8,320
32	S2/R2	S65E	1.90	2	2	5.27	33,160	920
33	S2/R2	S65E-Sub	3.23	4	5	13.16	57,389	8,712
34	S2/R2	S65E-Sub	7.00	1	2	5.27	23,687	613
35	S2/R2	S65E Marvel	6.70	2	3	7.90	30,283	1,617

36	S2/R2	S65E Marvel	6.70	3	4	10.53	48,059	1,625
37	S2/R2	pS65	1.85	5	6	15.80	55,591	6,653
38	S2/R2	pS65	5.70	3	4	10.53	45,225	4,274
39	S2/R2	pS65-Sub	4.50	3	4	10.53	45,807	4,127
40	S2/R2	pS65-Sub	5.50	3	4	10.53	34,690	3,486
41	S2/R2	pS65 Marvel	3.40	4	5	13.16	35,054	4,069
42	S2/R2	pS65 Marvel	3.60	4	6	15.80	37,656	5,180
43	S2/R3	WT	5.50	3	4	10.53	44,716	4,381
44	S2/R3	WT	7.00	3	4	10.53	38,409	4,701
45	S2/R3	WT Marvel	7.90	3	4	10.53	35,414	2,059
46	S2/R3	WT Marvel	4.40	3	4	10.53	40,032	3,946
47	S2/R3	S65E	1.00	1	1	2.63	14,813	258
48	S2/R3	S65E	5.70	3	4	10.53	43,474	4,509
49	S2/R3	S65E Marvel	2.00	2	2	5.27	26,311	977
50	S2/R3	S65E Marvel	6.80	3	4	10.53	28,849	3,971
51	S2/R3	S65E-Sub	4.50	3	4	10.53	44,523	4,451
52	S2/R3	S65E-Sub	4.20	3	4	10.53	32,165	3,652
53	S2/R3	S65E-Sub Marvel	1.52	3	3	7.90	33,518	2,755
54	S2/R3	S65E-Sub Marvel	4.20	2	3	7.90	36,382	1,769
55	S2/R3	pS65	1.38	3	3	7.90	22,775	2,123
56	S2/R3	pS65	4.00	1	2	5.27	19,263	571
57	S2/R3	pS65 Marvel	3.00	3	3	7.90	100,163	5,913
58	S2/R3	pS65 Marvel	4.80	4	6	15.80	51,549	6,218
59	S2/R3	pS65-Sub	1.22	3	3	7.90	24,554	2,227
60	S2/R3	pS65-Sub	1.47	4	4	10.53	30,178	4,368
61	S2/R3	pS65-Sub Marvel	1.54	4	5	13.16	53,595	5,964
62	S2/R3	pS65-Sub Marvel	4.60	2	3	7.90	19,492	1,176

Table A4.1. The samples submitted to NGS and the sequencing yield per sample. The pfu/mL of NGPD-derived samples was recorded and used to determine the weighted amount of dsDNA submitted to NGS. This resulted in a varying number of reads and sequences per sample.

A4.2 ^{1}H-^{15}N HSQC protein NMR spectra relating to ^{15}N-ubiquitin

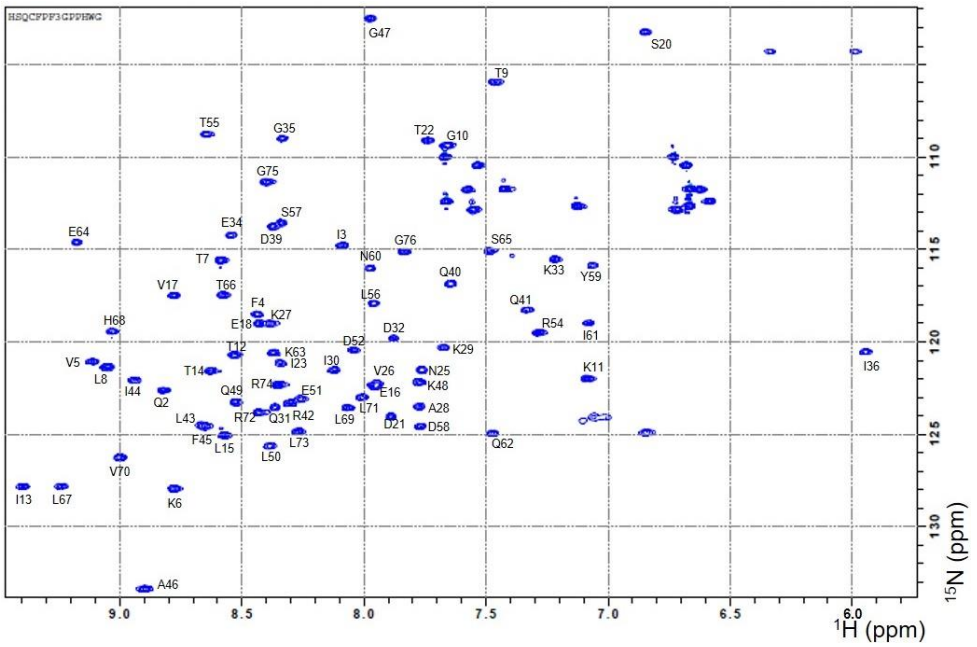


Fig.A4.2.1. The assigned ^{1}H-^{15}N HSQC spectra for ^{15}N-ubiquitin.

