Recognition of unanchored polyubiquitin by natural and engineered ubiquitin-binding proteins

Daniel Scott BSc.

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

July 2016

Abstract

The covalent post-translational modification of selected substrates with the ubiquitin protein has emerged as a central regulatory mechanism, governing protein stability, activity and localisation, and accordingly an array of cellular processes. Ubiquitin signalling versatility arises owing to the diverse nature of (poly)ubiquitin modification, with distinct modifications subsequently transduced in a specific manner by ubiquitin-binding domains found in ubiquitin-binding proteins. In recent years the notion that ubiquitin exerts influence solely via the covalent modification of substrates has been challenged, with unanchored, or substrate-free polyubiquitin chains emerging as key regulators of cellular physiology.

The investigations described in this thesis seek to exploit the inherent selectivity of natural and engineered ubiquitin-binding proteins, to afford the purification of endogenous unanchored polyubiquitin, probing the molecular composition and interactions of this biologically significant ubiquitin pool. In *chapter 3* by utilising the deubiquitinating enzyme USP5, which contains multiple ubiquitin-binding domains, and is normally responsible for the selective disassembly of unanchored ubiquitin, we purify unanchored polyubiquitin from mammalian cell extracts. Subsequently, we apply both ubiquitin-selective antibodies and mass spectrometry-based analyses to examine the polyubiquitin profile of the mammalian unanchored ubiquitin pool. In *chapter 4* we then assess the mechanisms of ubiquitin recognition by USP5, presenting a structural mass spectrometry-based framework to probe and quantify ubiquitin: ubiquitin-binding domain interactions. Finally in *chapter 5*, based on the conclusions we draw from

USP5-ubiquitin recognition in *chapter 4*, that multiple domains in suitable arrangement yield specificity for polyubiquitin chains, we design and synthesize a synthetic protein to favour the capture of unanchored polyubiquitin chains of defined topology, namely Lys-48 linked diubiquitin (and longer polyubiquitin chains), from mammalian cell extracts.

We conclude that strategies for the rational design and engineering of polyubiquitin chain-selective binding in non-biological polymers are possible, paving the way for the generation of reagents to probe the unanchored polyubiquitin chains of defined topology, and more widely the 'ubiquitome'.

"Nature never appeals to intelligence until habit and instinct are useless. There is no intelligence where there is no need of change."

H.G. Wells

Acknowledgements

First and foremost I would like to acknowledge and sincerely thank both of my supervisors, Dr. Robert Layfield in the School of Life Sciences, and Professor. Neil Oldham, in the School of Chemistry, whose support and guidance has made the investigations in this thesis possible. Both have truly been fantastic in supporting my development, and also making the experience thoroughly enjoyable – Thanks!

I would like to extend my gratitude to all those who I have had the fortune of sharing various labs with, and have imparted various bits of lab-based wisdom (or otherwise) upon me. Specifically in Rob's group I would like to thank Dr. Jo Strachan (see…remembered this time) and Barry Shaw, who showed me the ropes, Dr. David Tooth, who has always been keen to discuss proteomics and (even) Dr. James Cavey who was always happy (I think) to listen to cloning woes. More broadly I would like to thank Dr. Varun Gopala, Dr. Farhana Haque, Dr. Alice Goode, Charlie Ducker, Emma Burton, and the extended D45 cast of Dr. Federico Dajas-Bailador, Dr. Alex Rathbone, Raquel Ribeiro, and Clare Martin, for providing excellent (soup-based) conversation during coffee breaks, entertainment, and often cake.

In Neil's group I am extremely appreciative to Dr. Kleitos Sokratous, Dr. Matthew Jenner, and Dr. Jose Afonso, who helped with all things MS when I first started, as well as Dr. Richard Elms, Daniel Rooney, Lucio Manzi, and Nkazi Tshuma, who collectively made the experience so enjoyable, and were always on hand to discuss MS, as well as the best methods to get one over Matt.

Thanks must also go to collaborators including Dr. Jed Long (who for a short period also proved to be a great house mate), Juliet Morgan and Alex Slater in Professor Mark Searle's group in the School of Chemistry, who kindly allowed me to utilise their protein purification facilities, and were always on hand to help out. Externally I must also thank former Searle group member Dr. Thomas Garner, who helped with MD simulations.

I would also like to acknowledge the friends I have had the fortune of meeting during my time at the University, who have provided welcome support, sometimes cake (re-occurring theme) and often a beer. This certainly includes the past and present members of the MonoQ's, who offered musical escape with classics like 'middle8'…. Thanks chaps. Following on from this, thanks must go to Dr. Matt Green, front man of the MonoQ's, and now grunt of the Firehouse brewery, whose eagerness to brew has provided a welcome distraction.

Special praise must also go to the Foxes, Leicester City FC, who as newly crowned premier league champions have provided sheer satisfaction, not only in the last year, but also during the course of my PhD, where almost continuous bragging rights over local rivals have been reserved, and enjoyed.

Finally I would like to extend my utmost thanks to my friends and family outside of the lab, especially including my parents and sisters who have been so patient and understanding, offering support during the tricky times. Perhaps most importantly I would like to thank Jess, without whose love, support, and encouragement this wouldn't have been possible.

Publications

Publications of relevance to arise from the investigations completed during doctoral studies at the University of Nottingham, from which some of the material presented in this thesis is re-produced.

Published

STRACHAN, J., SHAW, B., KRISHNA, V., **SCOTT, D.** *et al* 2013. Broad Utility of an Affinity enrichment Strategy for Unanchored Polyubiquitin Chains. *J Proteomics Bioinform*. S7, 1-7.

SCOTT, D., OLDHAM, N. J., STRACHAN, J., SEARLE, M. S. & LAYFIELD, R. 2015. Ubiquitin-binding domains: Mechanisms of ubiquitin recognition and use as tools to investigate ubiquitin-modified proteomes. *Proteomics*, 15, 844-861.

SCOTT, D., LAYFIELD, R. & OLDHAM, N. J. 2015. Ion mobility-mass spectrometry reveals conformational flexibility in the deubiquitinating enzyme USP5. *Proteomics*, 15, 2835-2841.

SCOTT, D., LAYFIELD, R. & OLDHAM, N. J. 2015. Structural insights into interactions between ubiquitin specific protease 5 and its polyubiquitin substrates by mass spectrometry and ion mobility spectrometry. *Protein Science*, 24, 1257-1263.

SCOTT, D., GARNER, TP., LONG, J., STRACHAN, J., MISTRY, SC., BOTTRILL, AR., TOOTH, DJ., SEARLE, MS., OLDHAM, NJ. & 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, D., STRACHAN, J., KRISHNA, VG., SHAW, B., TOOTH, DJ., SEARLE, MS., OLDHAM, NJ. & LAYFIELD, R. 2016. Method for the purification of endogenous unanchored polyubiquitin chains. *Methods Mol Biol.* 1449, 203-213.

In press

MANZI, LM., BARROW, A., **SCOTT, D.**, LAYFIELD, R., WRIGHT, TG., MOSES, JE & OLDHAM, NJ. 2016. Carbene Footprinting Accurately Maps Binding Sites in Protein-Ligand and Protein-Protein Interactions. *Nat Commun.*

Abbreviations

ADH	Alcohol dehydrogenase
BSA	Bovine serum albumin
CCS	Collisional cross section
CID	Collision-induced dissociation
CNBr	Cyanogen bromide-activated
CSD	Charge state distribution
DC	Direct current
DDA	Data-dependent acquisition
DMEM	Dulbecco's Modified Eagle Medium
DTT	Dithiothreitol
DUB	Deubiquitinating enzyme
ECL	Enhanced Chemiluminescence
ESI	Electrospray ionisation
FPLC	Fast protein liquid chromatography
FRET	Fluorescence resonance energy transfer
FTICR	Fourier transform ion cyclotron resonance
FUBE	Free ubiquitin binding entity
GAB	Gel application buffer
GFP	Green fluorescent protein
GST	Glutathione-S-Transferase
HDX	Hydrogen-deuterium exchange
HEK293T	Human embryonic kidney cell line containing the SV40 Large T-antigen
hHR23A	Human homologue of Rad23A
HRP	Horseradish peroxidase
IAA	Iodoacetamide
IKK	I-κB kinase
IMS	Ion mobility spectrometry
IM-MS	Ion mobility-mass spectrometry

IP	Immunoprecipitation
IPTG	Isopropyl-β-D-thiogalactopyranoside
ITC	Isothermal titration calorimetry
K_d	Dissociation constant
LB	Luria broth
LC	Liquid chromatography
LTQ	Linear ion trap
m/z	Mass/charge
MALDI	Matrix assisted laser desorption ionisation
MD	Molecular dynamic
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
NEM	*N*-Ethylmaleimide
nESI	Nano Electrospray ionisation
NF-κB	Nuclear factor kappa-light-chain-enhancer of activated B cells
NMR	Nuclear magnetic resonance
NZF	Npl4 zinc finger
OT	Orbitrap
PA	Projection approximation
PBS	Phosphate-buffered saline
PDB	Protein data bank
PSAQ	Protein standard absolute quantification
PTM	Post-translational modification
Q	Quadrupole
RF	Radio frequency
SAXS	Small angle X-ray scattering
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SILAC	Stable isotope labelling with amino acids in cell culture
SUMO	Small ubiquitin-like modifier
TBS	Tris-buffered saline
TOF	Time-of-flight
TUBE	Tandem ubiquitin-binding entity
tUIM	Tandem UIM domains
TWIMS	Travelling-wave ion mobility spectrometry
UBA	Ubiquitin-associated domain
UBAN	UBD in Abin and Nemo

Ub-AQUA	Ubiquitin absolute quantification
UBD	Ubiquitin-binding domain
UiFC	Ubiquitylation-induced fluorescence complementation
USP	Ubiquitin specific protease
WT	Wild-type
ZnF	Zinc finger

Contents

Chapter 1 - Introduction **4**
1.1 The Ubiquitin code 6
1.1.1 Mechanistic insights into protein ubiquitylation 6
1.1.2 Achieving ubiquitin signalling diversity 8
1.1.3 Deciphering the ubiquitin code 11
1.1.4 Regulation of the ubiquitin code 17
1.2 Strategies to probe protein ubiquitylation 18
1.2.1 Routes to (poly)ubiquitin 18
1.2.2 Ubiquitin antibodies 19
1.2.3 Expression of heterologous ubiquitin 19
1.2.4 UBDs as tools 20
1.3 Mass spectrometry 24
1.3.1 Principles of mass spectrometry 24
1.3.2 Experimental applications of mass spectrometry 30
1.3.3 Applications of mass spectrometry to probe the ubiquitome 35
1.4 Unanchored polyubiquitin chains 40
1.4.1 An unanchored ubiquitin reservoir 40
1.4.2 Unanchored polyubiquitin and the proteasome 41
1.4.3 Unanchored polyubiquitin in aggresome biogenesis 41
1.4.4 Unanchored polyubiquitin and innate immune signalling 42
1.4.5 The unanchored polyubiquitin code 44
1.5 Summary and objectives 46

Chapter 2 - Materials and methods **48**
2.1 Molecular biology methods 48
2.1.1 Bacterial expression vectors 48
2.1.2 t-UBD cloning 50
2.1.3 DNA electrophoresis 50
2.1.4 Heat shock transformation 51
2.1.5 Plasmid stocks, isolation and sequencing 51

2.1.6 Site-directed mutagenesis 51
2.2 Protein expression and purification techniques 53
2.2.1 Small-scale bacterial protein expression tests 53
2.2.2 Large-scale bacterial expression 54
2.2.3 Bacterial cell lysis 54
2.2.4 Protein purification strategies 55
2.2.5 SDS-PAGE 57
2.2.6 Desalt / buffer exchange 58
2.2.7 Protein quantification 58
2.3 Recombinant protein affinity chromatography 59
2.3.1 Protein immobilisation 59
2.3.2 Unanchored polyubiquitin chain affinity enrichment 60
2.3.3 Deubiquitylation assays 61
2.3.4 Western blotting 61
2.4 Mass spectrometry 63
2.4.1 Bottom-up proteomics 63
2.4.2 Carbene footprinting 65
2.4.3 Native mass spectrometry 68
2.5 Molecular dynamics 75
2.5.1 Determining theoretical CCS 75
2.5.2 Designing the t-UBD 76

Chapter 3 – Purification of endogenous unanchored polyubiquitin chains 78
3.1 Purification of endogenous unanchored polyubiquitin chains from HEK293T cells utilising the FUBE 82
3.1.1 Preparation of unanchored polyubiquitin-enriched cell lysate 82
3.1.2 Expression and purification of the ZnF-UBP domain of USP5 83
3.1.3 Analysis of FUBE enriched unanchored polyubiquitin from HEK293T cell extracts 85
3.2 Overcoming the limitations of the FUBE as an enrichment tool for unanchored polyubiquitin 88
3.2.1 Expression and purification of USP5 89
3.2.2 USP5-ubiquitin binding properties of recombinant USP5 92
3.2.3 C335A-USP5 as a second generation FUBE 93
3.3 Mass spectrometry analysis of endogenous unanchored polyubiquitin 97
3.4 Conclusions 102

Chapter 4 - Exploring the mechanisms of polyubiquitin recognition by USP5 . 106
4.1 Probing the conformational dynamics of USP5 111
4.1.1 Charge-state distribution analysis of USP5 111
4.1.2 Ion mobility-mass spectrometry as a tool to examine USP5 conformers . 113
4.1.3 Computational interrogation of USP5 structures 125
4.1.4 Conclusions from native ESI-IM-MS analysis and MD simulations 127
4.2 Exploring USP5 structure with respect to (poly)ubiquitin recognition 128
4.2.1 Expression and purification of USP5 and ubiquitin mutants 128
4.2.2 Probing USP5-monoubiquitin interactions 131
4.2.3 Carbene footprinting 138
4.2.4 Probing USP5-tetraubiquitin interactions 143
4.3 Discussion 146

Chapter 5 - Designing a tandem UBD construct for the selective recognition of unanchored polyubiquitin 149
5.1 Designing the t-UBD 151
5.2 Synthesising the t-UBD 154
5.3 Confirming the functionality of the t-UBD 156
5.4 Optimising the t-UBD linker length 162
5.5 Purification of endogenous unanchored polyubiquitin chains by the t-UBD .. 164
5.6 Discussion 168

Chapter 6 - General discussion 172
6.1 Purifying endogenous unanchored polyubiquitin from HEK293T cell extracts 173
6.2 Characterising the mechanism of (poly)ubiquitin recognition by USP5 175
6.3 Designing an artificial tandem UBD construct for the selective purification of Lys-48 linked unanchored polyubiquitin 178
6.4 Future work 179

Chapter 7 – References 182

Chapter 1 - Introduction

The post-translational modification (PTM) of selected proteins is a central mechanism by which the genetic coding capacity of eukaryotes is expanded, to mediate transient cellular response to internal and external stimuli. Proteins are subject to a wide variety of often reversible amino acid-specific modifications, including the covalent addition of small chemical groups, as well as sugars, lipids or indeed other proteins [1]. The consequence of PTM on protein structure and function is diverse, and as one might expect, dependent upon the modification in question. However typically, modification results in the alteration of protein surface properties or conformation, which in turn can be exploited by PTM-specific binding domains to influence protein activity, stability or localisation [2].

One of the most significant and complex PTMs to emerge is protein ubiquitylation, originally identified by Avram Hershko, Aaron Ciechanover and Irwin Rose, who were later awarded the 2004 Nobel Prize for Chemistry for their efforts. In their seminal investigations in the 1970s and early 1980s, at a time when the lysosome was considered the principal degradation machinery and the study of protein synthesis was *en vogue*, they sought to probe the underlying mechanisms of intracellular protein degradation. Initially, Hershko and colleagues identified the ~8.5 kDa ubiquitin protein (then termed APF-1), and then for the first time revealed that ATP-dependent modification of proteins, with ubiquitin, signalled lysosomal-independent degradation [3-5]. In the ensuing years the enormity of this discovery emerged, with covalent protein ubiquitylation found to not only govern the selective degradation of master regulatory

proteins such as transcription factors [6] and cell cycle regulators [7], but also being implicated, via alternative 'types' of ubiquitin signal, in non-proteolytic roles including in DNA repair [8, 9] and transcriptional regulation [10]. Accordingly, the concept of the 'ubiquitin code' has emerged, which is central to the regulation of the cellular *milieu*.

In recent years the notion that ubiquitin signalling is mediated solely by the covalent modification of substrates has been challenged. Indeed in parallel, unanchored, or substrate-free, polyubiquitin has emerged as an important regulator of numerous aspects of cellular physiology, via direct non-covalent interactions with substrates [11]. The investigations detailed in *chapters 3 - 5* seek to develop strategies to probe the molecular composition and interactions of the unanchored polyubiquitin pool. Accordingly in *chapter 1* we first introduce the 'ubiquitin code' and review the previous experimental tools and strategies that have been utilised to probe the wider ubiquitin-modified proteome (ubiquitome). We then complete *chapter 1* by reviewing insights into the unanchored polyubiquitin pool afforded to date.

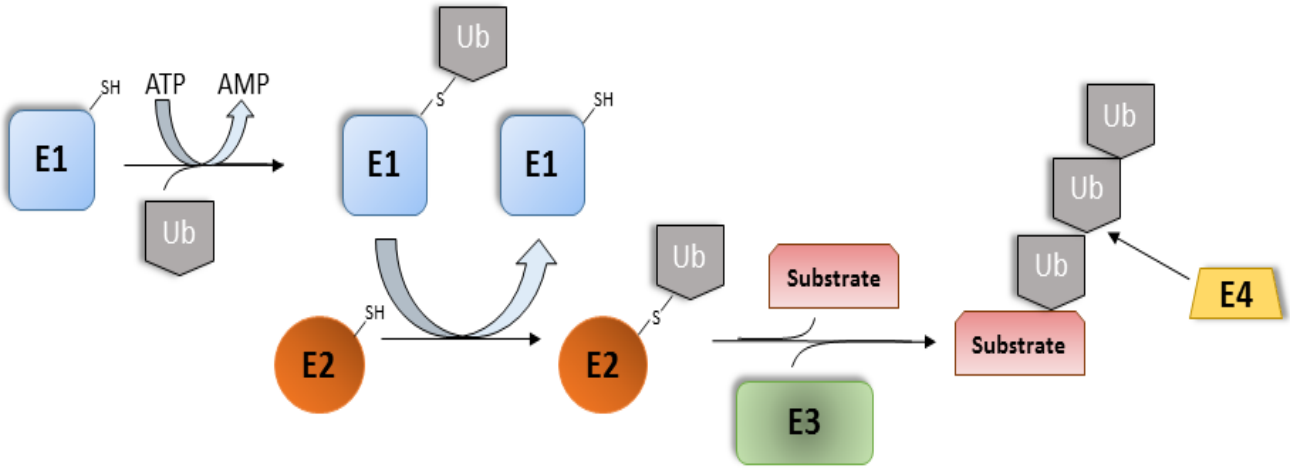
1.1 The Ubiquitin code

Tight regulation of protein PTM is critical to the maintenance of cellular homeostasis. Indeed, in the case of protein ubiquitylation, misregulation has been implicated in the pathogenesis of a range of diseases including cancers [12, 13], as well as neurodegenerative disorders [14, 15]. Consequently, understanding the dynamic interplay between the proteins responsible for writing, reading and erasing the so called 'ubiquitin code' has been the subject of intense research [16], which is summarised in *section 1.1*.

1.1.1 Mechanistic insights into protein ubiquitylation

The 76-residue monoubiquitin protein is covalently conjugated, through its C-terminal Gly-76 residue, to (usually) the ε-amino group of selected Lys side chains on substrates. Ubiquitylation is catalysed by an enzymatic cascade comprising an E_1-activating, an E_2-conjugating and an E_3-ligase enzyme, as summarised in Figure 1.1 [17, 18]. In the first ATP-dependent step, an E_1 enzyme activates ubiquitin via a ubiquitin-adenylate intermediate, forming a thioester with the carboxyl group of Gly-76. Subsequently, thioester transfer of ubiquitin from E_1 to a cognate E_2 takes place, to yield an E_2-ubiquitin thioester intermediate. Finally, an E_3 ligase binds to both the E_2-ubiquitin thioester and the substrate to catalyse the transfer of ubiquitin from the active site cysteine of the E_2 (directly or indirectly) to the substrate [19]. The E_3 ligase can be considered the determinant of substrate specificity, which is reflected in the existence of >600 mammalian E_3 enzymes, compared to ~60 E_2 and two E_1 enzymes [20]. Additionally, a subclass of E_3-like proteins have been identified, dubbed E_4 enzymes, which can act as polyubiquitin chain elongation factors [21].