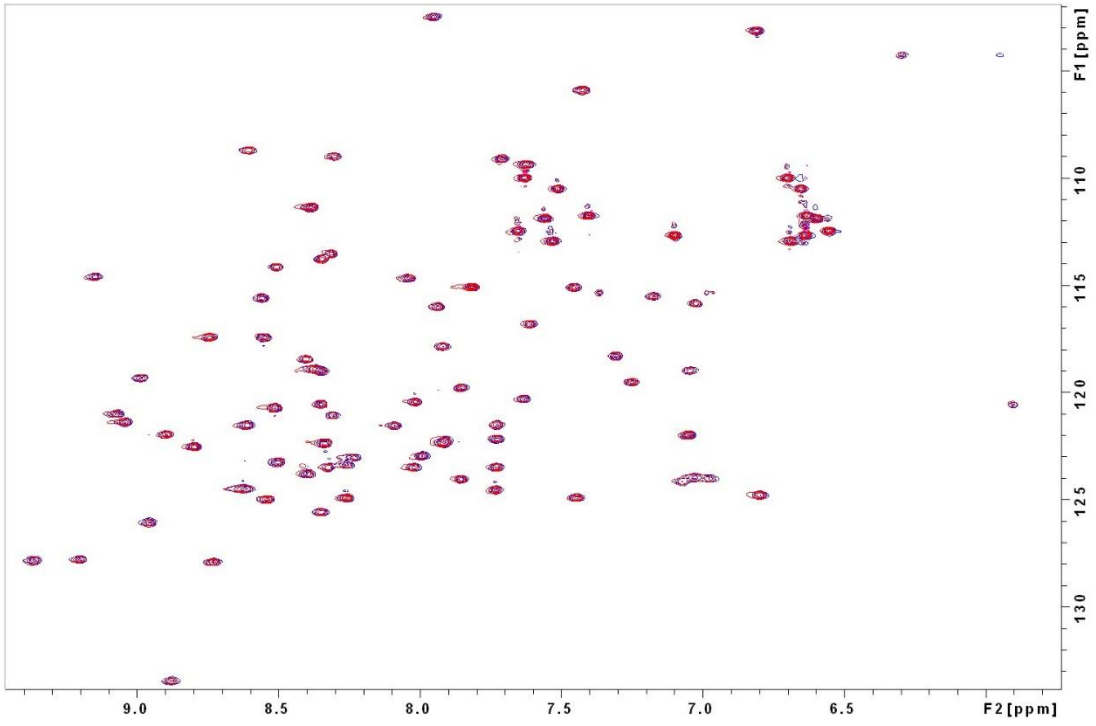


Fig.A4.2.3. The overlaid ^{1}H-^{15}N HSQC NMR spectra for ^{15}N-ubiquitin in the presence and absence of an MIU. Spectra for 50µM ^{15}N-ubiquitin sampled at 281K on a Bruker Ascend^{TM} 800 MHz spectrometer in the absence (blue) and presence (red) of 135µM Rabex-5-MIU. Data was captured as 2D ^{1}H-^{15}N HSQC NMR in 10% D_2O pH 7.5. The reference point was the water line.

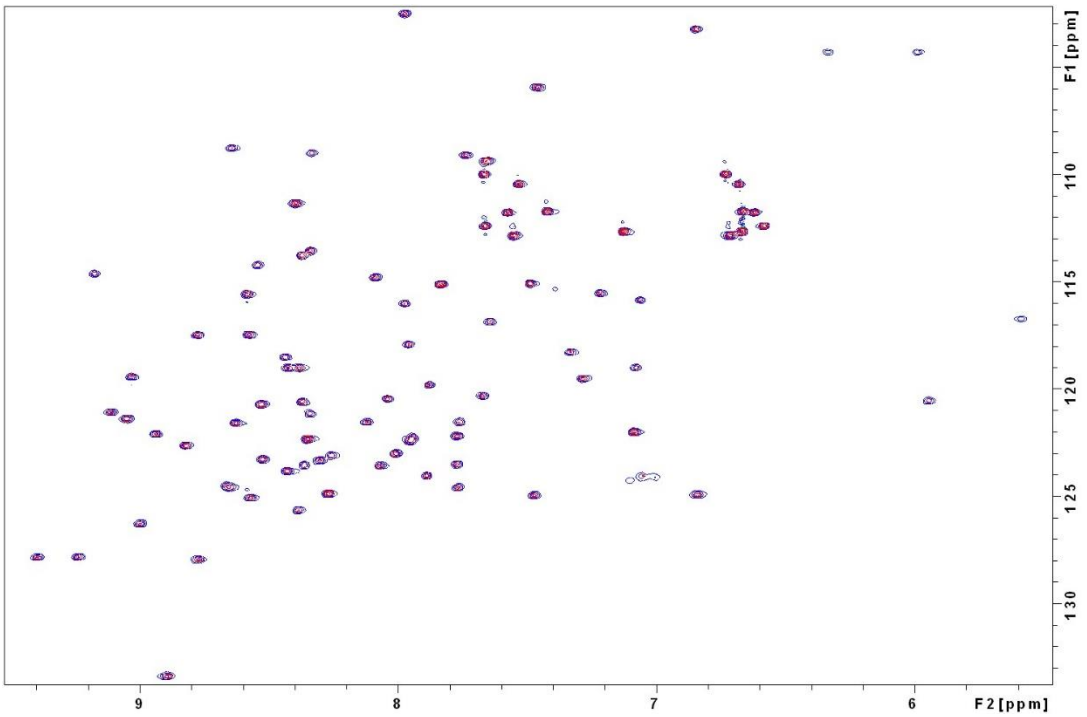


Fig.A4.2.4. The overlaid ^{1}H-^{15}N HSQC NMR spectra for ^{15}N-ubiquitin in the presence and absence of an MIU. Spectra for 50μM ^{15}N-ubiquitin sampled at 281K on a Bruker Ascend^{TM} 800 MHz spectrometer in the absence (blue) and presence (red) of 400μM Peptide 45. Data was captured as 2D ^{1}H-^{15}N HSQC NMR in 10% D_2O pH 7.5. The reference point was the water line.

A.5 Proteomic Identification of Ubiquitin-Binding Domains and Ubiquitin-Binding Proteins

A5.1 Proteins affinity chromatography purifying with ubiquitin-Sepharose in the presence and absence of calcium

6.7mg of total protein from pig brain lysate was captured on Sepharose coupled to human ubiquitin at 2mg/ml for 2hrs with end-over-end rotation at room temperature in the presence or absence of 250μM $CalCl_2$. Captured proteins were identified by LC-MS/MS and mapped to human homologues. The proteins that occurred in both conditions were discussed in the main text. Proteins that only bound ubiquitin-Sepharose in a single condition are listed below.

Table A5.1.1. Proteins that bound ubiquitin-Sepharose in the absence of calcium only:

Ubiquitin-Sepharose			
Total Spectrum Count	Gene	Protein	
4	RPL5	RL5	60S ribosomal protein L5
4	PSMD7	PSMD7	26S proteasome non-ATPase regulatory subunit 7
4	TIPRL	TIPRL	TIP41-like protein
3	ACADVL	ACADV	Very long-chain specific acyl-CoA dehydrogenase, mitochondrial
3	HEBP2	HEBP2	Heme-binding protein 2
2	AP1M1	AP1M1	AP-1 complex subunit mu-1
2	MTHFD1	C1TC	C-1-tetrahydrofolate synthase, cytoplasmic
2	RABGGTA	PGTA	Geranylgeranyl transferase type-2 subunit alpha
2	SDCBP	SDCB1	Syntenin-1
2	DCUN1D1	DCNL1	DCN1-like protein 1
2	CRPPA	ISPD	D-ribitol-5-phosphate cytidylyltransferase
2	PSMD5	PSMD5	26S proteasome non-ATPase regulatory subunit 5
2	PRXL2B	PXL2B	Prostamide/prostaglandin F synthase
2	SLC25A1	TXTP	Tricarboxylate transport protein, mitochondrial
2	EIF3J	EIF3J	Eukaryotic translation initiation factor 3 subunit J
2	UBE2K	UBE2K	Ubiquitin-conjugating enzyme E2 K
2	RPL13A	RL13A	60S ribosomal protein L13a
2	RRBP1	RRBP1	Ribosome-binding protein 1
1	TARS	SYTC	Threonine--tRNA ligase, cytoplasmic
1	PSMD12	PSD12	26S proteasome non-ATPase regulatory subunit 12
1	RPS5	RS5	40S ribosomal protein S5
1	VAPA	VAPA	Vesicle-associated membrane protein-associated protein A
1	DNAJB4	DNJB4	DnaJ homolog subfamily B member 4
1	DDX39B	DX39B	Spliceosome RNA helicase DDX39B
1	AKAP5	AKAP5	A-kinase anchor protein 5