A



B

MQIFV**K**TLTG**K**TITLEVEPSDTIENV**K**A**K**IQD**K**EGIPPDQQR

LIFAG**K**QLEDGRTLSDYNIQ**K**ESTLHLVLRLRGG

Figure 1.1 Overview of the ubiquitylation reaction mechanism. (**A**) The ubiquitylation cascade begins with ATP-dependent charging of the E_1 enzyme and results in formation of a thioester bond between the ubiquitin C-terminus and the E_1 active site cysteine. Ubiquitin is then transferred to the E_2 active site cysteine in a transthioesterification reaction. An E_3 ligase finally catalyses transfer of the ubiquitin from the active site cysteine of the E_2, to a primary amine on a lysine side chain (or protein N-terminus). Ubiquitin chains can be extended by E_4 enzymes. (**B**) The primary sequence of monoubiquitin, with the residues modified by further ubiquitin moieties highlighted in red, this includes Met-1, Lys; 6,11,17,19,33,48 and 63.

1.1.2 Achieving ubiquitin signalling diversity

According to the collection of E_1, E_2 and E_3 enzymes sequestered, ubiquitin can be appended to substrates as a monomer, at a single or multiple sites, or as polyubiquitin chains assembled to the preceding ubiquitin moiety through internal Lys residues found in positions; 6, 11, 27, 29, 33, 48 and 63, or the N-terminal Met-1 residue (as indicated in Figure 1.1). This polymerisation of ubiquitin moieties, in the simplest form, yields eight unique polyubiquitin modifications [22].

While the conjugation of single or multiple ubiquitin monomers to substrates was later found to be significant [23-26], modifications with Lys-48 linked polyubiquitin chains were the first to be probed. Indeed pioneering studies revealed that Lys-48 linked polyubiquitin chains, of four moieties in length, acted as the optimal 'tag' for the recognition and degradation of ubiquitin-modified substrates by the 2.5 mDa protease, termed the 26S proteasome [27-29]. Subsequently, Lys-63 linked polyubiquitin chains were implicated in roles independent of the 26S proteasome, including in intracellular transport processes [30], DNA damage response [9, 31], cytokine signalling [32], and as an activator of autophagy, an alternative degradatory pathway which promotes clearance of protein aggregates [33]. Consequently, the 'ubiquitin code' has emerged, with the central dogma that the appendage of unique (poly)ubiquitin modifications, which differ according to chain length and linkage, to selected substrates, results in specific signalling outcomes. Recent proteomic investigations have supported this notion, with all eight ubiquitin polymers described *in vivo*. Critically, the relative abundances of each polyubiquitin linkage proves dynamic with respect to the organism and cell type in question [34-36], as well as in response to pharmacological intervention (i.e. proteasome inhibition) [34, 35, 37], or at cell cycle checkpoints [38].

Insights into the roles of polyubiquitin chains linked by residues other than Lys-48 or 63 have proved challenging to elucidate. However, the emergence of strategies to probe

ubiquitin modifications in a linkage specific manner, such as polyubiquitin linkage selective antibodies (discussed further in *section 1.3*) has mediated the evaluation of selected polyubiquitin linkages. Indeed a Met-1 specific antibody has permitted the discovery of the LUBAC, a 600 kDa E3 ligase complex which with exquisite specificity produces Met-1 linked polyubiquitin chains. The functional significance of this protein complex and the Met-1 linked chains produced remains under investigation, but appear fundamental to the regulation of NF-κB and thus immune signalling [39-42]. Concurrently, a Lys-11 linked linkage selective antibody has revealed that assembly of such chains by the APC/C, mediate and thus regulate the selective degradation of cell cycle substrates [38, 43]. Furthermore, as alluded to in Figure 1.2, roles for the other 'atypical' linkages, polyubiquitin chains linked by Lys-6, 27, 29 and 33, are also tentatively emerging [44-51].

Intriguingly, new layers of complexity in the 'ubiquitin code' have in recent years begun to surface (reviewed in detail in [52]). Heterogeneous polyubiquitin chains, i.e. polyubiquitin chains with more than one linkage type, have been implicated in specialised signalling roles, for example including altering the recognition and processing of ubiquitylated substrates at the proteasome [53-57]. Furthermore ubiquitin is now understood to be not only modified by further ubiquitin moieties, but also with ubiquitin-like proteins, including the small ubiquitin-like modifier (SUMO) family of proteins [58], ISG15 [59] and NEDD8 [60], as well as small chemical groups including acetyl [61] and phosphate [62] groups, which contributes to further expansion of ubiquitin signalling capabilities. Finally, as alluded to in the introductory remarks, unanchored or substrate-free polyubiquitin chains have emerged as novel regulators of cellular physiology [11], completing second-messenger-like functions (discussed further in *section 1.5*), indicating that research into the 'ubiquitin code' is still very much in its infancy.

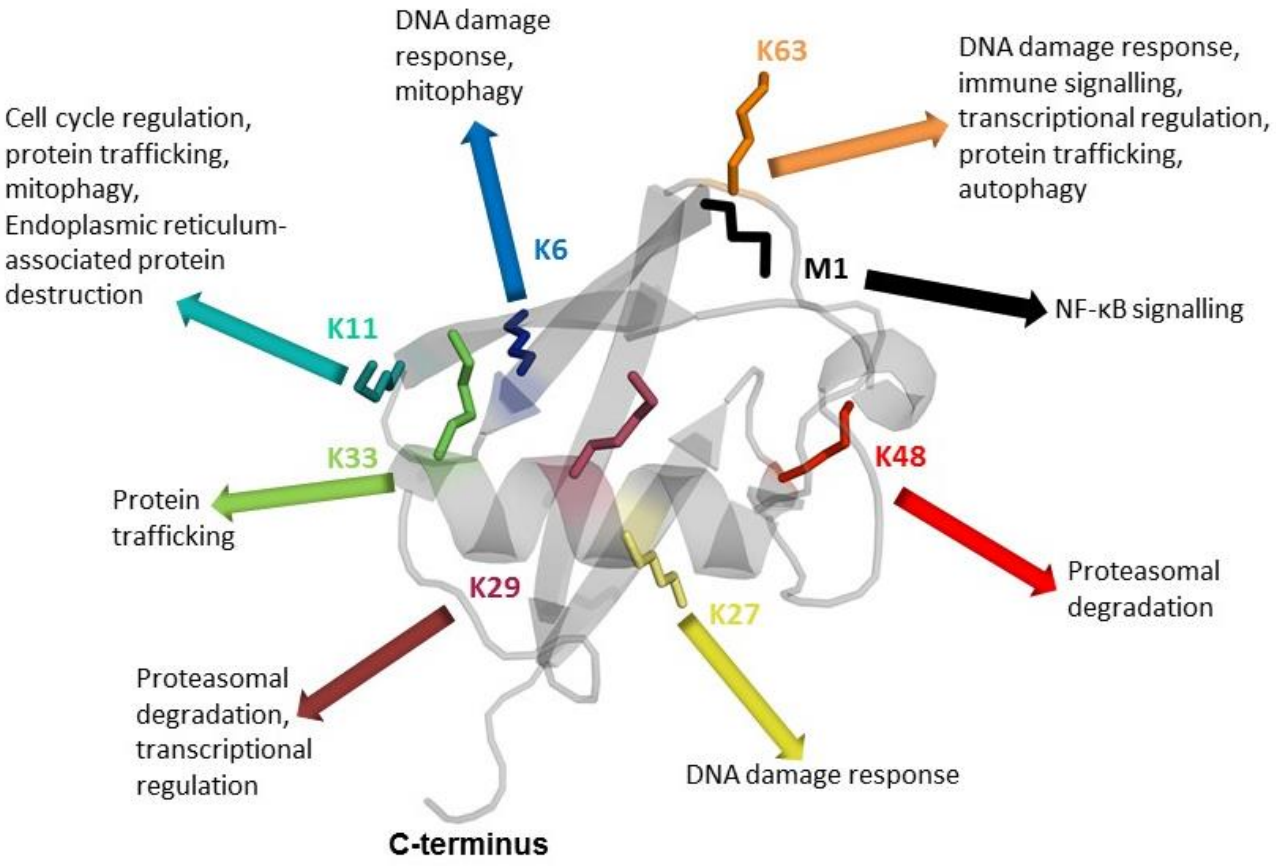


Figure 1.2 The emerging roles of protein ubiquitylation. The structure of ubiquitin (grey, PDB 1UBQ) indicates that all seven lysine residues (side chains coloured) reside on different surfaces of the molecule. Note that Met-1 (black) is spatially close to Lys-63 (orange). The emerging role for each polyubiquitin linkage is indicated (as reviewed in [52]).

1.1.3 Deciphering the ubiquitin code

Diverse (poly)ubiquitin modifications are recognised and transduced by ubiquitin-binding domains (UBDs) found in intracellular ubiquitin-binding proteins. These modular domains, of which ~20 families have been identified, participate in non-covalent interactions with ubiquitin surface patches, presented by ubiquitin modifications. These well-defined surface patches include a hydrophobic surface centred on Ile-44, an acidic hydrophilic Asp-58-centred surface, as well as residues at the C-terminus of ubiquitin (when not conjugated to targets) [63]. Furthermore, additional interactions with residues in the vicinity, or surrounding these sites have also been noted [64]. Table 1.1 highlights some of the UBDs of relevance to the investigations in this thesis (for extensive description of UBD families and their respective members, see [63]).

Table 1.1 Selected UBDs and there binding specificity (adapted from [65]).

Ubiquitin-binding domain (UBD)		**Representative protein**	**Ubiquitin surface**	**Polyubiquitin selectivity**
Ubiquitin-interacting motif	UIM	S5a/RPN10	(UIM-1 and UIM-2) Ile-44 (UIM-2 only) Leu-8, Val-70, Leu-71	Longer polyubiquitin chains (relative to monoUb), both Lys-48 and Lys-63 linked. Other linkages not determined (N.D.).
		Rap80	Ile-44	Lys-63 linked, relative to Lys-48 linked polyubiquitin chains. Increased affinity for longer chains. No binding to linear polyubiquitin chains.
Ubiquitin-associated domain	UBA	hHR23A-UBA2	Ile-44	Lys-48 linked polyubiquitin chains
		Ubiquilin-1	Ile-44	No reported selectivity.
Ubiquitin binding in ABIN and NEMO	UBAN	ABIN	Ile-44 and Phe-4 or Gln-2/Glu-16	Linear, relative to Lys-48 and Lys-63 linked polyubiquitin chains.
Npl4 zinc finger	NZF	TAB2 and TAB3	Ile-44	Lys-63 linked, relative to Lys-48 linked and linear polyubiquitin chains.
A20 zinc finger	ZnF A20	ZNF216	Asp-58	No reported selectivity.
Zinc finger ubiquitin-binding domain	Znf UBP (BUZ)	IsoT/USP5	C-terminal GlyGly	No reported selectivity.

UBDs, as exemplified in Figure 1.3, typically recognise (poly)ubiquitin with modest affinity (3-500 μM), via a simple single-site mode of interaction, in a Lys-linkage independent manner. However, recognition of ubiquitin surface patches is by no means mutually exclusive, for example the A20 Zinc-finger (ZnF) domain of ZNF216 (Asp-58 binding) and the ubiquitin-associated domain (UBA) domain of p62 (Ile-44 binding) can co-occupy a single ubiquitin moiety [66].

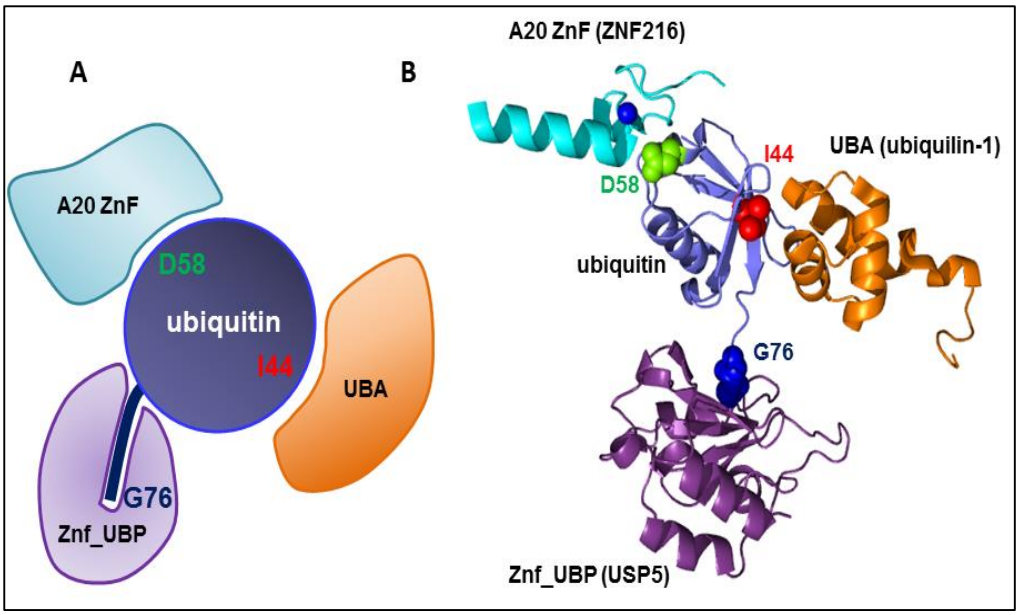


Figure 1.3 Overview of the sites of ubiquitin recognition by UBDs. (A) Cartoon representation of the key ubiquitin surfaces, which are targeted by UBDs. This includes an Ile-44 (red) centred hydrophobic patch, an acidic hydrophilic Asp-58 (green) centred patch and residues (blue) at the C-terminus of ubiquitin (when not conjugated to substrates). Representative UBDs that recognise these surfaces are indicated. (B) Structural model of a single ubiquitin in complex with three different UBDs (parent proteins indicated in brackets), generated from the overlay of the experimentally determined binary complexes, demonstrating that simultaneous interactions at different sites on ubiquitin are not mutually exclusive (model derived from PDB 2L00, 1WR1 and 2G45) (Adapted from [65]).

Central to the notion of Lys-linkage specific roles *in vivo*, UBDs can also transduce ubiquitin modifications in a Lys-linkage specific manner, with structural insights, afforded by nuclear magnetic resonance (NMR), X-ray crystallography, and single-molecule fluorescence resonance energy transfer (FRET), having proved critical to rationalising mechanisms of ubiquitin recognition, which we will now consider.

Inspection of the structures of *in vitro* generated diubiquitin (all polyubiquitin linkages except Lys-27, see (Figure 1.4)), reveals that the site of ubiquitin conjugation dictates the relative orientation of ubiquitin moieties and the extent of intermolecular contacts made between ubiquitin moieties. Indeed, diubiquitin chains linked by Lys-63 and Met-1 adopt similar extended conformations with no contacts between ubiquitin moieties (excluding the covalent linkage), making the structure flexible, and interaction with the Ile-44 centred surface freely accessible. Conversely, diubiquitin polymers linked by Lys-6, 11, 33 and 48 adopt closed or compact structures, with significant, albeit distinct contacts made between ubiquitin moieties. Accordingly, we can conclude that the site of ubiquitin conjugation determines the interaction surfaces presented by ubiquitin polymers, and that it is likely that each of the eights simple ubiquitin polymers present unique topologies, which underlie polyubiquitin linkage specific roles [56, 67-71].

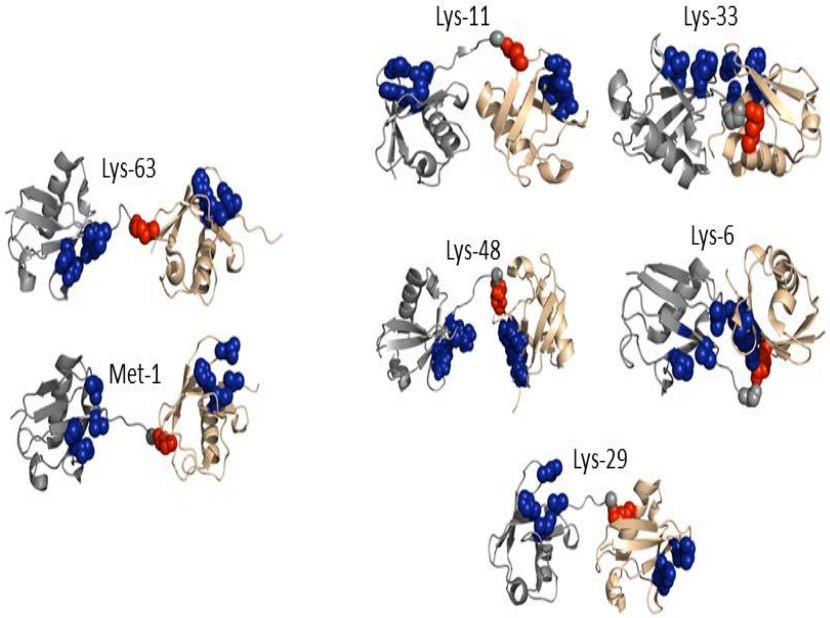


Figure 1.4 The available structures of *in vitro* generated diubiquitin. Structures (linkage indicated above structure) fall under two broad conformational distributions, one wherein the dimer adopts a compact conformation, as typified by Lys-6/11/33/48 linkages, the second, typified by Lys-63 and Met-1 (and to an extent Lys-29) linkages, wherein the dimer adopts an extended conformation. Note that the positioning of the residues surrounding the Ile-44 patch (Leu-8/Ile-44/His-68/Val-70) (blue) differ significantly depending upon isopeptide linkage (Met-1, peptide) (red). (PDB 2W9N, 2XK5, 2XEW, 4S22, 4XYZ,1Z06 and 2JF5 (Structures adapted from [56, 67-71]).

The inherent flexibility of the isopeptide linkage however, means that each ubiquitin polymer is likely to be associated with an ensemble of dynamic structures in conformational exchange [72]. This is evidenced in the recent structures of Lys-33 linked ubiquitin polymers, where the di- and tri-ubiquitin polymers of this linkage adopt markedly different compact and extended conformations respectively [67]. Critically this intrinsic malleability of ubiquitin polymers has been established as an important factor in conformational selection by UBDs, which sample pre-existing conformations, or remodel polyubiquitin chains, to facilitate the linkage-specific recognition of ubiquitin polymers [73]. An exemplar of conformational selection by UBDs relates to Lys-29 linked diubiquitin, which in isolation assumes an extended conformation similar to Met-1 and Lys-63 linked chains, albeit with minor contacts between ubiquitin moieties (see Figure 1.4). However, when co-crystallised with the Lys-29/33 selective Npl4 zinc finger (NZF) 1 UBD of TRABID, the dimer undergoes significant conformational rearrangement. This permits recognition of both the proximal ubiquitin (i.e. ubiquitin moiety with a free C-terminus) at residues around the Lys-29 linkage, and the distal ubiquitin at the Ile-44 patch, and thus linkage selective recognition [56].

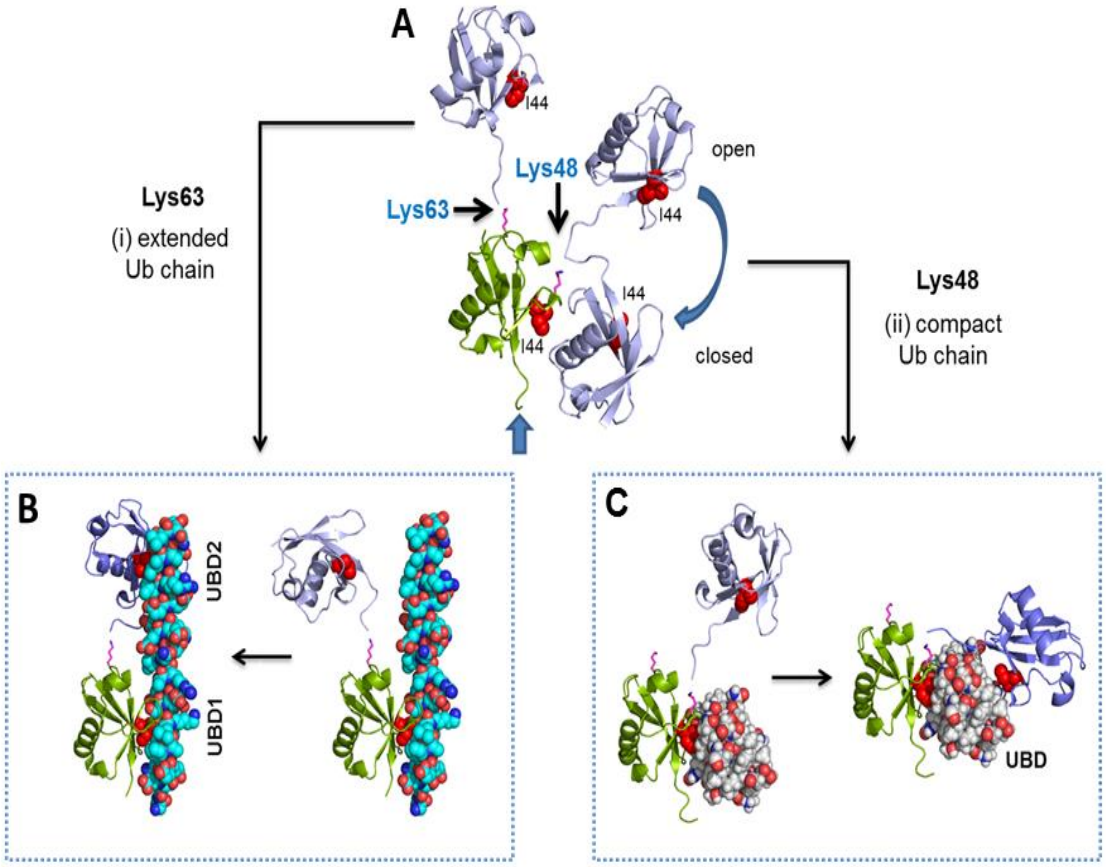
The conformational selection of specific polyubiquitin topologies by isolated UBDs can also be extended to other members of the NZF family, found in a number of different proteins including TAB2 and TAB3, which display Lys-63 linkage selectivity [74, 75], and HOIL-1, which displays Met-1 selectivity [76]. Furthermore, a number of alternate families of UBDs utilise conformational selection to mediate interactions with a specific ubiquitin topology. For example, this includes the isolated ubiquitin-associated (UBA) 2 domain of the proteasome shuttling factor human homologue of RAD23A (hHR23A) [70] (and an orthologue in yeast, Mud 1 [77]), which exploits and selects the compact Lys-48 linked diubiquitin structure, to interact with Ile-44 patches on both the proximal and distal ubiquitin moieties, to mediate the selective recognition of Lys-48 linked

polyubiquitin chain (see Figure 1.5). While the A20 ZnF4 domain of the A20 protein selectively recognises Lys-63 tri-ubiquitin [78], and the UBD found in ABIN and NEMO (UBAN) favours the recognition of linear diubiquitin [79].