1	CAPNS1	CPNS1	Calpain small subunit 1
1	SRI	SORCN	Sorcin
1	USO1	USO1	General vesicular transport factor p115
1	DHX9	DHX9	ATP-dependent RNA helicase A
1	COPA	COPA	Coatomer subunit alpha
1	ERP44	ERP44	Endoplasmic reticulum resident protein 44
1	VBP1	PFD3	Prefoldin subunit 3
1	LRRC47	LRC47	Leucine-rich repeat-containing protein 47
1	NADK2	NAKD2	NAD kinase 2, mitochondrial
1	HEPACAM	HECAM	Hepatocyte cell adhesion molecule
1	VPS29	VPS29	Vacuolar protein sorting-associated protein 29
1	ACAD10	ACD10	Acyl-CoA dehydrogenase family member 10
1	NAXE	NNRE	NAD(P)H-hydrate epimerase
1	VARS	SYVC	Valine--tRNA ligase
1	ERC2	ERC2	ERC protein 2
1	GPD1	GPDA	Glycerol-3-phosphate dehydrogenase [NAD(+)], cytoplasmic
1	ANLN	ANLN	Anillin

Table A5.1.2. Proteins that specifically bound ubiquitin-Sepharose in the presence of calcium only:

Ubiquitin-Sepharose + Calcium			
Total Spectrum Count	Gene	Protein	
5	RAB15	RAB15	Ras-related protein Rab-15
3	AKR7L	ARK74	Aflatoxin B1 aldehyde reductase member 4
3	GAS7	GAS7	Growth arrest-specific protein 7
3	APPL1	DP13A	DCC-interacting protein 13-alpha
2	RAP1A	RAP1A	Ras-related protein Rap-1A
2	HRNR	HORN	Hornerin
2	ECI2	ECI2	Enoyl-CoA delta isomerase 2, mitochondrial
2	KIF1A	KIF1A	Kinesin-like protein KIF1A
2	SCN2B	SCN2B	Sodium channel subunit beta-2
2	LYZ	LYSC	Lysozyme C
2	GPX4	GPX4	Phospholipid hydroperoxide glutathione peroxidase
2	CAPN2	CAN2	Calpain-2 catalytic subunit
2	UBL7	UBL7	Ubiquitin-like protein 7
2	MT3	MT3	Metallothionein-3
2	NDUFA7	NDUA7	NADH dehydrogenase
2	ENOPH1	ENOPH	Enolase-phosphatase E1
2	EIF5	IF5	Eukaryotic translation initiation factor 5
2	SMAP1	SMAP1	Stromal membrane-associated protein 1
2	ABCE1	ABCE1	ATP-binding cassette sub-family E member 1
1	PGM3	AGM1	Phosphoacetylglucosamine mutase
1	RPS7	RS7	40S ribosomal protein S7

A5.2 Using NMR to investigate the interaction between peflin, ubiquitin and calcium

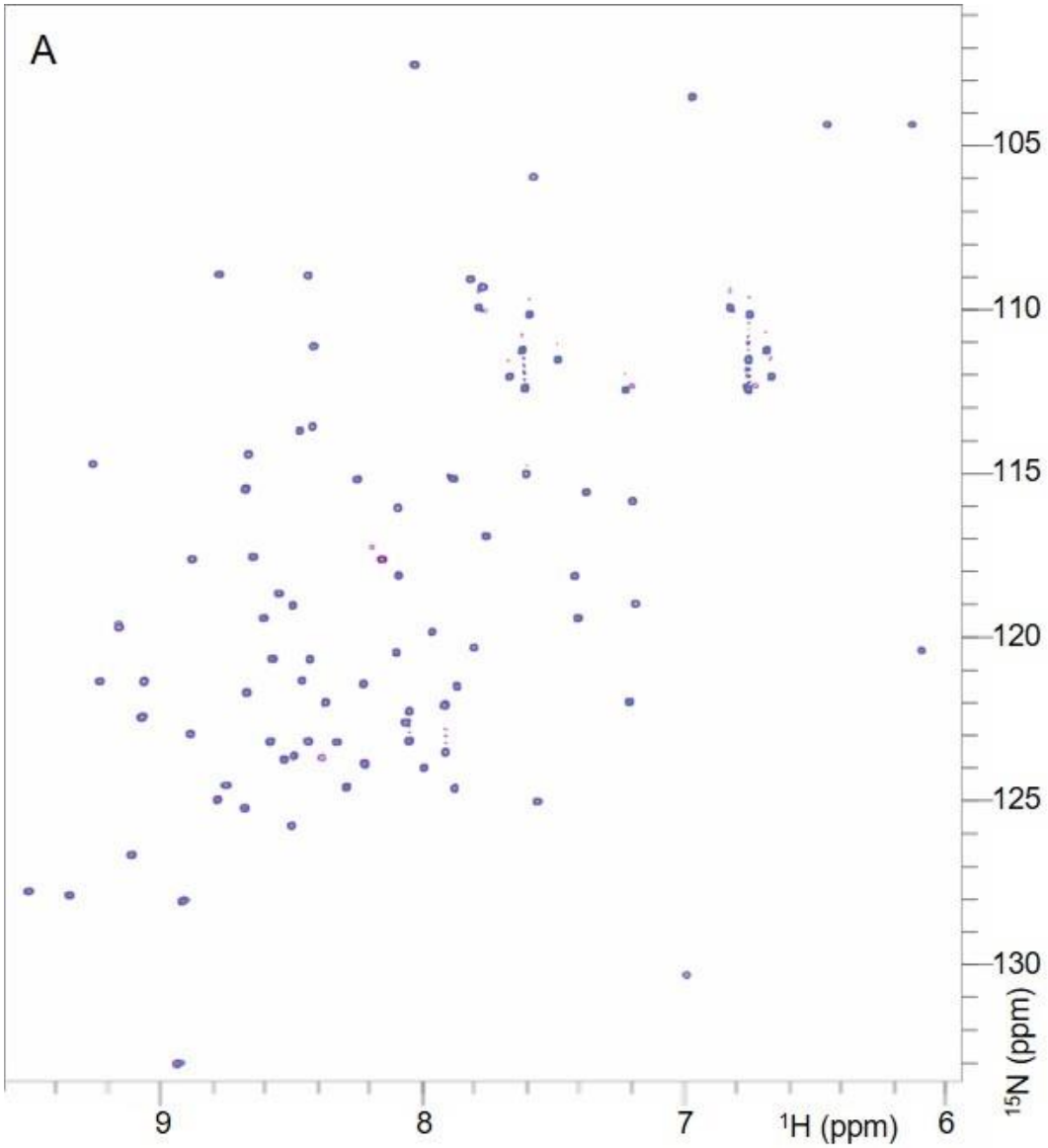


Fig.A5.2.1. The effect of GST-peflin on ^{15}N-ubiquitin NMR spectra. HSQC spectra of 50μM ^{15}N-ubiquitin were acquired at 298K in minimal capture buffer with 10% D_2O in the absence (blue) and presence (purple) of 400μM GST-peflin.