1.1.3.3 Avidity affects as determinants of polyubiquitin specificity

In addition to conformational selection by isolated UBDs, polyubiquitin linkage selective binding by UBDs can also be achieved through the use of multiple UBDs, as either repeats of the same motif or a collection of different and complementary domains, arranged in such a way to favour multivalent interactions with specific polyubiquitin chain topologies. This is exemplified by RAP80, a component of the BRCA-1A complex central to DNA double strand break repair [80]. RAP80, presented in Figure 1.5, is composed of tandem UIMs (tUIM), which are rigidly connected by an α-helix, to exploit the distance between the Ile-44 patches on the distal and proximal ubiquitin, achieving avid interactions with Lys-63 linked diubiquitin, but not Lys-48 or Met-1 linked polymers [81]. Similarly, Lys-48 linked polymers can also be selectively recognised by arrangements of UIM domains, as typified by the tUIMs of the proteasome subunit S5a/RPN10 [82], or indeed the proteasome-associated AIRAPL protein [83].

Arrays of UBDs can also be positioned to favour linkage independent interactions, thus acting as a ‘molecular ruler’ to select on the basis of chain length rather than linkage. This is exemplified by the deubiquitinating enzyme (DUB) ubiquitin specific protease 5 (USP5), which contains four UBDs, and is thought to bind to (poly)ubiquitin independent of linkage to achieve the selective disassembly of unanchored polyubiquitin chains of assorted polyubiquitin linkage via avidity effects [84]. Similarly, ubiquitin chain length (and indeed linkage) is critical at the proteasome, with regards to substrate recognition, processing and ultimately degradation [85].



Tandem UIMS (RAP80) – Lys-63 Ub_2

UBA2 (hHR23A) – Lys-48 Ub_2

Figure 1.5 Mechanisms of polyubiquitin recognition by UBDs. (A) Lysine linkage determines topology and thus interaction surface presentation by diubiquitin. Lys-63 linked dimers adopt an extended conformation, with the Ile-44 patch (red) accessible in both the proximal (green) (i.e. ubiquitin moiety with a free C-terminus) and distal (blue grey) ubiquitin moieties. Whilst the Ile-44 patches on the proximal and distal ubiquitin moieties of Lys-48 diubiquitin interact weakly, resulting in equilibrium between extended and compact conformations. (B) RAP80 interacts specifically with Lys-63 diubiquitin using two rigidly connected UIMs in tandem within a single extended helix (PDB 3A1Q). (C) The UBA2 of hHR23A interacts with Lys-48 diubiquitin by selecting for and binding to a compact conformation (PDB 1Z06) (adapted from [65]).

1.1.4 Regulation of the ubiquitin code

In conclusion, the prevailing model is that structurally unique polyubiquitin modifications are synthesised and appended by the E_1-E_3 enzyme cascade, and subsequently recruit different ubiquitin-binding proteins via interactions with their UBDs, to influence the fate of the modified protein. This process is tightly regulated (see Figure 1.6) by the activity of DUBs, which with the exception of the JAMM metallo-protease family, are cysteine proteases. DUBs regulate or fine tune ubiquitin interactions by removing or editing such modifications [86]. An exemplar, previously alluded to, is the DUB USP5, central to the investigations in *chapters 3 and 4* (where USP5 structure and function is further reviewed), which is responsible for the selective disassembly of unanchored (or substrate free) polyubiquitin chains [87]. We can conclude that ubiquitylation is a dynamic, reversible modification central to cellular homeostasis.

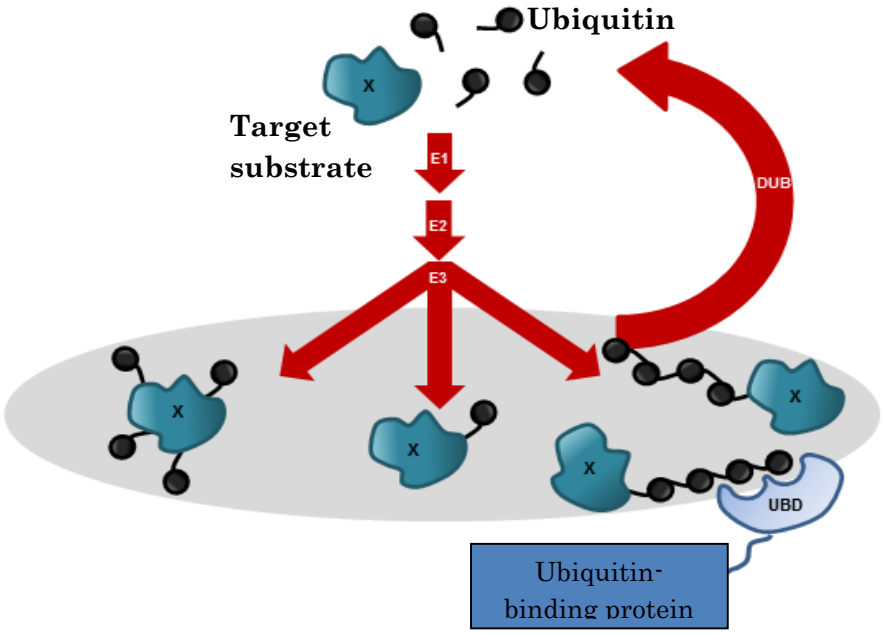


Figure 1.6 Cartoon summary of the ubiquitin system. Ubiquitylation of selected substrates (denoted x) with ubiquitin (black 'tadpoles'), is catalysed by the concerted activity of E_1-E_3 enzymes. This results in proteins that are singly or multiply monoubiquitylated, or modified with polyubiquitin chains of different topologies. The ubiquitin code is 'read' by more than 20 different families of UBDs within intracellular ubiquitin-binding proteins. Ubiquitylation can be reversed or edited by the actions of DUBs (adapted from [65]).

1.2 Strategies to probe protein ubiquitylation

While the initial investigations of Hershko and colleagues relied upon elegant biochemistry and enzymology, the necessity to dissect the ubiquitome and Lys-linkage specific roles within this sub-proteome, has expedited the development of a number of molecular tools, which are reviewed in *section 1.2*.

1.2.1 Routes to (poly)ubiquitin

Constructs that afford the expression of monoubiquitin, as well as polymers of linear, Met-1 linked polyubiquitin have long been feasible in bacterial expression systems, permitting large scale preparation and purification. Furthermore, the cognate E_2/E_3 pairs responsible for the linkage specific assembly of Lys-48 and 63 linked polyubiquitin chains rapidly emerged [88]. However, the enzymes responsible for the linkage specific assembly of the remaining polyubiquitin chains remained, until very recently, largely unknown and thus proved limiting in the dissection of linkage-specific roles [52]. Accordingly, advances in the high fidelity chemical synthesis of polyubiquitin has been key, and has resulted in the availability (although costly) of all eight polymers of ubiquitin, allowing the crystal structures, highlighted in Figure 1.4 to be obtained.

The availability of recombinant ubiquitin has also accelerated the development of a number of synthetic reagents, which utilise ubiquitin as a 'scaffold', to probe ubiquitin interacting partners. For example, this includes activity-based probes, which react specifically with the active-site cysteine of cysteine protease DUBs (excellently reviewed in [89]), and thus can be utilised in competition assays to uncover DUB specificity [90] and activity in response to inhibitors [91].

Of late strategies to produce ubiquitinated substrates have also been developed. This includes native chemical ligation approaches [92], or alternately reconstitution of the

eukaryotic ubiquitin system in bacteria [93], wherein candidate $E_1/E_2/E_3$ enzymes and substrate are co-expressed to yield ubiquitylated proteins in milligram quantities [94].

1.2.2 Ubiquitin antibodies

The isolation of ubiquitin, alluded to in *section 1.2.1*, has driven the production of a number of 'general purpose' ubiquitin antibodies. Furthermore, the subsequent availability of all eight ubiquitin polymers has, in the case of chains linked by Lys-11, 48, 63 and Met-1, facilitated the development of linkage-selective antibodies which, as discussed in *section 1.1.2*, has aided the dissection of the roles of these chains *in vivo* [95]. However, as yet linkage selective antibodies for the remaining polyubiquitin linkages have failed to emerge

1.2.3 Expression of heterologous ubiquitin

One of the first approaches towards the dissection of polyubiquitin linkage specific roles *in vitro*, was the ubiquitin replacement strategy. In this approach, endogenous ubiquitin genes are replaced with a series of Lys to Arg point mutations, driving selective expression of defined polyubiquitin chains. This methodology proved insightful in budding yeast (*S. cerevisiae*), confirming the significance of the Lys-48 polyubiquitin linkage with respect to the 26S proteasome, and implicating Lys-63 linked chains in DNA repair [35, 96, 97]. However, application of the ubiquitin replacement strategy to mammalian systems has proved limited, owing the abundance (estimated concentration of 20 μM, as the product of 4 genes) and stability of ubiquitin, as well as lethality upon knockout. Furthermore, perturbation of the ubiquitin surface has been noted to have wider implications on global ubiquitin recognition, specifically in the case of the Lys-6-Arg mutant, where ubiquitin conjugates accumulate [98] and the assembly of Lys-11 [43] and heterotypic Lys-11/Lys-63 linked polyubiquitin chains is inhibited [99]. A promising alternative to this strategy, developed by Xu *et al* [100], has utilised RNA interference to replace endogenous Lys-63 linked polyubiquitin with a Lys-63-Arg

mutant in a mammalian system, to successfully investigate the roles of this linkage in immune signalling [100].

Attempts to identify substrates of the ubiquitin system, and more broadly catalogue the ubiquitome, has prompted the development of constructs which permit the overexpression of N-terminally epitope tagged (including myc, Hisx6, HA, GST, FLAG or a Biotin-accepting domain [101]) ubiquitin in cell culture models. This approach permits denatured and thus robust purifications of ubiquitin modified proteins [35], and is particularly powerful when combined with mass spectrometry (MS) (discussed further in *section 1.3*).

1.2.4 UBDs as tools

In addition to epitope-tagged ubiquitin, which permits stringent purification of ubiquitin-modified proteins, the inherent binding specificity of UBDs for ubiquitin, and the stable, covalent nature of ubiquitin modification, has led to the exploitation of UBDs as tools in the affinity enrichment of the ubiquitome (as reviewed in [65]). Figure 1.7 presents a summary of the applications of UBDs, which will now briefly discuss.

One of the first examples of the application of UBDs saw a recombinant GST-fusion of the S5a/Rpn10 proteasome subunit (which contains two UIMs) coupled to glutathione Sepharose, used to purify ubiquitinated proteins from a range of sources [102-105]. Subsequently, additional GST-fusions have found utility, including alternative full-length proteasome-associated proteins, including Rad23 and Dsk2 [106], as well as isolated UBDs which target the Ile-44, Asp-58 centred patches as well as the C-terminal residues [107, 108]. A key consideration, highlighted by Nakayasu *et al* [109], is that the portion of the ubiquitome enriched differs significantly depending upon the UBD-based affinity resin used.

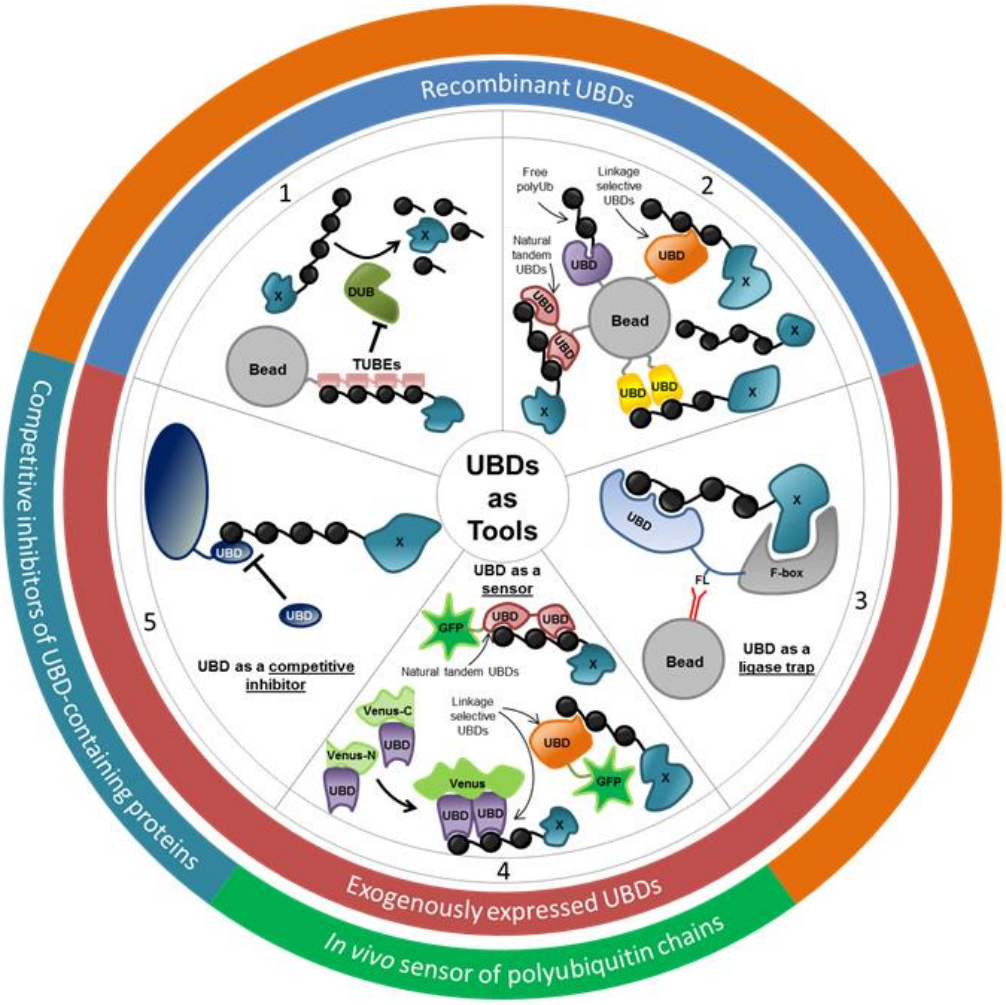


Figure 1.7 Overview of the different applications of UBDs as tools to investigate the ubiquitome. (1) Artificial tandem ubiquitin-binding entities (TUBEs) can be used to directly purify (poly)ubiquitin-modified proteins (denoted x, black 'tadpoles' represent ubiquitin) and also act to inhibit DUB activity (and proteasomal degradation) in cell extracts. (2) Individual UBDs, or natural tandem UBD repeats, with specificity for different ubiquitin interaction surfaces, can be used to purify polyubiquitin and ubiquitin-modified proteins. Free (unanchored) polyubiquitin can be captured by UBDs with specificity for ubiquitin's free C-terminus (purple). In some cases, isolated UBDs retain binding specificity on beads; in other cases artificial avidity effects may mask intrinsic specificity. (3) UBDs form part of the Ubiquitin Ligase Substrate Trapping approach, where they are artificially fused to different F-box proteins (SCF components) and bind the nascent ubiquitin chain of the modified ligase substrate(s), which are then subject to MS-based identification. (4) Fusing linkage-selective individual and tandem UBDs to fluorescent tags allows ubiquitin-signalling events to be monitored in living cells. Ubiquitylation-induced fluorescence complementation (UiFC) involves two tandem arrays of three Epsin UIMs fused to terminal fragments of the Venus fluorescent protein. Simultaneous binding of both arrays to Lys48-linked polyubiquitin chains brings the two Venus fragments into close proximity reconstituting fluorescence. (5) Expression of individual UBDs in mammalian cells can be used to competitively inhibit selected ubiquitin mediated processes (adapted from [65]).

A major advance in the isolation of ubiquitin-conjugates followed the introduction of engineered tandem ubiquitin-binding entities (TUBEs), wherein multiple repeats (normally four) of the same non-selective UBD are placed in tandem to exploit avidity affects, to favour the affinity purification of ubiquitinated substrates [110]. These commercially available reagents are typically based upon multiple repeats of either the UBA1 domain of hHR23A (TUBE 1) or UQ1-UBA of ubiquilin-1 (TUBE 2), which exhibit 100- to 1000-fold enhancement in affinity for Lys48- and Lys63-linked tetraubiquitin chains respectively, when compared to the isolated UBA domains. Furthermore, in addition to acting as tools for the efficient purification of ubiquitin-modified proteins, TUBEs have also been shown to exert a protective effect towards proteasomal degradation and DUB activity in cell extracts [110], allowing for example, the recovery of otherwise labile, or short-lived polyubiquitin conjugated proteins, including p53 and IκBα. The success of the TUBE approach has augmented the development of further UBD fusions, including: E_3-UBD (assorted UBA domains) [111] and E_3-ubiquitin [112] constructs, to affinity purify substrates of the E3 ligase in question. Moreover the E_2-activating enzyme, UBE2S, has been fused to the ZnF-UBP UBD of USP5, to aid the production and capture of free Lys-11 linked chains for initial structural studies [68].

UBDs whether organised in artificial tandem arrays [113], as with TUBEs, or in isolation [114], have also found additional miscellaneous applications in the dissection of ubiquitin-mediated processes, as highlighted in Figure 1.7 (and extensively reviewed in [64]). Indeed UBDs with intrinsic selectivity for Lys-63 linked (including the UIMs of RAP80 and Vps27 or the TAB2 NZF) or linear (UBAN) polyubiquitin have been fused to green fluorescent protein (GFP), to produce reagents that can selectively monitor the localisation and accumulation of the polyubiquitin chains *in vitro*. An alternative ubiquitin biosensor approach based on fluorescence complementation has also been developed [115], wherein biomolecular-induced fluorescence

complementation is employed for the specific detection of Lys48-linked polyubiquitin chains *in vivo* [66]. Owing to the intrinsic specificity of such UBD based tools for ubiquitin substrates, the aforementioned ubiquitin biosensors has also found application in the inhibition of specific ubiquitin mediated pathways to further aid dissection of specific ubiquitin topologies and there functional significance (this idea is further examined in [65]). Finally, the intrinsic linkage specificity of DUBs, afforded by their UBDs, can be exploited to define the nature of ubiquitin modifications decorating a substrate. In this approach, termed the 'ubiCREST strategy', a panel of DUBs are incubated with free ubiquitin or ubiquitylated substrates, reactivity is subsequently assessed by Western blotting [116].

In conclusion, there are a number of tools available to probe the ubiquitome, which have offered original insights into the minutiae of ubiquitin modification. However, reagents which permit the linkage selective purification of defined polyubiquitin topologies (chain linkage and length), are to date limited.

1.3 Mass spectrometry

Since the pioneering investigations of J.J Thomson, Francis Aston and Arthur Dempster [117-119], significant technical advances have meant MS has emerged as one of the most powerful techniques with regards to the study of protein primary sequence, PTM and interactions, including of specific relevance to this thesis, the ubiquitome. MS approaches are applied extensively in the investigations in this thesis, accordingly in *section 1.3* we review the fundamentals of mass spectrometry, and the application to the analysis of proteins. We then review previous applications of MS to probe the ubiquitome.

1.3.1 Principles of mass spectrometry

Fundamentally, MS is utilised to determine the molecular mass of an element or compound, following conversion to gas-phase ion, and manipulation according mass-to-charge (*m/z*) ratio. By nature, a mass spectrometer consists of three central components; an ion source, where gas-phase ions are produced, a mass analyser, which separate ions according to *m/z* ratio, and finally a detector, wherein ions are measured proportional to abundance. A typical cartoon schematic is illustrated in Figure 1.8, we will now discuss the key components.

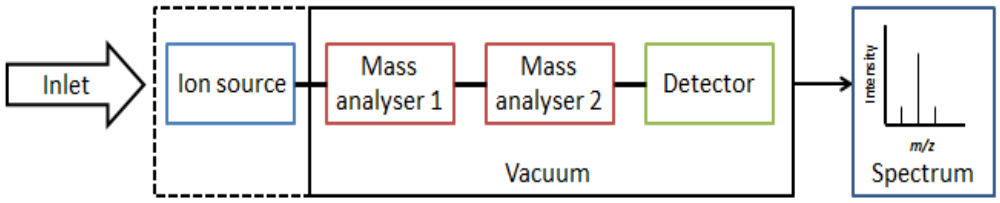


Figure 1.8 Schematic of a (tandem) mass spectrometer. The sample inlet can include separation technology such as liquid chromatography (LC), or more simply a syringe pump depending upon the nature of the experiment. Gas phase ions are generated in an ion source, and transferred via ion optics to a mass analyser, where ions are separated according to their *m/z* ratios. Ions are then detected and the signals are converted into a mass spectrum, where charge deconvolution takes place to determine the mass of the analyte. A continuous pressure gradient is used from the source to the detector to aid ion transfer through the instrument.

1.3.1.1 Generating gas-phase ions

A number of ionisation techniques have been developed that transfer analytes to gas-phase ions. However, historical methods including for example; electron ionisation and chemical ionisation, yield fragmented molecular ions and additionally are not applicable to large, non-volatile, thermally labile biomolecules, such as proteins.

1.3.1.1.1 Electrospray ionisation

Accordingly a critical advance, with respect to MS-based analysis of proteins, was the development in the 1980's of 'soft' ionisation techniques such as electrospray ionisation (ESI) (and also matrix assisted laser desorption ionisation (MALDI)) [120, 121]. ESI permits the ionisation and gentle desolvation of proteins from solution, into the gas-phase without the input of excessive internal energy, preserving covalent and potentially non-covalent bonds [122, 123].

1.3.1.1.2 Analysis of proteins by electrospray ionisation

Analytes are first dissolved in a volatile, polar buffer that will readily undergo electrochemical reactions (i.e. water or methanol solutions), and then pumped at low flow rate (typically 1–20 μL/min) into a conductive capillary [123]. When completing ESI-MS analysis of proteins, mass spectrometers are typically operated in positive ion mode, thus the tip of the capillary, illustrated in Figure 1.9, is held at a high potential difference of +1.5-3 kV, relative to a counter electrode, to generate a strong electric field (~10^6 V/m). Consequently positive charges (typically protons) accumulate at the surface of the tip, causing the liquid to be deformed into a highly turbulent Taylor cone, which in turn results in the ejection of an aerosol, or fine jet, of highly charged droplets [124].

The stability of the charged droplets is governed by surface tension, which acts to maintain the spherical shape of the droplet, and coulombic force, which proceeds to separate the droplet owing repulsion of 'like' charges at the droplet surface. As the droplet traverses the space between the capillary tip and the inlet (see Figure 1.9), rapid evaporation of solvent, aided by heated gas, takes place. Accordingly, whilst total charge on the droplet remains constant, droplet shrinkage results in higher charge density, leading to an increase in coulomb repulsion between the 'like-charges' at the droplet surface. At a given droplet radius, defined as the Rayleigh limit, Coulombic repulsion surpasses the surface tension holding the droplet together, leading to Coulomb fission, or the disintegration of the parent droplet into smaller charged progeny [125]. The theoretical notion of charged droplet fission was supported by Gomez and Tang [125], although their experimental evidence additionally indicated that droplets can alternate between oblate and prolate shapes. Elastic vibration between such states results in the ejection of progeny droplets prior to reaching the Rayleigh limit, effectively reducing the coulombic repulsion required for Coulombic fission [125, 126].

Cycles of charged droplet desolvation and fission proceed to eventually yield nanodroplets, from which gas-phase ions are formed and released, in a mechanism that remains to be fully elucidated. Two prevailing models which describe this process have emerged, the charged-residue model (CRM) and the ion evaporation model (IEM), although both are the subject of conjecture (reviewed in [127]).

ESI of proteins in positive ion mode results in the protonation of protein analytes at accessible, basic residues. Accordingly, a feature of an ESI-MS spectrum of a given protein is the presence of multiply charged and thus quasi-molecular ions, where each charge state is represented by $[M+zH]^+$ (where z is the number of charges).

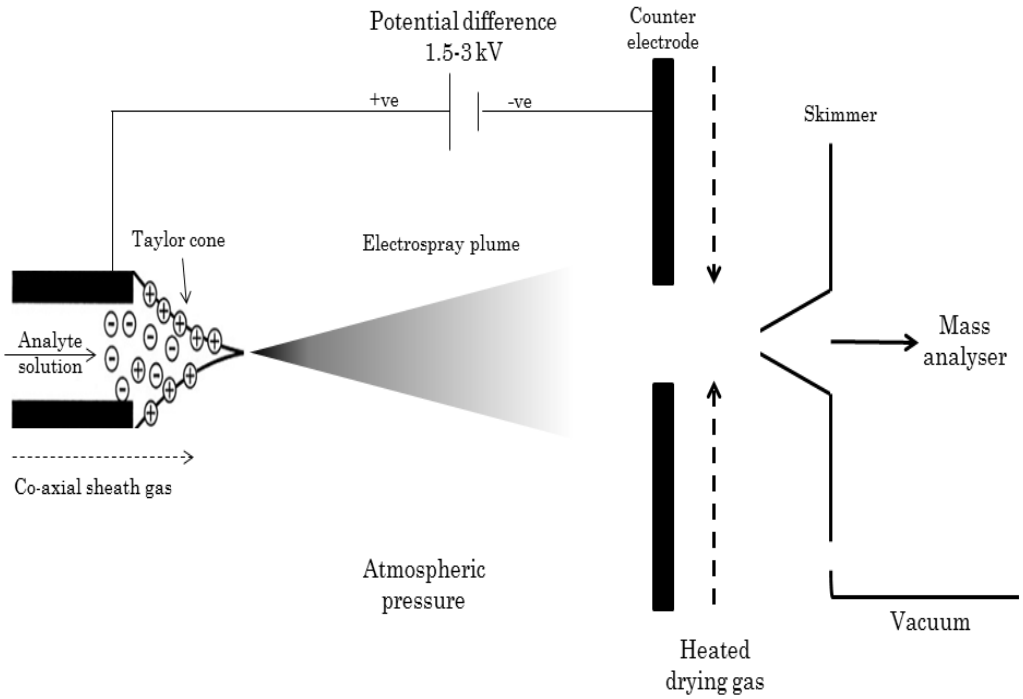


Figure 1.9 Schematic of an ESI source. The strong electric field applied to the tip of the ESI capillary results in the formation of a Taylor cone, from which a fine jet of charged droplets is ejected. Nebulisation is aided by co-axial sheath gas (usually nitrogen). The positively charged ions are attracted to the counter electrode, where they are accelerated into the mass spectrometer (adapted from [128]).

A further extension of ESI, nano(n)ESI (utilised extensively in the investigations in this thesis), has been introduced which proceeds by the same ionisation mechanism as ESI, but permits lower flow rates and thus reduced sample consumption. The ESI capillary indicated in Figure 1.9 is replaced with a fine borosilicate glass capillary, coated in a conductive metal (typically gold or silver).

1.3.1.2 Mass analysers

Following ionisation, gas-phase ions are subject to manipulation in static, or dynamic, electric and magnetic fields in mass analysers, according to *m/z* ratio. A number of analysers have been developed which rely on differing mechanisms of separation, as summarised in Table 1.2. Ion motion and manipulation according to *m/z* in these mass analysers is well established and has been reviewed extensively in [129].

Table 1.2 Typical mass analysers used in MS instruments (adapted from [129])

Analyser	Symbol	Principle of separation
Quadrupole	Q	*m/z* (trajectory stability)
Linear ion trap	LTQ	*m/z* (resonance frequency)
Time-of-flight	TOF	Velocity
Fourier transform ion cyclotron resonance	FTICR	*m/z* (resonance frequency)
Orbitrap	OT	*m/z* (resonance frequency)

1.3.1.2.1 Hybrid mass spectrometers

Each mass analyser offers contrasting features, operating at differing mass range limits, resolving power, analysis speed, transmission and mass accuracy. Consequently, evolution of MS instrument design, as illustrated in Figure 1.8, has seen the coupling of multiple mass analysers, to negate the limitations of isolated mass analysers, and thus create hybrid instruments. With regards to the analysis of proteins, common examples of hybrid instruments include the Quadrupole (Q)-Time-of-flight (TOF) instrument, which exploits the theoretically unlimited mass range of the TOF analyser to permit the analysis of large proteins and protein-complexes. Conversely, hybrid instruments that couple linear ion traps (LTQ) to either Orbitrap (OT), or Fourier transform ion cyclotron resonance (FTICR) analysers, offer unrivalled mass accuracy and resolution, albeit on a limited mass range, and thus prove excellent for the accurate determination of peptide mass, which has applications in bottom-up proteomics (see *section 1.3.2.1*).

1.3.1.2.2 Tandem mass spectrometry

Tandem arrangements of mass analysers whether in space, as with hybrid instruments, or indeed time, permits multiple rounds of mass analysis (MS/MS or MS^n), which when coupled to an ion dissociation process allows the *de novo* sequencing of proteins (or typically peptides).

In the first stage of mass analysis, a ‘precursor’ peptide ion of specific *m/z* is isolated, and is then activated (often in a collision cell) to yield a series of ‘product’ ions. The most common dissociative process is collision induced dissociation (CID), where ions are accelerated into inert buffer gas (often nitrogen), leading to the transfer of collisional kinetic energy to internal energy. This promotes the fragmentation of the ‘precursor’ ion in a well-documented manner at the amide bond (see Figure 1.10) [130], to yield a series of ‘product’ ions, which typically differ by one amino acid. The ‘product’ ions then undergo a second round of mass analysis and are detected (see *section 1.3.1.3*). From the resulting MS/MS spectrum, based on the mass difference between ‘product’ ions, peptide primary sequence can be determined. While this can be achieved manually, typically MS/MS raw data is submitted to bioinformatics tools which match observed peptide ‘product’ ion series to protein sequence databases. A score is then derived from this comparison to infer the statistical significance of the match between the spectrum and the database sequence, thus affording protein identification [131].

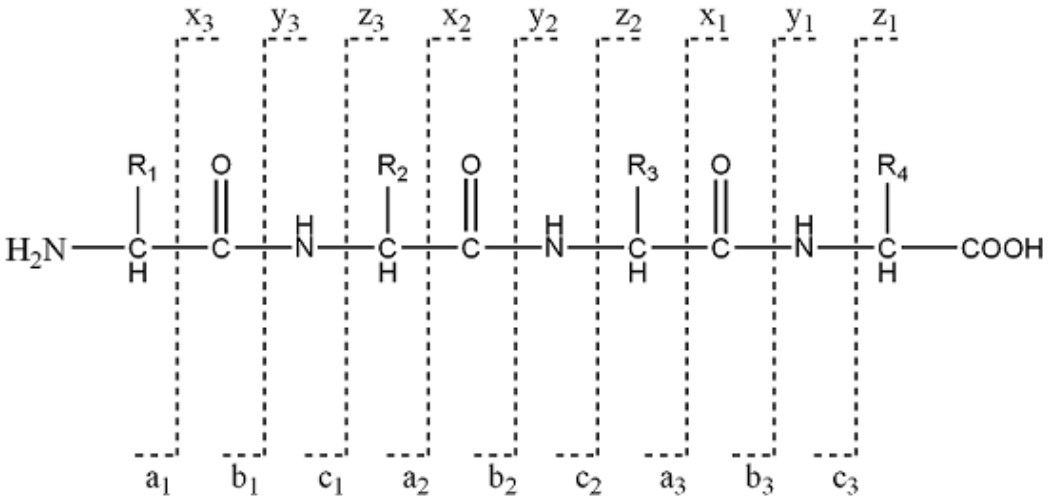


Figure 1.10 Peptide fragmentation mechanisms. Low energy CID results in the fragmentation of amide bonds into b and y ions, whereas higher energy conditions can result in alternative fragmentation pathways [130].

1.3.1.3 Ion detection

Following successful transmission through the mass analyser, ions are detected and transformed into a mass spectrum. The mechanism of detection is dependent upon the design of the instrument, and indeed the mass analyser utilised in question. The classic method of detection utilises electron multipliers, which detect and amplify the signal of incident ions as they strike the detector plate, to give an electric current. However, OT or FTICR mass analysers utilise a pair of metal plates within the mass analyser, which are in close proximity to ion trajectories, to monitor image current, a time-dependant signal of ion motion [129].

1.3.2 Experimental applications of mass spectrometry

With the key components of the mass spectrometer established in *section 1.3.1*, we now discuss the applications of MS with respect to the analysis of proteins.

1.3.2.1 Proteomics

To date, the evolution of MS has perhaps had the greatest impact on the field of proteomics, wherein MS has been utilised to identify and catalogue proteomes from a range of organisms. In the classic 'bottom-up' proteomics approach, the specificity of proteases (often trypsin, which cleaves C-terminally of Arg and Lys residues) is exploited to digest complex mixtures of proteins (i.e. cell lysates or less complex fractions) to generate predictable peptides. The peptide mixture is subsequently fractionated by LC (or alternative separation techniques) [132, 133], and analysed by MS/MS, as described in *section 1.3.1.2.2*. Using this approach, entire proteomes containing thousands of proteins can be sequenced during a single analysis. Furthermore the development and implementation of quantitative bottom-up MS approaches, has permitted the detection of dynamic changes in such proteomes.

Quantitative techniques include those that utilise stable isotope labelling or alternately are label free, such as spectral counting [134].

1.3.2.2 Structural mass spectrometry

Alongside proteomic strategies, which permit the high-throughput identification and quantitation of protein interaction networks [135], MS can also be used to probe protein structure and interactions, which will now investigate [136, 137].

1.3.2.2.1 Preserving protein quaternary structure in the gas-phase

Following the introduction of ESI, research in the early 1990's demonstrated that protein assemblies could be preserved and transferred intact from solution into the vacuum of the mass spectrometer, permitting mass analysis [123]. However, as the gas-phase represents an environment devoid of water and thus stabilizing hydrophobic effects, questions arose regarding the extent to which native protein structure is retained. Critically, studies summarised by Breuker and McLafferty [138], indicated that while structural alterations (typically compaction), owing to surface side-chain collapse, can occur within picoseconds of desolvation, for many proteins gross structural reorganization does not appear to take place on the time frame (<50 ms) of MS measurements. Furthermore, structural alterations in the gas phase appear, as one might expect, to be protein dependent. Indeed, flexible proteins, with large voids or cavities, appear prone to considerable collapse post desolvation, as typified by the chaperone protein GroEL [139], amongst others [140]. However, with the exception of these extreme examples, providing instrument conditions are carefully optimised, proteins can survive the ESI process in solution-like conformations to yield valuable structural information [141-147].

This has paved the way for the field of gas-phase structural biology and in particular native MS, in which large, dynamic or heterogeneous protein complexes can be detected

in a rapid and sensitive manner. As this technique has developed, it has found application to increasingly larger and more intricate multi-protein assemblies, extending from large (upto several megadaltons [148]) recombinant protein complexes, including chaperone complexes [149-151], intact viruses [144, 152, 153] and polydisperse proteins such as αB-crystallin [154], to endogenously expressed protein complexes [155-160]. Furthermore, in addition to offering insight into the oligomeric state of assemblies, native MS has also been utilised to provide a qualitative or indeed quantitative assessment of the stoichiometry of ligand binding [161].

1.3.2.2.2 Ion mobility-mass spectrometry

Alongside native mass spectrometry, there has been interest in developing MS-based approaches for obtaining three-dimensional information on proteins and their complexes. This has led in recent years to the coupling of ion mobility spectrometry (IMS) to MS (IM-MS).

IMS is essentially a gas-phase electrophoretic technique, in which ions are separated according to differential mobility in a drift cell containing an inert gas, which acts to impede the propagation of ions through the cell. Three IMS techniques are typically used in IM-MS (and are reviewed in [162]): drift-time ion mobility spectrometry (DTIMS), field-asymmetric ion mobility spectrometry (FAIMS) and travelling-wave ion mobility spectrometry (TWIMS), which is utilised in the investigations in this thesis [163].

The TWIMS device and the principle of separation are depicted in Figure 1.11. Essentially the TWIMS device permits separation of ions according to their averaged collisional cross section (CCS), a measure of the effective area of interaction between a given ion and the neutral gas through which the ion is travelling. This CCS can in turn can be related to the three dimensional conformation (size/shape) in absolute values of angstrom or nanometers [164]. Importantly, comparative studies have revealed that the

TWIMs device does not compromise mass spectrometer sensitivity [165], thus IM-MS, where the mass spectrometer has an in-built TWIMS device, permits separation on the size-to-charge axis, as well as the mass-to-charge axis. Accordingly, this approach has found extensive application in the characterisation of protein structure and interactions [146, 166-168].

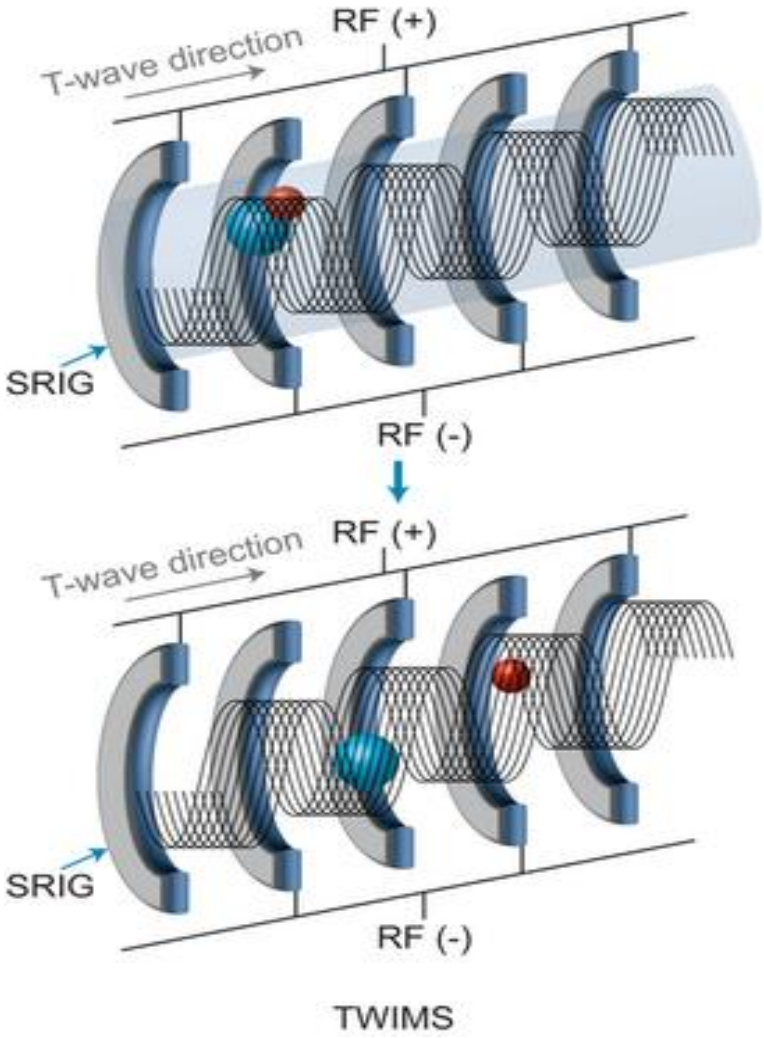


Figure 1.11 Cartoon of a TWIMS cell. The gas-filled TWIMS mobility cell comprises a set of stacked ring ion guides (SRIG), through which ions are transmitted. Ions are confined radially by radio frequency (RF) voltages, applied of opposite phase to consecutive rings. Direct current (DC) voltage ‘pulses’ are applied sequentially to each electrode, and act as travelling waves which axially pushes ions from one end of the device to the other. By manipulating the parameters of the DC travelling waves, separation of ions is achieved. Higher-mobility ions are 'carried' by the wave, whereas ions of lower mobility 'roll over' the wave, thus taking longer to move through the gas-filled mobility cell. Blue and red spheres represent large and small ions respectively (adapted from [162]).