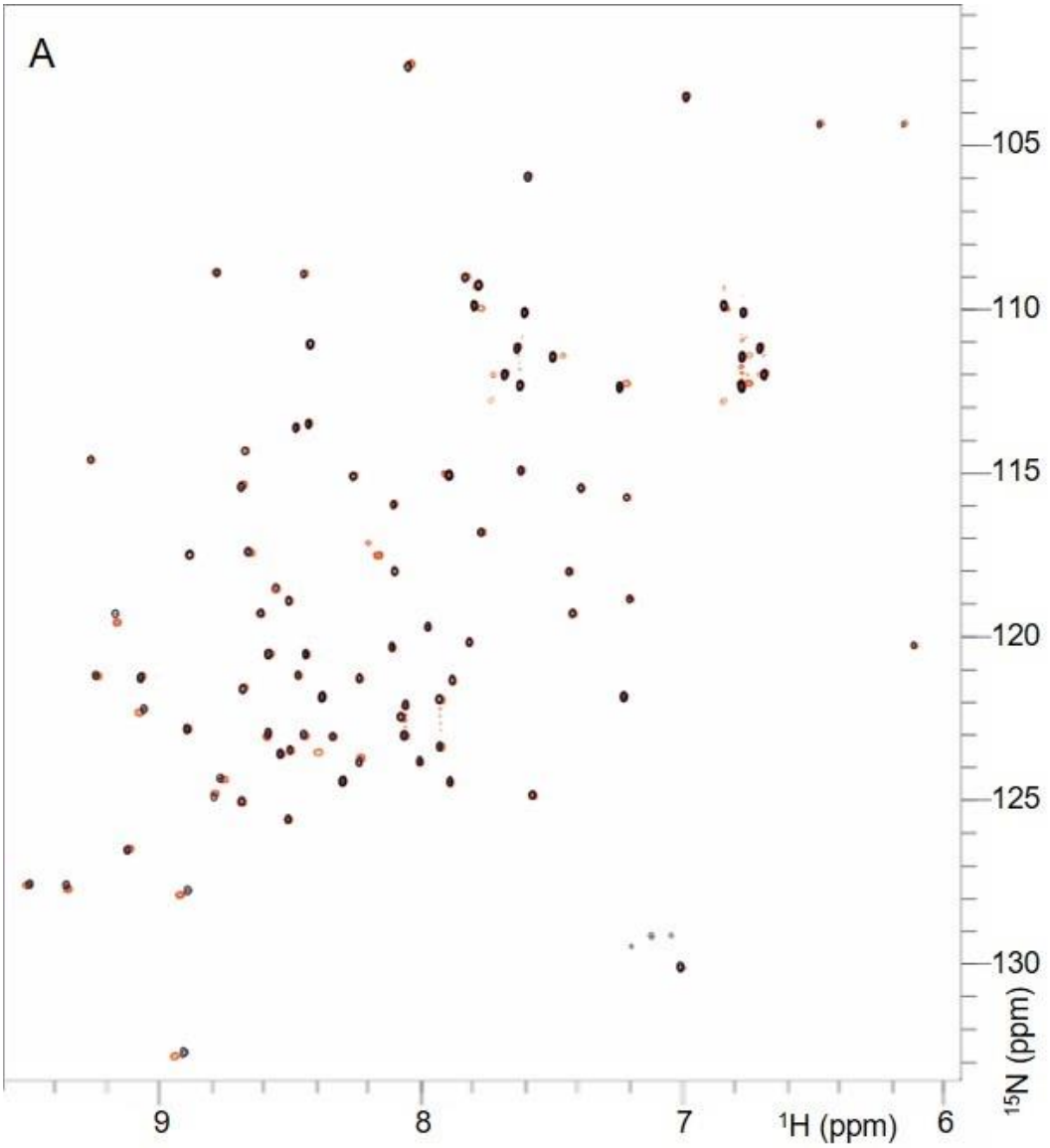


Fig.A5.2.2. The effect of GST-peflin on ^{15}N-ubiquitin NMR spectra in the presence of calcium. HSQC spectra of 50μM ^{15}N-ubiquitin were acquired at 298K in minimal capture buffer with 10% D_2O and 250μM $CaCl_2$ in the absence (black) and presence (red) of 400μM GST-peflin.