1.3.2.2.3 Molecular dynamics

The ability to draw meaningful conclusions from experimental IM-MS data has been supported in recent years, following the development and application of protein molecular dynamics (MD) approaches. These computational simulations model the physical movements of atoms, the co-ordinates of which have been derived by high-resolution structural techniques (NMR, X-ray etc), within designated parameters (including *in vacuo*), to probe the accessible conformational landscape of a protein. The output structures from MD can be inputted into any of three algorithms that have emerged, to calculate theoretical CCS of a protein. This theoretical CCS can in turn be compared to IM-MS derived experimental CCS values, to probe gas-phase structure/conformation. The three algorithms include the projection approximation method (PA), the trajectory method (TM) and the exact hard sphere scattering (EHSS). While the TM method proves to be the most realistic model of protein CCS, typically the PA method is used to determine CCS, as this proves to be the most accessible computationally, and can in turn be converted to a TM-like number by applying an empirical correction factor. Critically, as indicated in a recent review by Benesch and Ruotolo [169], excellent correlation is displayed between experimentally determined CCS, measured by IM-MS, and calculated values resulting from the above *in silico* methods, demonstrating that solution protein quaternary structure in many cases is retained in the gas phase. Furthermore, where the conformation of a protein, or a protein assembly is unknown, the combination of experimentally derived CCS data, with those obtained by molecular modelling and/or molecular dynamics simulations can provide a good prediction of the real structure or conformation.

1.3.2.2.4 The structural mass spectrometry tool box

In parallel to native IM-MS which explicitly determines the shape/size of proteins of interest, a number of complementary MS-based techniques have emerged, including

hydrogen-deuterium exchange (HDX), covalent labelling and cross-linking, which contribute to the dissection of protein structure and interactions. These approaches are largely beyond the scope of the investigations in this thesis, but are reviewed extensively elsewhere [170, 171]. Despite information afforded by such structural MS approaches being of inherent low resolution, these approaches offer a number of unique advantages with respect to the classic high-resolution techniques including; NMR, Cryo-EM and X-ray crystallography. Indeed while these established high-resolution structural techniques provide valuable detail on protein interactions at the molecular (and atomic) level, with few exceptions [172], analysis of large, multi-protein complexes remains challenging. Accordingly, the area of greatest potential impact of structural MS in structural biology resides in areas where other, more established techniques are of limited value, due to protein/protein-complex size, heterogeneity, or time.

1.3.3 Applications of mass spectrometry to probe the ubiquitome

Due to the aforementioned developments in MS strategies to probe protein networks and interactions, MS has progressed into an excellent platform for the systematic characterization of polyubiquitin modification. We now review the MS strategies and subsequent insights afforded by previous MS-based analysis of the ubiquitome.

1.3.3.1 Ubiquitin trypsinolysis

As alluded to in *section 1.3.2.1*, trypsin, which cleaves C-terminally of Lys or Arg residues, is often used to enzymatically cleave protein mixtures into peptides mixtures prior to MS/MS analysis in the classic 'bottom-up' approach. This strategy has been widely utilised to catalogue the proteins and indeed the sites of ubiquitylation in the ubiquitome.

Trypsin cleavage of ubiquitin-modified proteins, as illustrated in Figure 1.12, results in the cleavage of the conjugated ubiquitin moiety at Arg-74 to yield a Gly-Gly modified lysine residue on the substrate. The Gly-Gly stub prevents trypsinolysis at the modified Lys residue, resulting in missed trypsin cleavage. Critically, the missed cleavage and the addition of the Gly-Gly stub, which has a mass of 114.029 Da, can be exploited by bioinformatics tools and thus forms the basis of identification of ubiquitylated proteins and modification sites by MS. In the same manner, modified Lys residues in polyubiquitin chains are also spared from trypsin cleavage to produce polyubiquitin signature peptides, accordingly these can be used to confirm and quantify sample polyubiquitin linkage profile.

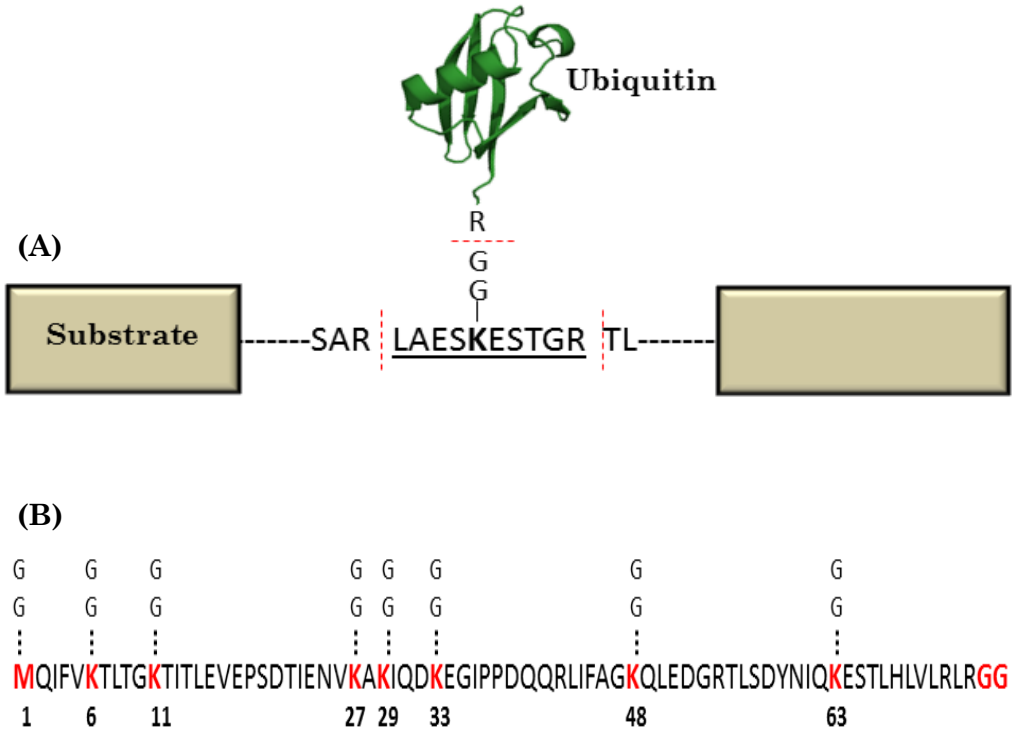


Figure 1.12 Ubiquitin trypsinolysis. (A) Schematic of the trypsin digest of a ubiquitin modified substrate. Red dashed line indicates sites of trypsin cleavage. (B) The primary sequence of ubiquitin. Sites of ubiquitin modification are highlighted in red, with Gly-Gly stub resulting from trypsin cleavage indicated.

1.3.3.2 Detecting ubiquitin modified proteins

As ubiquitylated proteins are often short-lived or of low abundance, prior to 'bottom-up' proteomics, enrichment strategies, as discussed in *section 1.2*, are required - including those which utilise epitope tagged ubiquitin, UBDs and antibodies. The first large-scale identification of ubiquitin-modified substrates utilised His-tagged ubiquitin to enrich ubiquitinated substrates in yeast. This approach led to the identification of 1075 proteins, with 110 ubiquitylation sites [34]. Subsequently, this approach has been applied with increasing sensitivity to mammalian cells and animal models [101, 173-175].

A key advance with regards to the identification of ubiquitylated proteins, which as of 2010 was ~250 in mammalian systems based on analysis of the Swiss-Prot database, was the development of an antibody specific to the Lys-Gly-Gly stub on ubiquitylated proteins, which permits the immunoprecipitation (IP) of these peptides post-tryptic digest [176]. This has led to a number of studies, most recently yielding the identification of ~19000 ubiquitylated sites, within ~5000 proteins from a mammalian cell line [60]. Despite also detecting ubiquitin-like proteins including NEDD8 and ISG15, which also impart a Gly-Gly stub to modified substrates, the same study indicated that modification with such ubiquitin-like proteins accounts for a negligible portion of the ubiquitome [60].

1.3.3.3 Quantitation of the ubiquitome

Dynamic changes to ubiquitylation patterns underlie many physiological processes, thus the application of quantitative MS has proved valuable. Relative quantification using label free [177] and stable isotope labelling with amino acids in cell culture (SILAC) approaches have been broadly used, for example in attempts to monitor ubiquitome dynamics in mammalian cells following the introduction of a range of E_3-ligase mutants [178]. However, owing to intrinsic LC-MS/MS variations, which

complicate label-free methods, SILAC has found more wide spread use [179]. A number of studies have utilised SILAC following; protease inhibition in mammalian cells [174], the introduction of ubiquitin mutants in yeast [100], and to probe a subset of ubiquitin conjugates recognized by Rpn10, a ubiquitin receptor of the yeast proteasome [180]. The gold standard for the identification and quantitation of the ubiquitome is now to combine SILAC with Lys-Gly-Gly IP. This strategy affords proteome-wide, site-specific quantification of endogenous ubiquitylation sites [181-184].

1.3.3.4 Detecting polyubiquitin chains

Quantitation has also been applied to probe the absolute quantity of each polyubiquitin signature peptide (highlighted in *section 1.3.3.1*) within a sample. In this approach, termed ubiquitin absolute quantification (Ub-AQUA), isotopically labelled peptides, which mimic the products of polyubiquitin trypsinolysis, are spiked into a sample prior to digestion and MS analysis [185]. This strategy has been utilised to confirm the presence of all seven ubiquitin isopeptide linkages in yeast [35] and mammalian cell extracts [186]. Additionally, Ub-AQUA has been useful in quantifying the products of *in vitro* ubiquitylation reactions [187], modifications on substrates *in vitro* [188], enriched from mammalian [189, 190] and yeast [191] extracts, as well as the roles of atypical ubiquitin chains in proteasomal degradation [35]. However, while the Ub-AQUA strategy has been insightful, the polyubiquitin landscape is likely to be dynamic, permitting response to specific stimuli (i.e. stages of the cell cycle). Furthermore, the method of purification is also likely to influence absolute values of chains. To account for loss of proteins through fractionation and processing, the ubiquitin protein standard absolute quantification (PSAQ) approach was developed [108]. Here, intact isotopically labelled ubiquitin or ubiquitin-modified proteins are spiked into samples, which are subsequently affinity enriched on non-selective UBDs including the UBA domain of human PLIC2 and the ZnF-UBP domain of USP5. This approach was utilised to

determine total ubiquitin concentration, and to monitor changes following the proteasome inhibition [108].

1.3.3.5 Middle/top down proteomics

The bottom-up approach, whilst insightful, proves limited with respect to the analysis of heterogeneous polyubiquitin chains, owing to the loss of connectivity upon trypsin digest. To address this issue, middle-down (limited proteolysis) and top-down (intact protein analysis) MS approaches have been developed [192-195]. These approaches were critical to identifying forked polyubiquitin chains, in which two ubiquitin moieties are linked to adjacent Lys residues on the proximal ubiquitin, [193, 196] and also cyclised Lys-48 linked diubiquitin which are lacking a free C-terminus [197]. Additionally top-down approaches contributed to determining the topology and structural organisation of the 19S proteasome lid in yeast [155].

1.3.3.6 Native mass spectrometry

Finally, as alluded to in *section 1.3.2.2*, native MS has emerged as a technique to probe protein structure and interactions, and has found application in the examination of UBD:(poly)ubiquitin interactions. Indeed Sokratous and colleagues [198] demonstrated that native MS could be applied in a rapid and sensitive manner to probe the affinity and linkage selectivity of UBDs, including the ZnF-UBP domain of USP5 and the hHR23A UBA2 domain, for mono- and polyubiquitin substrates. Reassuringly, the binding affinities established by alternative techniques, such as Isothermal titration calorimetry (ITC), were in good agreement with those derived by MS in this study.

We can conclude that MS based strategies to probe the ubiquitome have proven extremely insightful, and that the continued development and refinement of such strategies will offer new insights into the ubiquitome.

1.4 Unanchored polyubiquitin chains

The notion that ubiquitin signalling is mediated solely by the covalent modification of substrates has, as alluded to in *section 1.1.2*, been challenged. Indeed, a landmark paper by Xia *et al* [11] revealed that unanchored, or substrate-free polyubiquitin chains could interact directly and non-covalently with the TAB2 ubiquitin receptor, to activate the protein kinase TAK1, to regulate downstream NF-κB signalling. Subsequently a number of studies have ensued, highlighting the physiological significance of the unanchored ubiquitin pool. Accordingly, as this unanchored ubiquitin pool is central to the investigations in this thesis, in *section 1.4* we review the literature to date, exploring the source and emerging roles of unanchored polyubiquitin.

1.4.1 An unanchored ubiquitin reservoir

Unanchored polyubiquitin has long been identified, indeed early investigations in *S. cerevisiae* by Kimura *et al* [199] sought to rationalise the presence of an unanchored ubiquitin pool as a simple source of monoubiquitin, for the ubiquitylation of substrates in a signal-dependent manner, for example following cellular insult. Kimura *et al* [199] noted that cellular heat stress resulted in the depletion of the unanchored polyubiquitin pool, as regulated by the DUB, Doa4, and its inhibitor, Rfu1. Subsequent studies have revealed depletion of the unanchored polyubiquitin pool to be a common feature of other cellular insults, including DNA damage and oxidative stress [200].

Interestingly Li et al [201] have also revealed that unanchored polyubiquitin chains can be directly conjugated *en bloc* to target substrates. Thus collectively, these studies support the notion that unanchored polyubiquitin may act as a pre-assembled reservoir of ubiquitin chains, which permits the rapid ubiquitylation of selected substrates. Indeed we can draw mechanistic parallels to the intracellular storage of glucose as glycogen.

1.4.2 Unanchored polyubiquitin and the proteasome

Concurrently, specific physiological roles have also emerged for unanchored polyubiquitin. Indeed, as early as 1997 unanchored Lys-48 linked polyubiquitin chains were implicated in the regulation of 26S proteasome activity (*via* competitive inhibition of substrate recognition) [202]. Subsequent studies in yeast [203] and mammalian cells [204] support this concept, revealing suppression of USP5, the DUB responsible for the selective disassembly of unanchored polyubiquitin, to result in the accumulation of unanchored polyubiquitin chains as well as critically, defects in ubiquitin-dependent proteolysis. Furthermore proteasome-associated ubiquitylation machinery, including DUBs (such as; UBP6, USP1, USP5, POH1 and Uba1 [205-208]) and E3-ligases (UFD4 and HUL5 [200, 209]) have been identified as capable of generating unanchored polyubiquitin chains in a signal-dependent manner, thus broadly supporting the notion that proteasomal activity may be regulated by unanchored polyubiquitin chains.

1.4.3 Unanchored polyubiquitin in aggresome biogenesis

Upon impairment of proteasomal function, unanchored polyubiquitin chains have also been implicated in the biogenesis of aggresomes, a key mechanism which ensures that degradation of misfolded, ubiquitylated protein aggregates is completed. In this pathway, illustrated in Figure 1.13, the action of DUBs, including POH1 and Ataxin-3 expose unanchored Lys-63 linked polyubiquitin chains in protein aggregates. Such unanchored chains are subsequently recognised by a ZnF-UBP domain, found in histone deacetylase 6 (HDAC6), which governs the formation of aggresomes. Indeed, the recruitment of HDAC6 facilitates the transport of protein aggregates to the aggresomes, using the cellular microtubule network, which subsequently dispose of the resulting protein aggregates by autophagy [210-212]. This aggresomal pathway, or the hijacking thereof, has also been highlighted to being critical to the pathogenesis of the Influenza A virus [213].

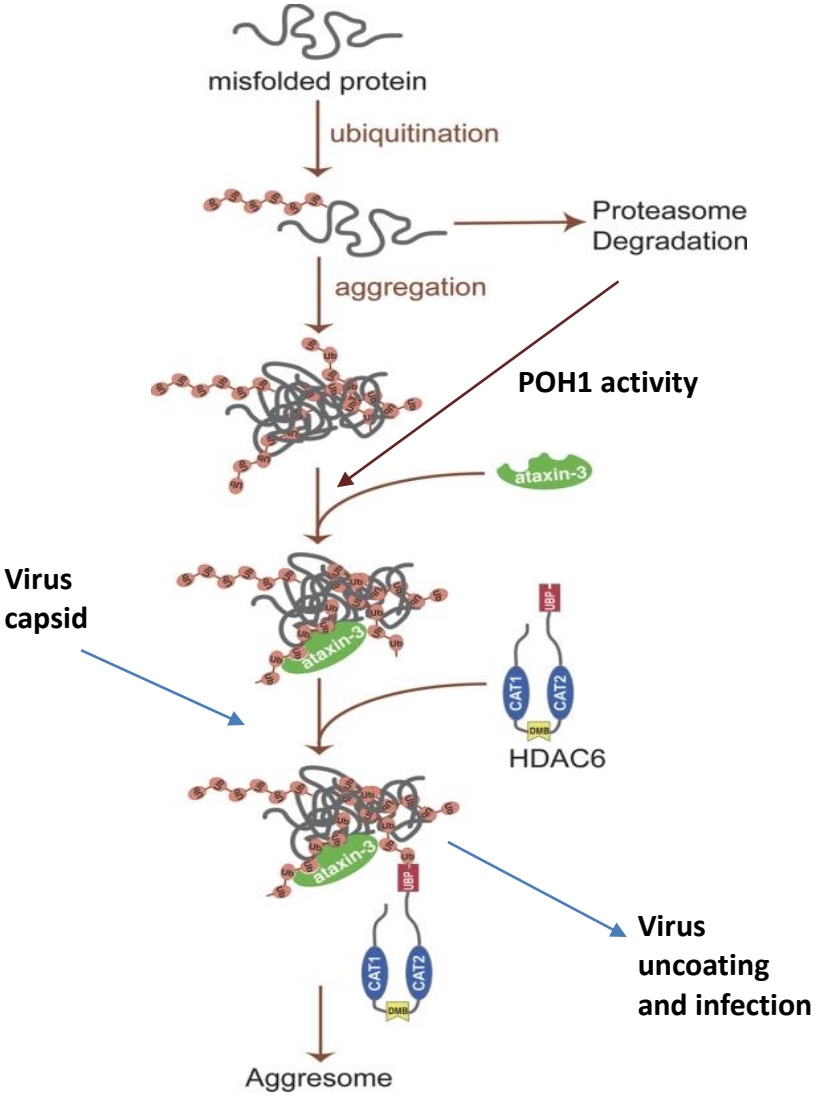


Figure 1.13 Unanchored ubiquitin and the aggresome response. Under normoxia, polyubiquitylated proteins are efficiently degraded by the proteasome. However upon impairment of proteasomal function, polyubiquitylated proteins accumulate to form protein aggregates. DUBs such as POH1 or Ataxin-3 interact with such aggregates to release unanchored ubiquitin, which in turn can be recognised by the ZnF-UBP domain of HDAC6, resulting in recruitment of aggregates to the dynein motor complexes and then aggresomes. The influenza virus has been demonstrated to exploit this pathway, resulting in virus uncoating and infection [213] (Figure adapted from [211]).

1.4.4 Unanchored polyubiquitin and innate immune signalling

Alongside roles in the processing of proteins for degradation, unanchored polyubiquitin chains have also been highlighted as important signalling molecules in the activation of NF-κB, as well as other critical transcription factors, central to the innate immune response.

Pioneering investigations by Xia *et al* [11] revealed unanchored Lys-63 linked polyubiquitin chains, generated by the E_3 TRAF6 and E_2 enzyme complex Ubc13-UEV1A, to be implicated in the activation of TAK1 following IL-1 stimulation, as illustrated in Figure 1.14 [11]. Furthermore, in the same study it was revealed that TNFα stimulation resulted in the activation of NF-κB, but was dependent upon Ubc5, an E_2 enzyme that is able to synthesize unanchored polyubiquitin chains with isopeptide linkages other than Lys-63 or Lys-48, thus suggesting that activation of NF-κB activation may utilize ubiquitin chains containing different linkages, depending upon the activation pathway. The activation of NF-κB by unanchored polyubiquitin has subsequently been probed in a number of studies, the findings of which are expertly reviewed in [214].

In addition to the unanchored ubiquitylation machinery highlighted by Xia *et al* [11], including TRAF6, a number of members of the TRIM E3-ligase family have been demonstrated to mediate innate immune cell signalling, and specifically the anti-viral response via the synthesis of unanchored polyubiquitin[215]. Indeed, TRIM5 [216, 217], TRIM6 [218]and TRIM25 [219-221] have each been implicated in the synthesis of unanchored polyubiquitin chains of assorted polyubiquitin linkage. These unanchored polyubiquitin chains can subsequently be delivered to receptors such as RIG-I [219, 220] or MDA5 [221], to initiate the activation of the NF-κB, IRF3 and AP-1 signalling cascades.

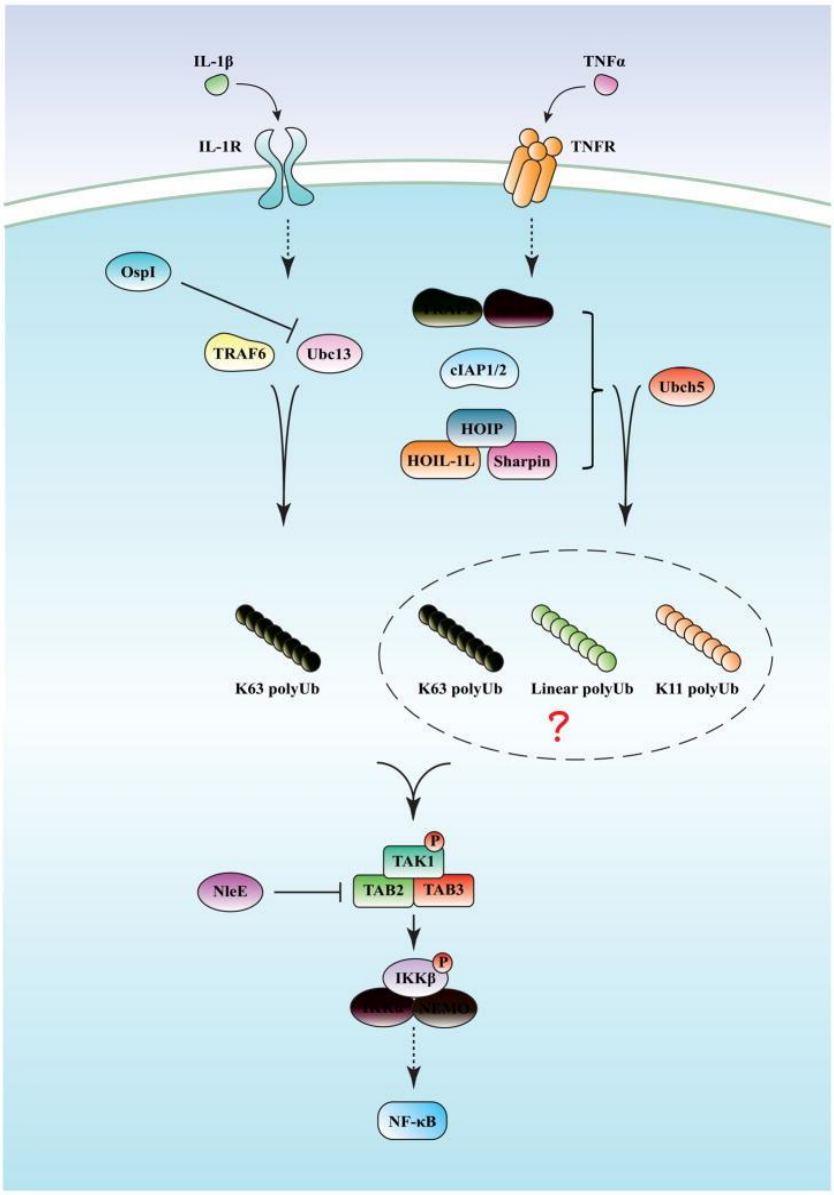


Figure 1.14 The role of unanchored polyubiquitin chains in the activation of NF-κB, following cytokine stimulation. Stimulation of cells with IL-1β leads to the activation of the ubiquitin E3 ligase TRAF6, which catalyses the synthesis of unanchored Lys-63 polyubiquitin chains that bind to the TAB2 subunit of the TAK1 kinase complex, resulting in TAK1 activation. TAK1 then phosphorylates IKKβ, leading to NF-κB activation. In the TNFα pathway, stimulation of the cells leads to the recruitment of several ubiquitin E3 ligases to the receptor, including TRAF2, TRAF5, cIAP1, cIAP2 and LUBAC (consisting of HOIP, HOIL-1, and Sharpin). These E3s and Ubch5 synthesize ubiquitin chains of different linkages, including K63, K11 and linear (M1). Two bacterial proteins, OspI and NleE, are able to inhibit NF-κB activation by targeting Ubc13 and TAB2/3, respectively (adapted from [214]).

1.4.5 The unanchored polyubiquitin code

Finally, while prominent roles for Lys-48 and 63 linked unanchored polyubiquitin chains have emerged, it is interesting to note that the E_2/E_3 pair Ubch5c/ TRAF6, utilised in the synthesis of unanchored Lys-63 linked polyubiquitin [11], have also been noted in the synthesis of unanchored polyubiquitin chains linked by Lys-11 [222] or Met-1 [223], hinting at broader roles for unanchored polyubiquitin chains of atypical linkage.

Also very recently the functional significance of unanchored mixed-linked chains has been highlighted in the activation of NF-κB signalling. Indeed, the E3 ligase Tax, in combination with an array E_2 enzymes (UbcH2, UhcH5c, or UbcH7) can synthesise unanchored mixed-linked polyubiquitin chains, which bind to the NEMO domain of IKK to activate this pathway. While conversely homogenous unanchored polyubiquitin chains linked Lys-48, 63 or Met-1, were found to be poor activators [224]. This highlights the significance of ubiquitin linkage within the unanchored ubiquitin pool, and indicates that mixed linked chains are functional

1.5 Summary and objectives

Just as conjugated polymers of ubiquitin, which vary in length and topology, can be considered to embody a 'ubiquitin code', it is feasible to theorise that this may be reflected in the unanchored ubiquitin pool. Indeed, we propose that there are pools of unanchored polyubiquitin in cells of unknown composition and function. However despite this, strategies that afford the purification of unanchored polyubiquitin have until recently, as we discuss in *chapter 3*, been unavailable. Thus, detail on the molecular composition and interactions of the unanchored polyubiquitin pool have been to date, lacking. Accordingly, in the investigations presented in this thesis, we seek to explore the biophysical and biochemical nature and composition of unanchored polyubiquitin, through the development of specifically engineered ubiquitin-binding proteins, and the use of mass spectrometry.

In first results chapter (chapter 3) we aim to develop strategies and tools to favour the selective purification of unanchored polyubiquitin, and thus complete the first comprehensive molecular analysis of the unanchored ubiquitin pool, as purified from mammalian cell extracts. To fulfil this aim, we utilise the UBDs of the DUB USP5 in an affinity enrichment strategy. We then probe the composition of the purified unanchored ubiquitin using both polyubiquitin linkage-selective antibodies and mass spectrometry.

Subsequently, in chapter 4 we seek to rationalise the mechanisms of USP5 unanchored ubiquitin recognition. This is achieved utilising a novel and comprehensive structural MS-based workflow, including native IM-MS and carbene footprinting.

Finally. in the last results chapter (chapter 5), we present the rational design and validation of a tandem UBD (t-UBD) hybrid protein, which for the first time utilises two different UBDs fused in a linker optimised manner, to exploit avidity effects that afford the recognition and purification of Lys48-linked unanchored polyubiquitin

chains. Utilizing native MS to quantitatively probe binding affinities, we confirm cooperative binding of the UBDs within the synthetic protein, and desired binding specificity for Lys48-linked ubiquitin dimers. Furthermore, MS/MS analyses indicate that the t-UBD, when applied as an affinity enrichment reagent, can be used to favour the purification of endogenous unanchored Lys48-linked polyubiquitin chains from mammalian cell extracts.

Chapter 2 - Materials and methods

In *chapter 2* the materials and methods used in the investigations are described. All solutions were prepared with Milli-Q grade deionised water. Reagents, unless stated, were obtained from Sigma Aldrich. Detailed buffer compositions can be found in Appendix I. Bench tops were routinely cleaned with Distel (Tristel) and wiped with 70 % ethanol. Bacteriological work was completed in close proximity to a Bunsen burner. Glassware, pipette tips and growth media were sterilised by autoclaving at 121 °C for 15 minutes. Heat sensitive reagents were filter-sterilised using 0.22 μM syringe filters (Sartorius). Disposable plasticware including Petri dishes, Universal tubes, Falcon tubes, syringes and needles were supplied sterile.

2.1 Molecular biology methods

2.1.1 Bacterial expression vectors

The bacterial expression vectors (plasmids) encoding the proteins investigated in this thesis are listed in Table 2.1, the primary sequences they encode are detailed in Appendix II. With the exception of the synthetic t-UBD gene (discussed in *section 2.1.2*), all constructs were previously cloned by members of the Layfield group.

The pGEX-4T-1 (GE Healthcare) vector was predominantly used to express proteins of interest (with the exception of ubiquitin) as N-terminal glutathione-S-Transferase (GST)-fusions, improving solubility and aiding purification. Owing to a thrombin recognition sequence, the GST tag can be cleaved from the protein of interest with thrombin, yielding the protein of interest with two residual amino acids at the N-

terminus (Gly-Ser) (as illustrated in Figure 2.1). All plasmids confer ampicillin resistance, which provides the basis for positive selection of plasmid transformed cells. Ampicillin was added to a final concentration of [286 μM] (equivalent to 100 μg/ml) to cultures.

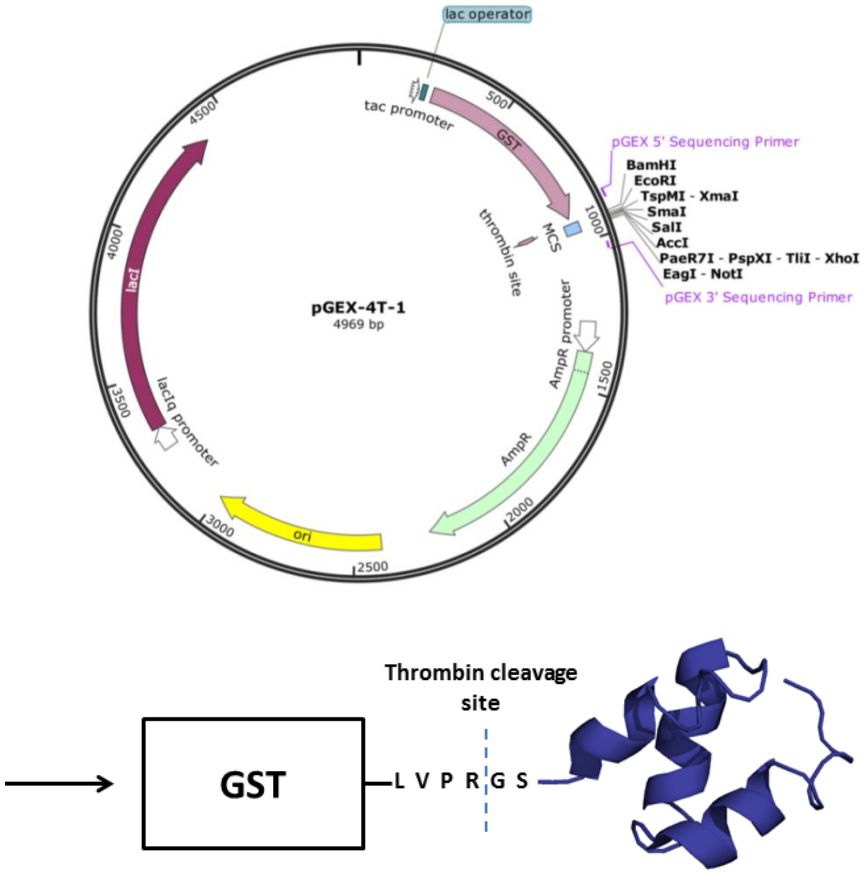


Figure 2.1 Plasmid map for pGEX-4T-1 bacterial expression vector. Overexpression results in the synthesis of N-terminally GST-tagged proteins. A Thrombin cleavage site is incorporated at the C-terminus of the GST-tag, which permits cleavage of the tag (indicated by dotted line) to yield protein of interest, with a Gly-Ser remnant at the N-terminus.

Table 2.1 Bacterial expression constructs. All cDNAs were isolated from *H. sapiens*.

Protein encoded	Plasmid	Cloning site	Fusion partner
USP5, Short Isoform (1-835)	pGEX-4T-1	*Bam*HI-*Xho*I	GST
ZnF-UBP (USP5) (163-291)	pGEX-4T-1	*Bam*HI-*Xho*I	GST
UBA2 (hHR23a) (314-362)	pGEX-4T-1	*Bam*HI-*Xho*I	GST
t-UBD (Synthetic protein, see *chapter 5*)	pGEX-4T-1	*Bam*HI-*Xho*I	GST
Monoubiquitin (1-76) and surface mutants (ΔGly75Gly76)	pkk223-3	*Eco*RI-*Hind*III	Untagged

2.1.2 t-UBD cloning

A synthetic tandem UBD (t-UBD) containing gene was designed and assembled as discussed in *chapter 5*. A synthetic gene, provided in a pBSK vector, was sourced commercially (Dundee Cell Products), with *Bam*HI and *Xho*I restriction sites at respective termini. 500 ng of empty pGEX-4T-1 and the synthetic t-UBD construct were digested, for 2 hours at 37°C, with 1 µL of *Bam*HI and *Xho*I restriction enzymes (NEB), 2 µL CutSmart buffer (NEB) and made up to 20 µL in water. Restriction digests were separated on a 1% Tris –acetate-EDTA (TAE)-agarose gel (see section 2.1.3), and extracted using a QIAquick gel extraction kit (Qiagen), according to the manufacturer's protocol, and DNA quantified using a NanoDrop ND-1000 spectrophotometer (Thermo Scientific). The empty pGEX-4T-1 vector and t-UBD insert were then ligated utilising the T4 DNA ligase (NEB), according to the manufacturer guidelines. Briefly ligations were setup as 25 µL reactions, with 50 ng of empty pGEX-4T-1 vector incubated with 16 ng of t-UBD insert (1:3, vector: insert molar ratio). Reactions were incubated for 16 hours at 16°C.

2.1.3 DNA electrophoresis

DNA electrophoresis was utilised to separate restriction digest products, as well as to approximate the size of the plasmid constructs. 1% agarose reconstituted in Tris-acetate-EDTA (TAE) buffer was microwaved until dissolved. After cooling, SYBR safe (5 µL/100 ml gel, Thermo Fisher) was added and the mixture poured into a gel unit, and a comb added (Geneflow). Once set, the gel was placed in a gel tank (Geneflow), and samples, diluted into DNA loading buffer (NEB), were loaded alongside molecular weight markers (NEB). Gels were electrophoresed at a constant voltage of 70 V until suitable separation was achieved. DNA bands were detected and imaged using a Gene Genius bioimaging system (Syngene).

2.1.4 Heat shock transformation

Plasmid DNA, encoding proteins of interest, were heat shock transformed into *in-house* prepared Inoue-competent *E. coli* cells (see [225]). Briefly 5 μL of ligation or mutagenesis product (see *section 2.1.5*), or 1 μL of purified plasmid DNA (~100 ng/μL) was mixed with 100 μL of Inoue-competent XL-10 gold cells, and incubated on ice for 30 minutes. The mixtures were then incubated at 42 °C for 45 seconds (heat shock) and then immediately cooled on ice. Reaction mixtures were made up to 1000 μL in Luria broth (LB) and incubated at 37 °C for 1 hour whilst shaking. 250 μL of the mixture was then plated onto an LB-agar plate (see Appendix I for preparation notes) containing ampicillin and grown for ~16 hours at 37 °C.

2.1.5 Plasmid stocks, isolation and sequencing

From LB-agar plates, single colonies were selected and cultured in 10 ml LB-ampicillin for ~16 hours at 37 °C, whilst agitating at 180 rpm. 0.5 ml of the overnight culture was mixed with 0.5 ml of 30 % (w/v) glycerol and stored at -80 °C as a glycerol stock. The remainder of the culture was utilised to isolate plasmid DNA, using a plasmid Miniprep kit (Qiagen) according to the manufacturer's guidelines. The integrity of all constructs was verified by Sanger sequencing completed *in-house*.

2.1.6 Site-directed mutagenesis

Mutations to wild type (WT) constructs, indicated in Table 2.2, were introduced by site-directed mutagenesis, according to the method of Liu *et al* [226], a modified QuikChange protocol (Stratagene). All reagents were part of the QuikChange kit (Stratagene), while oligonucleotide primers were purchased from Sigma Aldrich. Mutagenesis reactions were set up in thin walled PCR tubes according to Table 2.3 in a Sensoquest thermocycler (Geneflow). Products were then incubated with *Dpn*I (NEB) for 1 hour at 37 °C to digest parental methylated DNA. All mutated constructs were

heat shock transformed into XL-10 gold Inoue competent *E. coli* cells, plasmid prepped and DNA sequenced.

Table 2.2 Mutations introduced to plasmid constructs

Protein	Mutation	Primer	Primer sequence (5' – 3')
USP5	Arg-$_{221}$-Ala	Forward	GTGGGGCACGCTACTTCGATG
		Reverse	TAGCGTGCCCCACAGAGGATG
	Cys-$_{335}$-Ala	Forward	GTAACAGCGCCTACCTCAACTCTG
		Reverse	AGGTAGGCGCTGTTACCCAGGTTC
	Met-$_{643}$-Glu	Forward	GCTGGTGGAGGAGGGATTCCCTATGGACGCC
		Reverse	AGGGAATCCCTCCTCCACCAGCTGGATGATG
	Met-$_{711}$-Glu	Forward	CATTGTCTCCGAGGGCTTCTCCCGGGACCAGG
		Reverse	GGGAGAAGCCCTCGGAGACAATGGTGGTCACAC
t-UBD	R_{110}A	Forward	GCTGCGGCGCCAACAGCCAGGAACAGCATGC
		Reverse	TTCCTGGCTGTTGGCGCCGCAGCCCTGATGG
	2 x Gly	Forward	TGATGGCGGCGTGTGCCGCCAT
		Reverse	AACTTTGATGATGGCGGCGTGT
	3 x Gly	Forward	TGATGGCGGCGGCGTGTGCCGCCAT
		Reverse	AACTTTGATGATGGCGGCGGCGTGT
	5 x Gly	Forward	GGCGGCGGAGTGTGCCGCCATATT
		Reverse	TCCGCCGCCGCCATCATCAAA
	6 x Gly	Forward	GGCGGCGGAGGTGTGTGCCGCCATATT
		Reverse	ACCTCCGCCGCCGCCATCATCAAA

Table 2.3 A typical (A) reaction set up and (B) PCR thermocycling conditions for site-directed mutagenesis

(A)

Component	Concentration	Volume (µL)
Reaction buffer	10 X	2.5
dNTP	10 mM	0.5
Forward primer	10 µM	2.5
Reverse primer	10 µM	2.5
Template	10 ng/µL	1
Pfu Ultra II DNA Polymerase	2 U/µL	0.5
dH2O	-	to 25 µL

(B)

Temperature (°C)	Time (minutes)	Cycles
95	3	1
95	1	12
60	1	
72	6	
55	1	1
72	30	1
4	hold	-

2.2 Protein expression and purification techniques

For protein overexpression, purified plasmid DNA was heat shock transformed into *in-house* Inoue competent BL21(DE3) *E. coli* cells (or C41(DE3) cells for ubiquitin expression) and incubated on LB-agar plates overnight. From the plates, single colonies were selected and cultured in 10 ml LB-ampicillin in universal tubes for ~16 hours at 37 °C, whilst agitating at 180 rpm. All bacterial incubations were completed in orbital shakers (Kuhner). These starter cultures were used to inoculate further growth media for protein expression.

2.2.1 Small-scale bacterial protein expression tests

Small-scale expressions were utilised to empirically determine the optimal expression conditions for each recombinant protein, as outlined in Table 2.4. 10 ml of LB-ampicillin in a universal tube was inoculated with 100 µl (1 in 100) of the 10 ml starter culture and grown at 37 °C, whilst agitating at 180 rpm, until optical density at 600 nm (OD_{600}) reached 0.6. Protein overexpression was then induced via the addition of [200 µM] Isopropyl-β-D-Thiogalactopyranoside (IPTG) and incubated for differing time points, at differing temperatures to determine the optimal conditions for protein over-expression. Upon completion, cells were harvested by centrifugation at 3000 g for 10 minutes (Eppendorf 5810 centrifuge). The supernatant was discarded and the pellet was re-suspended in Tris-buffered saline (TBS)-0.1% Triton (Tr). Bacterial cell pellets were then sonicated (Soniprep 150, MSE) on ice for 3 minutes, with repeated cycles of 30 seconds at 10 microns, and 30 seconds cooling. The lysate was then clarified by centrifugation at 16000 g for 10 minutes (Biofuge pico, Hearaeus). Fractions were separated by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) (see *section 2.2.5*) and Coomassie stained, to reveal the over-expression conditions under which soluble protein yield was greatest.

Table 2.4 Optimal protein over-expression conditions, established by small-scale experiments.