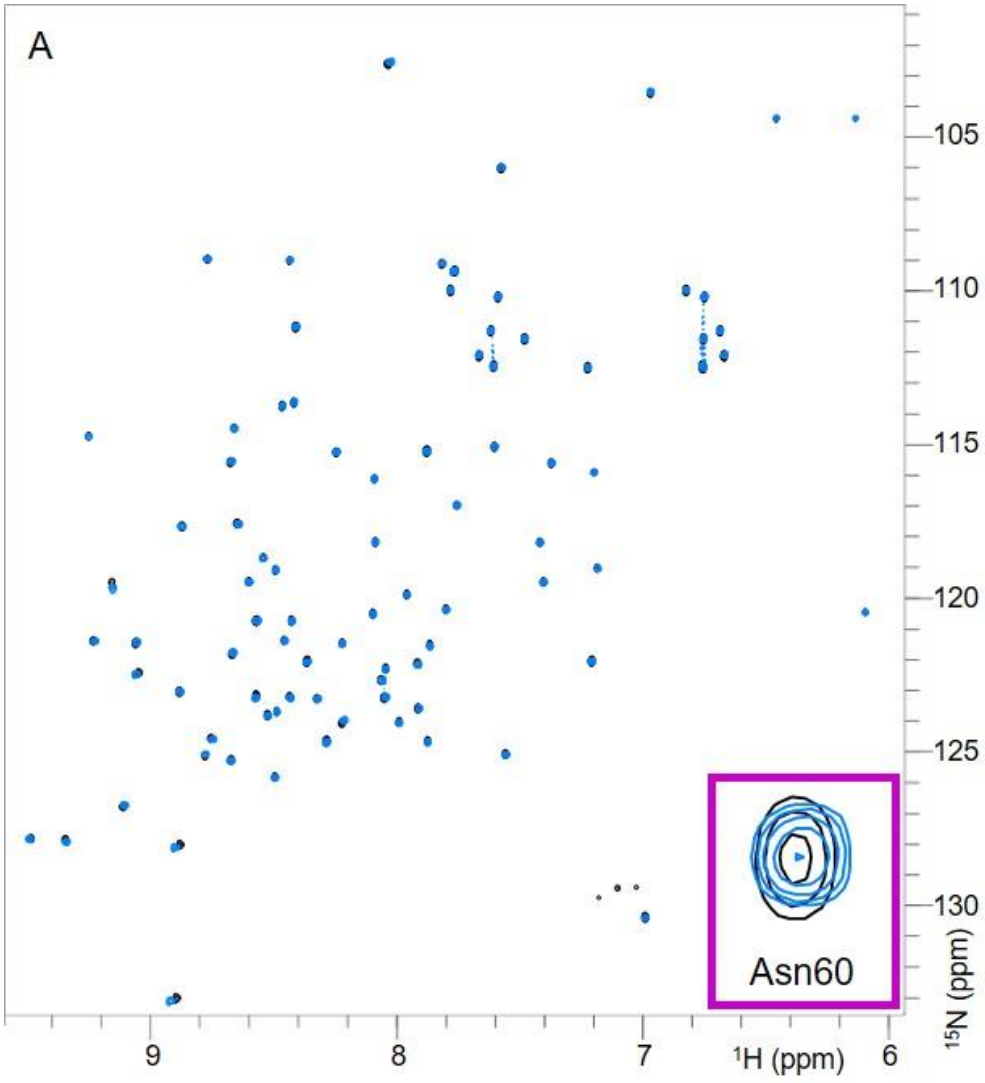


Fig.A5.2.3. The effect of calcium on ^{15}N-ubiquitin NMR spectra. HSQC spectra of 50μM ^{15}N-ubiquitin were acquired at 298K in minimal capture buffer with 10% D_2O in the absence (black) and presence (red) of 250μM $CaCl_2$. A single residue is shown to demonstrate the average CSP seen at most residues (inset).