Protein encoded	Induction temperature (°C)	Induction length (hours)
USP5	15	40
ZnF-UBP	20	16
hHR23a UBA2	20	16
t-UBD	20	16
Monoubiquitin	30	16

2.2.2 Large-scale bacterial expression

Large-scale cultures (1 L of LB) were completed in 2 litre baffled flasks. Each litre of growth media was inoculated with one 10 ml (1:100 dilution) starter culture and grown at 37 °C with agitation until the OD_{600} reached 0.6. Overexpression was induced via the addition of [200 μM] IPTG and grown according to the established optimal conditions (see Table 2.4). Upon completion cells were harvested by centrifugation at 3000 g for 20 minutes in a Sorvall RC3BP centrifuge (Thermo). The supernatant was discarded and the pellet for each litre of culture was re-suspended in 20 ml TBS-Tr and transferred to a Falcon tube before centrifuging at 3000 g for 20 minutes (Eppendorf 5810 centrifuge). The supernatant was discarded and the bacterial cell pellet stored at -80 °C until purification.

2.2.3 Bacterial cell lysis

The bacterial cell pellet resulting from one litre of culture of the protein of interest was thawed on ice, and supplemented with a 100 μL of DNAseI and 100 μL of protease inhibitor cocktail II. Typically pellets were then re-suspended in 30 ml of TBS-Tr (see Table 2.5). Cell resuspensions were lysed by sonication (Soniprep 150, MSE) on ice for 5 minutes, with repeated cycles of 30 seconds at 10 microns, and 30 seconds cooling. The lysate was then clarified by centrifugation in Avanti J-26 XP, using JA-25.50 rotor (Beckman Coulter), completing two spins at 35000 g for 30 minutes at 4 °C.

Alternately the bacterial cell pellets resulting from ubiquitin overexpression were re-suspended in 30 ml 5 x AE buffer (pH 4), to precipitate many of the co-expressed proteins. Lysis and clarification was completed as above, albeit between centrifugation steps a further 10 ml of 5 x AE buffer was added to the supernatant.

2.2.4 Protein purification strategies

Following lysis and clarification, proteins were purified utilising empirically determined strategies, summarised in Table 2.5. Typically fast protein liquid chromatography (FPLC) using an AKTA prime or AKTA start FPLC system (GE Healthcare), in conjunction with a variety of chromatographic media was utilised. All buffers were filtered using a 0.2 μM cellulose-nitrate filter (Whatman) and degassed under vacuum. All columns were pre-equilibrated in running buffer. Protein fractions were monitored by changes in the absorbance at 280 nm. Fractions shown to contain protein were analysed by SDS-PAGE (see *section 2.2.5*) to confirm purity.

Table 2.5 Purification strategies employed for proteins of interest

Protein encoded	Lysis buffer	Purification 1	Purification 2
USP5	TBS-Tr	GST-affinity	Gel filtration
ZnF-UBP	TBS-Tr	GST-affinity	Anionic exchange
hHR23a UBA2	TBS-Tr	GST-affinity	Anionic exchange
t-UBD	TBS-Tr	GST-affinity	Anionic exchange
Monoubiquitin	5 x AE buffer	Cationic exchange	Gel filtration

2.2.4.1 GST-affinity chromatography

The first stage of recombinant protein purification (with the exception of ubiquitin, see Table 2.5) exploited the affinity of the N-terminal GST-fusion tag for glutathione. Cell lysates containing GST-tagged proteins from 1 L of culture were incubated with 1 ml of glutathione-Sepharose 4B (GE Healthcare) for 1 hour at 4°C whilst rotating. Following the binding step, the Sepharose was transferred to a 15 ml gravity column

(Polypropylene, Qiagen) and washed with three column volumes of TBS-Tr, and then three column volumes of thrombin cleavage buffer (TCB) and incubated with thrombin (10 U) in 2 ml of TCB for 16 hours at 4°C to release the protein of interest with a Gly-Ser remnant at the N-terminus. The flow through was collected and the column washed in 3 x 1 ml TCB, all eluate was pooled to yield 5 ml, which was further purified.

2.2.4.2 Ion-exchange chromatography

Ion-exchange chromatography exploits the amphoteric nature of amino acids in solution, wherein at pH values above or below the proteins intrinsic isoelectric point, proteins retain a net charge, and thus can be separated from mixtures on oppositely charged resins.

All samples were diluted into 150 ml of water prior to loading onto either an anionic or cationic ion exchange column. Proteins were then eluted from the respective ion exchange columns by incrementally increasing the concentration of NaCl, utilising an automated gradient on the AKTA Start. Anionic exchange chromatography was completed following GST-affinity chromatography, using a HiTrap Q-HP column (GE Healthcare). Once the sample was loaded, the column was washed with 30 ml anionic loading buffer. Elution was completed by running a 2-20% gradient of anionic elution buffer (containing 2 M NaCl) over 300 ml. Cationic exchange chromatography was utilised to purify untagged ubiquitin directly following lysis of bacterial cell pellets, and was completed on a HiTrap SP-HP column (GE Healthcare). Once the sample was loaded, the column was washed with 30 ml cationic loading buffer. Elution was completed by running a 0-100% gradient of cationic elution buffer (containing 2 M NaCl) over 300 ml. Fractions containing ubiquitin were identified by SDS-PAGE (see 2.2.5) and subsequently lyophilised by snap freezing in liquid nitrogen and freeze drying under vacuum.

2.2.4.3 Gel filtration

Gel filtration separates proteins according to size. Following GST-affinity chromatography, USP5 (and mutagenic variants) was separated from further impurities using a HiLoad Superdex 200 column (GE Healthcare). Separation was completed in USP5 GF buffer. Ubiquitin, following ion-exchange chromatography, was re-suspended in 2 ml Milli-Q water and separated from remaining impurities on a HiLoad Superdex 75 column (GE Healthcare) column. Fractions containing ubiquitin where lyophilised prior to desalting (see section 2.2.5).

2.2.5 SDS-PAGE

SDS-PAGE was used to separate protein mixtures, approximate denatured protein mass and determine relative protein purity. SDS-PAGE was completed on 5-20% gradient gels, cast *in house*. The resolving gel was prepared according to Table 2.6 and poured using a gradient mixer, wherein 5 and 20% acrylamide solutions were prepared and mixed to produce a continuous gradient. After pouring, the gel was topped with water-saturated butanol to ensure a level surface. Once polymerised, the stacking gel was poured and a comb added to form wells. Gel application buffer (GAB) was added to protein samples, typically adding 20 µL GAB to 50 µL sample, which were heated to 100 °C for 5 minutes. The gel was electrophoresed at 40 mA (EC13590 powerpack, Thermo) for ~2 hours in electrode buffer. BLUeye Prestained protein markers (Geneflow) were loaded on the gel in order to approximate the denatured molecular weights of proteins within the samples. Gels were either incubated for 1 hour in Coomassie blue stain and destained using Coomassie blue destain solution, or proteins from the gel were transferred to nitrocellulose membrane for subsequent immunoblotting (see Section 2.3.4).

Table 2.6 SDS-PAGE gel solution preparation

Component	5% gel solution	20% gel solution	Stacking gel
Buffer A	-	3.33 ml	-
Buffer B	3.33 ml	-	-
Stacking buffer	-	-	5 ml
Water	5 ml	-	-
30 % (w/v) Acrylamide-Bisacrylamide (37.5:1)	1.67 ml	6.67 ml	1 ml
10% SDS (w/v)	100 µL	100 µL	-
Tetramethylethylenediamine (TEMED)	10 µL	10 µL	10 µL
10% Ammonium persulfate (w/v)	100 µL	100 µL	100 µL

2.2.6 Desalt / buffer exchange

Upon completion of protein purification, proteins were typically desalted. Lyophilised ubiquitin was re-suspended in water and desalted on a HiTrap desalting column (GE Healthcare) into water. Fractions containing ubiquitin were lyophilised for a final time. The remaining proteins were desalted by ultrafiltration using Vivaspin 20 columns (3/10 kDa MWCO) (GE Healthcare) into 25 mM ammonium acetate (pH 7) according to the manufacturer's instructions, snap frozen and stored at -20 °C.

2.2.7 Protein quantification

Prior to use, recombinant purified proteins were thawed on ice and quantified using a BCA assay (Pierce), with bovine serum albumin (BSA) utilised as a standard, according to the manufacturer's guidelines. Reactions were incubated in a 96-well plate for 30 minutes at 37 °C and analysed on a MultiSkan FC plate reader (Thermo). Alternately if the sample proved limiting, protein concentration was quantified using the NanoDrop ND-1000 spectrophotometer (Thermo Scientific).

2.3 Recombinant protein affinity chromatography

In addition to GST-affinity chromatography described in *section 2.2.4.1*, recombinant purified proteins were coupled to Cyanogen bromide-activated Sepharose (CNBr-Sepharose) 4B and utilised to enrich both commercial purified polyubiquitin and unanchored polyubiquitin from cell lysates. The associated methods are indicated in *section 2.3*.

2.3.1 Protein immobilisation

CNBr-Sepharose permits the covalent coupling of primary amine containing ligands to cynate ester/ imidocarbonate immobilised agarose 'beads'. Coupling was completed according to the manufacturer's guidelines. Briefly CNBr-Sepharose was hydrated for 15 minutes in excess 0.1 M HCl solution, and transferred to a 15 ml gravity column (polypropylene, Qiagen). CNBr-Sepharose was then washed with 5 bed volumes of coupling buffer. Proteins to be coupled, including ZnF-UBP, hHR23A UBA2, USP5 and t-UBD were diluted into coupling buffer and incubated with CNBr Sepharose for 3 hours at 4 °C on a rotator. CNBr-Sepharose was then washed and incubated for 16 hours at 4 °C on a rotator in 1M ethanolamine. Subsequently the beads were washed with 5 bed volumes of coupling buffer and then alternately (4 times) with 5 bed volumes of Tris buffer and acetate buffer. Finally beads were equilibrated in the assay running buffer.

Protein-coupled Sepharose was utilised to enrich either commercial polyubiquitin (Boston Biochem) or to purify endogenous unanchored polyubiquitin chains from HEK293T cell extracts (see *section 2.2.2*).

2.3.2 Unanchored polyubiquitin chain affinity enrichment

2.3.2.1 Mammalian cell culture

Mammalian cell culture was completed in a Walker class II microbiology safety cabinet. Reagents were stored at 4 °C and pre-warmed to 37 °C prior to use. Studies in this thesis utilised cultures of an adherent human embryonic kidney cell line containing the SV40 Large T-antigen (HEK293T). Cells were grown in Dulbecco's Modified Eagle Medium (DMEM), supplemented with 10% (v/v) foetal bovine serum and 1% (v/v) penicillin-streptomycin, in 75 cm^2 cell culture flasks (Corning) at 37 °C with 5 % CO_2, until ~80 % confluent (assessed by light microscope). Cells were then washed in Dulbecco's phosphate buffered saline (PBS) and treated with 2 ml of trypsin-EDTA to detach the cells. The trypsin-EDTA was then diluted using DMEM media, bringing the final volume to 8 ml. 2 ml of the mixture was returned to the flask (1:4 split) and made up to a final volume of 10 ml in DMEM and further cultured. The remainder was pelleted by centrifugation (Eppendorf 5810 centrifuge) at 1000 g for 2 minutes and stored at -20 °C prior to unanchored polyubiquitin chain purification.

2.3.2.2 Preparing HEK293T cell lysate for the purification of endogenous unanchored polyubiquitin chains

Typically cell pellets from four 75 cm^2 cell culture flasks (cultured as above) of HEK293T cells were re-suspended in 4 ml of homogenising buffer and lysed by sonication. The lysate was clarified by centrifugation (16000 g, 10 min, 4°C) and the supernatant passed through glass wool. The protein concentration of the clarified supernatant was determined (typically 2 mg of total protein/flask) using a Thermo Scientific Pierce BCA Protein Assay Kit (as per *section 2.2.7*), prior to quenching excess NEM (with DTT, [10 mM]). The lysate was heated at 75°C for 20 minutes and centrifuged (16000 g, 10 min, 4°C). The heated lysate was subsequently incubated with protein coupled Sepharose overnight. Following enrichment, Sepharose was then

washed in homogenising buffer and then Sepharose wash buffer prior to subsequent analysis by DUB assay, MS or immunoblotting.

2.3.3 Deubiquitylation assays

In vitro deubiquitylation assays were used to assess the activity of the catalytic null USP5 on commercial polyubiquitin chains, as well as to confirm the unanchored nature of purified endogenous unanchored polyubiquitin chains. The deubiquitylation assay was performed under conditions (1:4 enzyme: substrate molar ratio) previously employed by Komander and co-workers to uncover polyubiquitin linkage selectivity of different DUBs [71]. Briefly, purified USP5 (WT or Cys-335-Ala, denoted as C335A) was diluted to 0.2 mg/ml in 10X DUB buffer (500 mM Tris, 1.5 M NaCl 100 mM DTT (pH 7.5)) , to yield a 1X DUB buffer, and pre-incubated at room temperature for 10 minutes. 10 µL of the pre-incubated USP5, was then mixed with either 3 µg of Lys-63 linked tetraubiquitin (Boston Biochem, Cambridge, MA) or 50 µL of the respective (ZnF-UBP, USP5 or t-UBD coupled) Sepharose, and made up to 30 µL, following the addition of 3 µL of 10X DUB buffer, with water. The deubiquitylation of Lys-63 linked tetraubiquitin was monitored over a time course, accordingly 6 µL aliquots were removed at the indicated time points and quenched via the addition of 9 µL of GAB. 50 % of each sample was resolved on a 5-20% SDS-PAGE, and western blotted (see *section 2.3.4*). Alternately, in the case of deubiquitylation assays for Sepharose-enriched unanchored polyubiquitin, assays were incubated for 16 hours, and upon completion the reaction quenched via the addition of 20 µL of GAB. 100% of the sample was resolved on a 5-20% SDS-PAGE, and western blotted (see *section 2.3.4*).

2.3.4 Western blotting

Western blotting was completed to probe the nature of purified unanchored polyubiquitin chains. Samples were first separated using SDS-PAGE according to *section 2.2.5*. Gels were then soaked in western transfer buffer and a gel/nitrocellulose

membrane (Amersham Protran, 0.45 μM, GE Healthcare) sandwich created, surrounded by layers of filter paper (Whatman). The transfer was completed by blotting in western transfer buffer overnight at room temperature (Fisherbrand Electrophoresis Unit) at 40 mA. The successful transfer of proteins to the nitrocellulose membrane was confirmed by Ponceau-S staining, which was subsequently washed off by three 5 minute washes with TBS-Tween (Tw) (0.1%) (and used in all subsequent wash steps).

Membranes were covered in blocking buffer and incubated for 1 hour at room temperature. Membranes were then probed with a range of primary antibodies (re-suspended in blocking buffer), washed for 3 x 5 minutes and incubated with a species specific horseradish peroxidase (HRP)-conjugated secondary antibody, before a final 3 x 5 minute washing step. Antibody dilutions and conditions are listed in Table 2.7. Blot detection was completed using the Western lighting Plus-ECL enhanced Chemiluminescence (ECL) system (Perkin Elmer), according to manufacturer's guidelines. Membranes were then exposed to imaging film (Kodak) for durations varying from 1 minute to up to 20 minutes (if required). The film was then manually developed via immersion in developing solution (PQ universal, Ilford), followed by subsequent washing in water and immersion in a fixer solution (2000rt, Ilford). The film was then washed and dried.

Table 2.7 Antibody conditions utilised during Western blotting.

Antibody	Dilution	Conditions
α-VU1 (Ubiquitin) (Life Sciences)	1:1000	Incubate overnight at 4°C
α-Apu2 (Lys-48 linkage specific) (Millipore)	1:1000	
α-HWA4C4 (Lys-63 linkage specific) (Millipore)	1:1000	
Rabbit HRP α-Mouse secondary (Dako)	1:2500	Incubate for 1 hour at room temperature
Mouse HRP α-Rabbit secondary (Dako)	1:2500	

2.4 Mass spectrometry

MS based approaches are employed extensively in *chapters 3-5*, accordingly in *section 2.4* we present these approaches.

2.4.1 Bottom-up proteomics

'Bottom-up' proteomics, in which the specificity of trypsin is exploited to generate peptides from complex mixtures of proteins, was employed to assess the polyubiquitin linkage profile of samples enriched on protein-coupled Sepharose.

2.4.1.1 Sample preparation

HEK-derived purified unanchored (poly)ubiquitin enriched on ZnF-UBP, t-UBD or C335A-USP5 Sepharose, were eluted into 0.1% formic acid (pH 2), and digested according to a modified FASP strategy [227]. Briefly, eluate was buffer exchanged (Amicon ultra, 0.5 ml, 10 kDa MWCO) into digestion buffer (2 M urea, 0.1 M ammonium bicarbonate (pH 8), 5% (v/v) acetonitrile), reduced (50 mM DTT), alkylated (100 mM chloroacetamide, Sigma, Dorset, UK) and digested with trypsin overnight at 37°C (sequencing grade modified porcine trypsin, Promega, Southampton, UK). After overnight digestion, peptides were extracted (70% acetonitrile, 0.1% formic acid) and vacuum centrifuged to dryness prior to MS analysis.

2.4.1.2 LC-MS/MS analysis

LC-MS/MS was first completed *in house* using an Ultimate 3000 nano-LC system (Dionex) and an LTQ FT Ultra mass spectrometer (Thermo Scientific). The injected samples (1-3 μL) were concentrated on a commercially available C_4 reverse-phase trap column (PepMap 300, Dionex) and then loaded onto a custom packed C_4 reverse-phase analytical column (5μm, 100Å, Phenomenex) with a Picofrit tip (75 μM i.d. x 150 mm, New Objective). Peptides were eluted from the analytical column using a linear gradient

of mobile phase B (acetonitrile:water, (95:5, v/v) with 0.1% formic acid) in mobile phase A (water:acetonitrile,(95:5, v/v) with 0.1% formic acid), ranging from 0% to 70% in 30 minutes, at a flow rate of 0.3 µL/min, followed by 90% for 5 minutes and then column equilibration for 20 minutes. Instrument control was performed via the Chromeleon 6.8 data system.

The output from the analytical column was sprayed directly into the nanospray ion source of the LTQ FT Ultra mass spectrometer, by applying a 1.7 kV voltage to the coated PicoTip emitter (New Objective). Other parameters included; capillary temperature, 275 °C, inner capillary voltage, 37 V, and tube lens voltage, 145 V. The LTQ FT Ultra mass spectrometer was set to acquire one scan event at 100,000 resolution over the *m/z* range 400-2000 Da in positive ion mode. MS/MS was conducted in data-dependent (DDA) mode, with the 5 most intense ions for each scan isolated within a window of 8 Th, and subjected to CID using a normalized collision energy of 30.0. Dynamic exclusion was enabled, and signals with +1 charge state were rejected. The .raw data file obtained from each LC-MS/MS acquisition was processed using Bioworks software (Thermo Scientific), searching each file in turn against a custom database including the ubiquitin sequence. Variable peptide modification with Gly-Gly (+114.042927) and Leu-Arg-Gly-Gly (+383.228103) on Lys residues was set. This permits direct identification of ubiquitin signature peptides.

To validate *in house* findings, analysis was also completed in collaboration with Dr Sharad Mistry and Dr Andrew Botrill (who analysed prepared samples) at the University of Leicester. LC-MS/MS was carried out using an RSLC nano-HPLC system (Dionex) and an LTQ OT Velos mass spectrometer (Thermo Scientific). Samples were loaded onto a C_{18} reverse-phase trap column (Dionex). Peptides were then eluted from the trap column at a flow rate of 0.3 µl/min, through a reverse-phase PicoFrit capillary column (75µm i.d. x 400mm) containing Symmetry C18 100 Å media (Waters, UK) that was packed *in house* using a high pressure device (Proxeon Biosystems, Denmark).

Peptides were eluted over a period of two hours, with the output of the column sprayed directly into the nanospray ion source of the LTQ OT Velos mass spectrometer. The LTQ OT Velos mass spectrometer was set to acquire a one scan event at 60,000 resolution over the m/z range 300-1800 Da in positive ion mode. Up to 10 data dependent HCD MS/MS were triggered from a scan event. The isolation width was 2.0 Da, normalised collision energy 42.5. Dynamic exclusion was enabled. The .raw data file obtained from each LC-MS/MS acquisition was processed using Proteome Discoverer (version 1.4.0.288, Thermo Scientific), searching each file in turn using MASCOT (version 2.2.04, Matrix Science Ltd.) [228] against a custom UniprotTB_20140416 database (unknown version, 3997 entries) [229], to which the human ubiquitin sequence was added. Variable modifications set as Gly-Gly (+114.042927) and Leu-Arg-Gly-Gly (+383.228103) on Lys residues. The output from Proteome Discoverer was further processed using Scaffold Q+S [230] (version 4.4.1.1, Proteome Software).