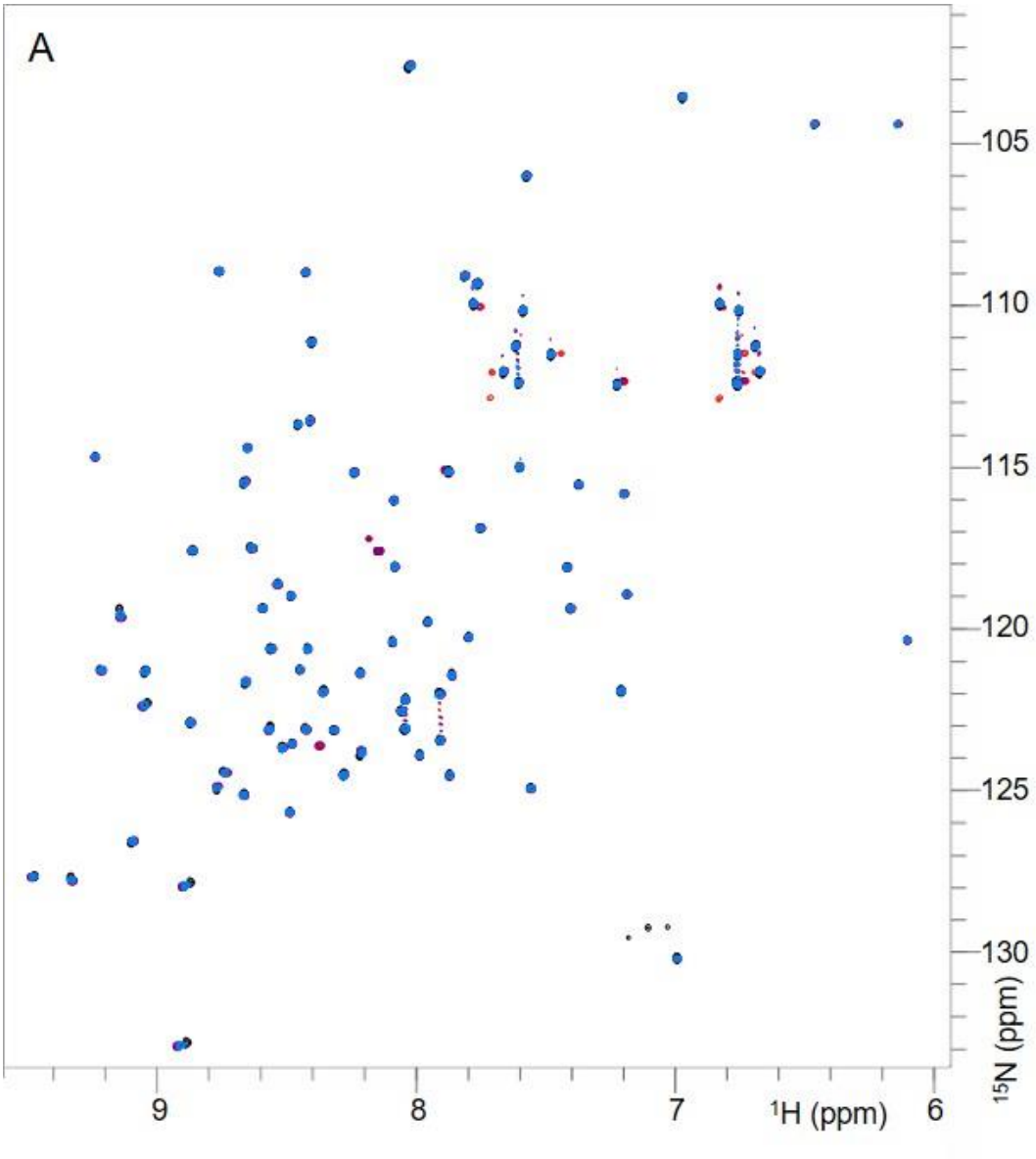


Fig.A5.2.4. The effect of both GST-peflin and calcium on ^{15}N-ubiquitin NMR spectra. HSQC spectra of 50µM ^{15}N-ubiquitin were acquired at 298K in minimal capture buffer with 10% D_2O with 400µM GST-peflin or 250µM $CaCl_2$ as appropriate. Four conditions were sampled; ^{15}N-ubiquitin alone (blue), ^{15}N-ubiquitin with calcium (black), ^{15}N-ubiquitin with GST-peflin (purple) and ^{15}N-ubiquitin with calcium and GST-peflin (red).

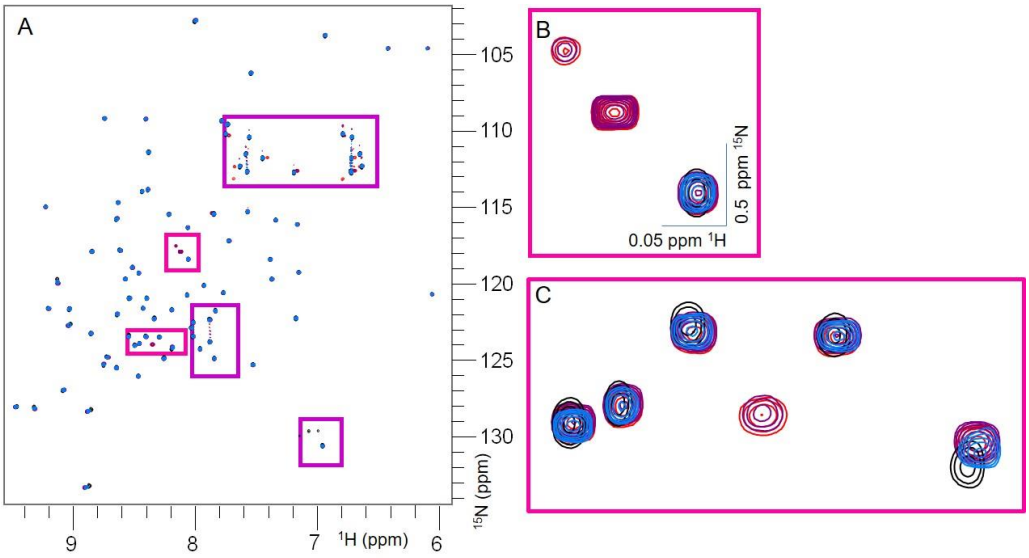


Fig.A5.2.5. Unexplained peak and sidechain movement in ^{15}N-ubiquitin NMR spectra under a variety of conditions. HSQC spectra for 50µM ^{15}N-ubiquitin in minimal capture buffer with 10% D_2O at 298K were captured under four conditions; ^{15}N-ubiquitin only (blue), +250µM $CaCl_2$ (black), +400µM GST-peflin (purple) and +400µM GST-peflin with 250µM $CaCl_2$ (red). Sidechain movement was to be expected and is seen on the right-hand side of the spectra (A; purple boxes). Additional peaks were also seen in the +GST-peflin ±$CaCl_2$ conditions (A, B and C; pink boxes).

These could represent ubiquitin sidechains that moved after addition of GST-peflin, however sidechains are not generally expected to appear in those sections of the spectra. Whilst sidechains are generally more mobile than the amino acid backbone peaks, they tend to appear on the right-hand side of the spectra, as can be seen in Fig.5.8.10-A (purple boxes). Instead this could be explained by some previously undetectable side chains being involved in hydrogen bonding interactions. This would mean that the -NH groups were no longer exchanging rapidly with the solvent and were more easily detectable by NMR. This could also apply to some backbone -NH groups in flexible regions of the protein structure or within the N- or C-termini that could become more ordered upon binding.

Alternatively, these ‘new’ peaks could indicate that there were two conformations of GST-peflin-bound ubiquitin present; one with the slightly shifted peaks discussed above and another where three peaks shifted so dramatically that it was no longer obvious to which residue they should be assigned. This scenario is unlikely, however, as no residues showed a drastic drop in peak height or volume corresponding to a ~50% occupancy. The final possible explanation is

that the GST-peflin itself contained some endogenous ^{15}N. This was highly unlikely; not only were the GST-peflin cultures prepared in a laboratory that does not use ^{15}N-labelling, but also the ‘new’ peaks are more distinct than what would be expected from background ^{15}N signals. In short, no acceptable reason could be identified for the appearance of the additional peaks in the GST-peflin conditions, although several possibilities exist.

Yeah, I got lab problems, that's the human in me
Bling bling, then I solve 'em, that's the scientist in me