2.4.2 Carbene footprinting

Diazirine-based protein carbene footprinting is utilised in *Chapter 4* to probe the USP5-Lys-48 diubiquitin interaction surface. In this MS-based approach the labelling of a protein, in this case C335A-USP5, with an *in house* synthesised aryltrifluromethylcarbene (Aryldiazirine 1) is monitored, in the presence and absence of ligand, Lys-48 diubiquitin, to permit the high resolution mapping of interaction sites. The workflow for this strategy is illustrated in Figure 2.2.

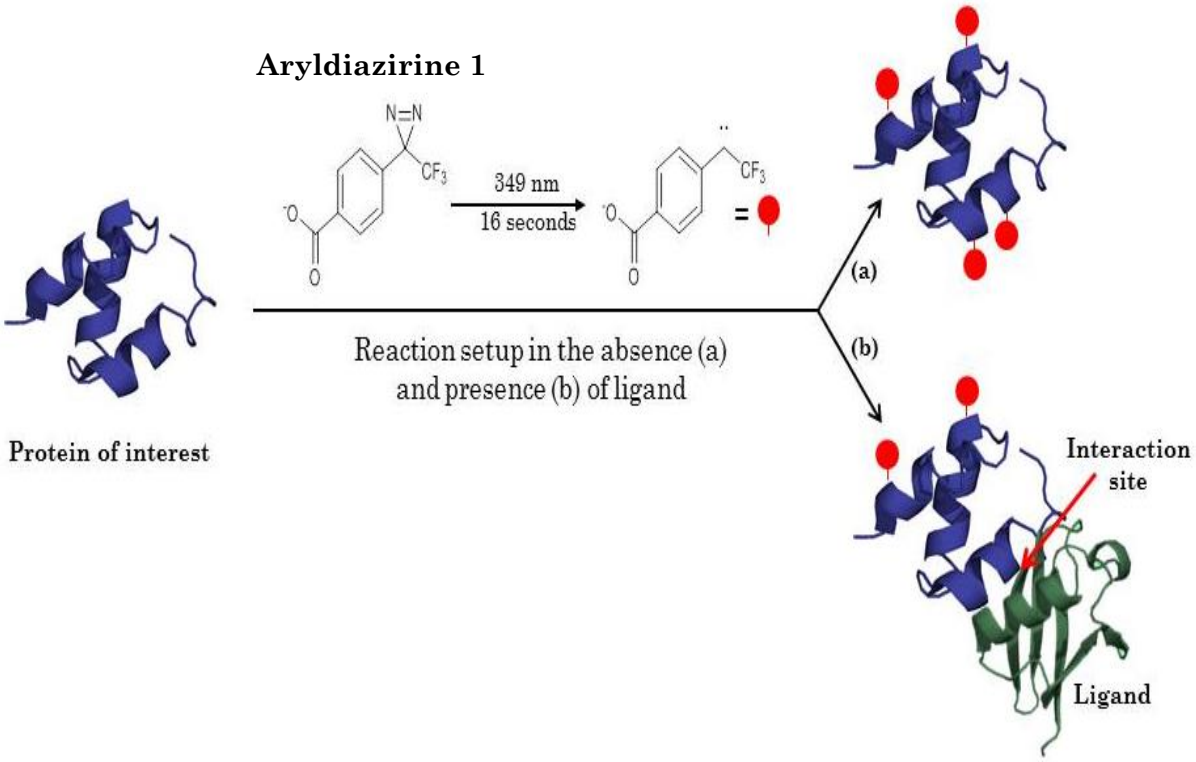


Figure 2.2 Diazirine-based protein carbene footprinting workflow. The protein of interest, in this case C335A-USP5, is incubated with Aryldiazirine 1, snap frozen in liquid nitrogen and irradiated at 349 nm for 16 seconds. This results in the photo-activation of the diazirine to yield a highly reactive carbene, which inserts between C-H bonds to irreversibly label the protein surface (label indicated by red circle). Owing to the structure of Aryldiazirine 1 this reagent favours labelling of basic or hydrophobic residues. The resulting mixture is separated by SDS-PAGE, stained in Coomassie and the bands corresponding to the labelled protein excised and digested. LC-MS/MS analysis is then completed to identify labelled peptides, which have clear mass shifts of 202.02 (± 0.02) Da. When completed in a differential manner, comparing the labelling pattern in the presence and absence of ligand, in this case Lys-48 diubiquitin, it is possible to deduce protein-ligand interaction sites using peptide MS or MS/MS.

2.4.2.1 Sample preparation

The carbene footprint of the C335A-USP5: Lys-48 diubiquitin complex was determined experimentally (footprinting analysis completed by Lucio Manzi). C335A-USP5, purified as indicated in *section 2.2.4*, was prepared to a final concentration of 30 μM, mixed with *in house* diazirine (in 20 mM Tris (pH 7.5), 150mM NaCl, to a final concentration of 10 mM) indicated in Figure 2.2, in the presence and absence of Lys-48 linked diubiquitin (30 μM in 20 mM Tris (pH 7.5), 150 mM NaCl). The mixture was

left to equilibrate for five minutes at room temperature. Aliquots (6 μL) of this mixture were then placed in 11 mm PP vial crimp/snap 250 μL sample vials (Thermo Fisher Scientific) and snap-frozen in liquid nitrogen.

The labeling reaction was then initiated by irradiation (16 seconds) of the mixture using a Spectra Physics Explorer 349 laser (an actively Q-switched Nd:YLF laser operating at 349 nm, with a repetition frequency of 1000 Hz, and a pulse energy of 125 μJ, Newport). The laser beam was directed into the open top of the sample vial using a small 45° mirror. The irradiated protein samples were subsequently separated by SDS-PAGE, on a 15 % acrylamide gel, and stained in Coomassie blue. Following destaining, bands corresponding to C335A-USP5 were excised, reduced (DTT, 10 mM in 100 mM ammonium bicarbonate), alkylated (iodoacetamide, 55 mM in 10 mM ammonium bicarbonate) and digested with trypsin (1:15 protease/protein ratio in 10 mM ammonium bicarbonate) for 16 hours at 37 °C. Upon completion, 0.5 μL of formic acid was added to inactivate the trypsin. The resulting solutions were analyzed by LC-MS/MS without further dilution.

2.4.2.2 LC-MS/MS analysis of the C335A-USP5:Lys-48 diubiquitin carbene footprint

Analysis of trypsin digested C335A-USP5 (in the presence and absence of Lys-48 linked diubiquitin, 1 μL) was completed out using an Ultimate 3000 nano-LC system (Dionex) and an LTQ FT Ultra mass spectrometer (Thermo Scientific) as in *section 4.2.1.2*, accordingly LC and MS parameters were the same. Identification of USP5 peptides was conducted using an automated DDA mode, in which the four most intense ions for each scan were isolated within a window of 8 Th and subjected to CID using a nominal energy of 35.0. The raw data was searched against a custom database including the C335A-USP5 sequence using Bioworks software (Thermo Fisher Scientific), and then manually examined.

2.4.2.3 Carbene footprinting data analysis

The proportion of labeling at the peptide level was determined by extracting the chromatograms for each labeled and unlabeled peptide ion. The fractional modification per peptide (P, see Equation 2.1), was then calculated using the approach described by Jumper *et al* [231], wherein the peak areas corresponding to labelled and unlabeled peptides are used to solve Equation 2.1. The values for each peptide derived from C335A-USP5 in the presence and absence of equimolar Lys-48 linked diubiquitin were compared (three biological replicas of each), to reveal the differential footprint which results from ligand binding. Significance in the difference of labelling, in the presence and absence of Lys-48 linked diubiquitin was determined by a student t-test.

Equation 2.1: $$P = \frac{A_{labelled}}{A_{labelled} + A_{unlabelled}}$$

2.4.3 Native mass spectrometry

Native mass spectrometry was employed extensively in *chapter 4 and 5*, to probe protein structure and interactions.

2.4.3.1 Sample preparation

To probe native protein structure and interactions by MS, non-denaturing conditions are required. Accordingly prior to analysis proteins are desalted, as discussed in *section 2.2.6*, into an aqueous solution of ammonium acetate (pH 7). Alternately to confirm protein construct integrity and mass, samples were denatured into 80 % MeCN/0.1 % formic acid using C_4 ZipTips™ according to the manufacturer's guidelines. This approach offers increased signal intensity and reduced peak adduction, ideal for accurate mass measurement. It is noteworthy that with larger protein constructs there can be up to a ~1% mass error, associated with ESI-MS mass measurement.

2.4.3.2 Synapt HDMS operating conditions

Native MS experiments were completed on the Waters Synapt HDMS, a hybrid quadrupole-ion mobility-orthogonal acceleration TOF instrument, the components of which are illustrated in Figure 2.3. The Synapt HDMS has an interchangeable source region, permitting both ESI and nESI analysis. The source temperature was maintained at 50 °C. For ESI, samples were injected using Hamilton syringes (Bonaduz) and a Harvard syringe pump (Holliston), which transferred the sample at typically a rate of 5 μL/minute to the capillary of the ESI probe. Ionisation was aided by a drying gas (200 L/h^{-1}) and sample cone gas (30 L/h^{-1}). For static nESI, capillaries were prepared from borosilicate glass capillaries (OD = 1 mm, ID = 0.78 mm, length = 10 cm) (Warner) using a Brown micropipette puller (Sutter instruments). For conductivity purposes, the tips were coated in silver using a vacuum evaporator built *in house*. Capillaries were loaded with 1-10 μL of sample and inserted into the standard waters nESI platform, which permits adjustment of capillary tips in the x,y and z axis. The back of the tip was connected to a low flow rate nitrogen supply (0.1-0.5 L/h^{-1}) to push the sample to the end of the tip. No nebuliser gas is required with nESI, which is driven completely by Coulombic repulsion.

The Synapt HDMS was operated in positive-ion mode and the TOF analyser operated in V-mode. Instrument parameters were optimised for each experiment to provide the highest relative signals. For native USP5 studies nESI was performed, whilst for native t-UBD investigations ESI was performed, the typical operating conditions are noted in Table 2.8. Ion activation conditions, including the sample cone voltage, the injection voltage into the trap cell, and the injection voltage into the IMS cell, were minimised to near the threshold for ion transmission, to avoid gas-phase unfolding and preserve non-covalent interactions. Collisional cooling was applied by partially closing a Speedi-valve attached to the source pumping line until a backing pressure of ~4-4.5 mBar was obtained.

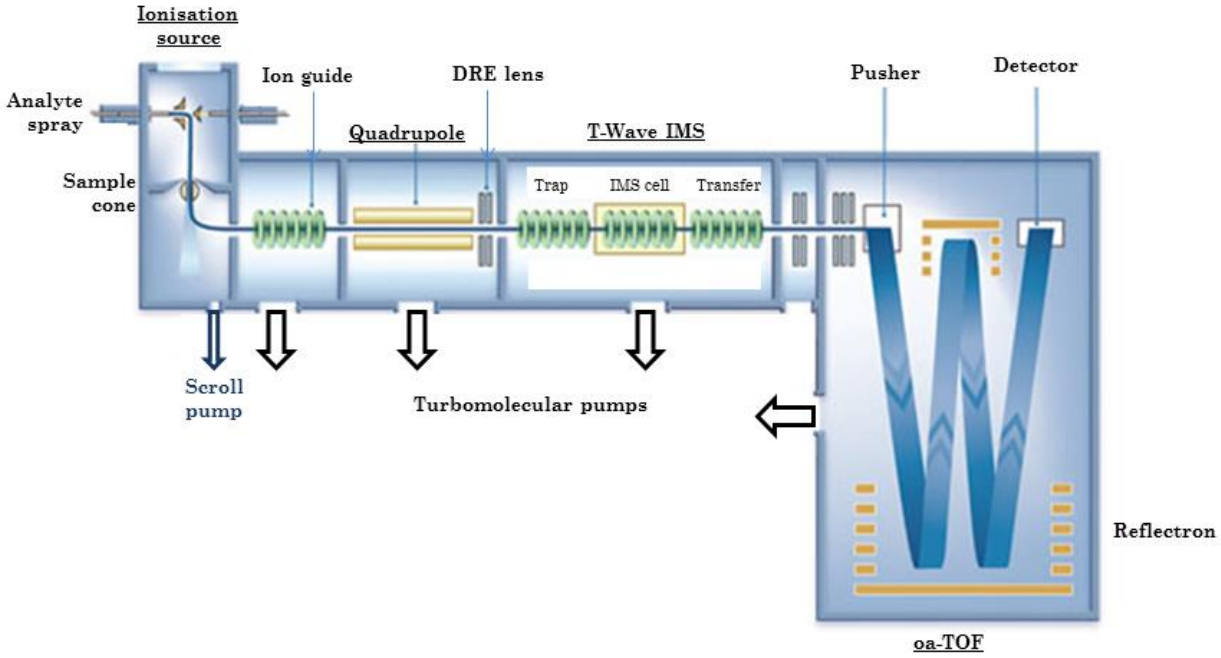


Figure 2.3 Schematic of the Waters Synapt HDMS, used throughout investigations in this thesis for native MS studies, as well as to confirm the mass of protein constructs following purification.

Table 2.8 Typical Synapt HDMS operating conditions for native MS investigations

	USP5 (nESI)	**t-UBD (ESI)**
	Voltages	
Capillary voltage (kV)	1.3	2.5
Sample cone voltage (V)	30	20
Extraction cone voltage (V)	5	5
Trap collision voltage (V)	10	8
Transfer collision voltage (V)	10	5
	Pressures	
Backing pressure (mbar)	~4.5	~4
Trap pressure (mbar)	2.5×10^{-2}	2.5×10^{-2}
IM pressure (mbar)	TOF mode - 1.6×10^{-5} IM mode - 4.5×10^{-1}	1.6×10^{-5}
TOF pressure (mbar)	1.8×10^{-6}	1.8×10^{-6}
	Scanning range	
TOF *m/z* range	1000 – 8000	500 - 4000

2.4.3.4 MS based K_d determination

Native ESI-MS is utilised in *chapters 4 and 5* to determine the apparent approximate binding affinity (K_d) of UBD-ubiquitin interactions, in an approach previously established by Sokratous *et al* [198]. Titration style experiments are completed, in which protein and ligand are prepared as 1:1 mixtures at a range of concentrations and analysed by ESI-MS. An example of this workflow is illustrated in Figure 2.4. Following acquisition, minimum smoothing and background subtraction is applied to the obtained spectra, prior to extracting the signal intensities corresponding to free protein and bound protein-ligand complex into Excel. The resulting ratio of unbound protein and bound protein-ligand complex can then be used to derive K_d, as outlined below.

The equilibrium expression for the dissociation of a protein complex is given by Equation 2.2, where $[P]_{eq}$, $[L]_{eq}$ and $[P{\cdot}L]_{eq}$ are the equilibrium concentration of protein, ligand and their complex, respectively.

Equation 2.2:
$$K_D = \frac{[P]_{eq}[L]_{eq}}{[P.L]_{eq}}$$

The free and the bound protein ions as determined by MS are $(P)^{n+}$ and $(P{\cdot}L)^{n+}$, respectively. Providing that the electrospray response factors for the bound and unbound protein ions are similar and that minimal dissociation of the complex occurs post-desolvation, it can be assumed that the ratio of the ion intensity (I) of the bound and unbound protein ions represent solution equilibria. Accordingly we can re-arrange Equation 2.2 (as detailed in [198]) to Equation 2.4, to determine the K_d from the ratio (R) (Equation 2.3) and the initial concentrations of protein and ligand.

Equation 2.3:
$$R = \frac{\Sigma I[P_{unbound}]}{\Sigma I[P_{bound}]}$$

Equation 2.4:

$$K_D = R\left([L]_0 - \frac{[P]_0}{1+R}\right)$$

It is noteworthy that in cases where the K_d is in the nanomolar range, sample concentration required for detection by MS (i.e µM), exceeds the K_d concentration, thus values cannot be determined.

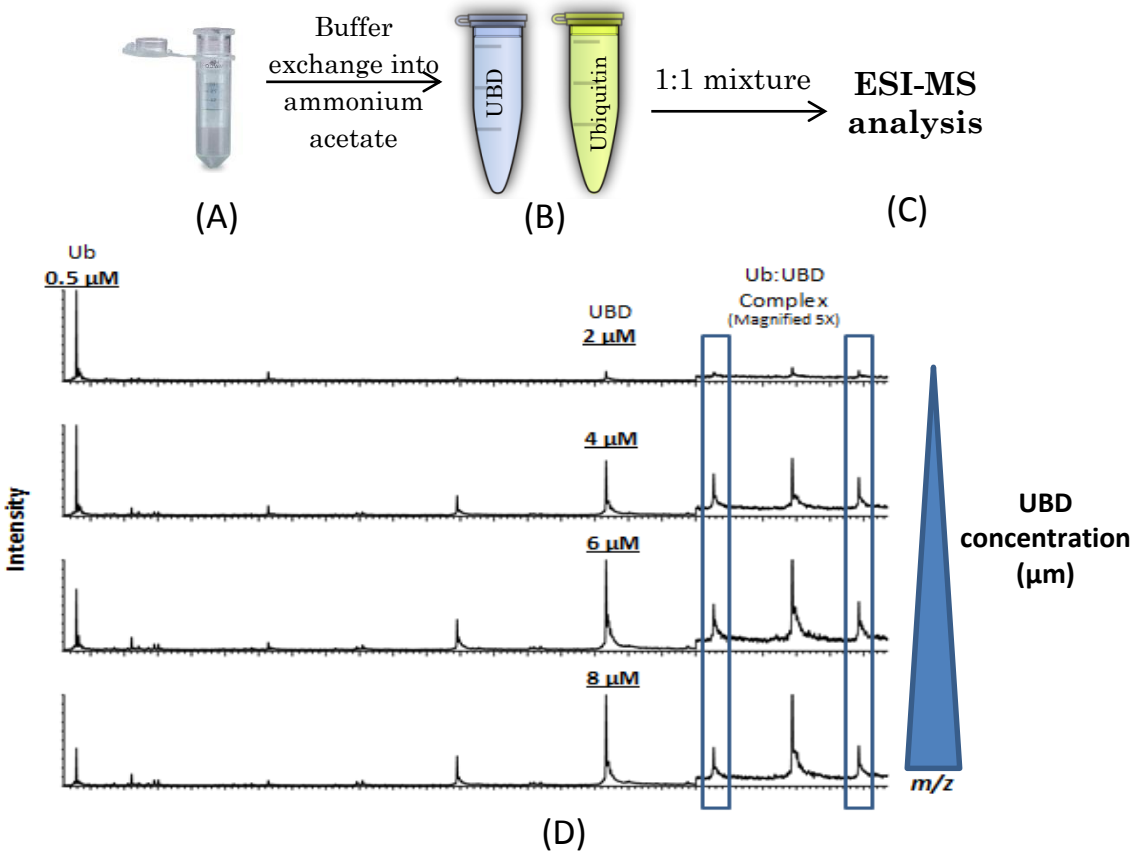


Figure 2.4 Workflow for K_d determination by ESI-MS. (A) Protein and ligand is desalted into a solution of (typically 25-200 mM) ammonium acetate. (B) Concentrations of protein and ligand are determined by absorption at 280 nm (using the NanoDrop ND-1000 spectrophotometer (Thermo Scientific)) and mixed in a 1:1 ratio, (C) prior to direct infusion into the Synapt HDMS. (D) A titration analysis is completed in which, in this example, the ubiquitin concentration is maintained at 0.5 µM, and the UBD concentration is varied (i.e. 2, 4, 6 and 8 µM). Three repeats at each titre value are completed. The intensity of peaks corresponding to unbound ubiquitin and bound complex are summed and utilised to determine binding affinity. Note that as UBD concentration increases free ubiquitin signals decreases and bound complex increases.

The Synapt HDMS has a built-in TWIMS device (see Figure 2.3), a series of stacked ring electrodes, which is sandwiched between the trap and transfer regions of the Synapt HDMS (see Figure 2.3). When operating the instrument in mobility mode, as opposed to the routinely used TOF mode, the trap acts to store ions before injecting ion packets into the mobility region. Subsequently the drift time, or time taken from the point of ejection at the trap, to the point of detection after TOF separation can be determined, alongside information on *m/z* and intensity. The ion drift time can then be converted to CCS, permitting comparison of gas-phase conformation to high resolution structures, determined by other biophysical techniques.

As alluded to in *chapter 1*, due to the complex nature of the electric field employed by the TWIMS device, calibration of the instrument must first be performed prior to CCS determination. This calibration strategy has previously been described [167, 232], and was adhered to in the calibration of the TWIMs device throughout investigations in thesis. The calibrants used were; β-lactoglobulin, BSA, and alcohol dehydrogenase (ADH) [232]. Measurement of analytes and calibrants was critically performed under identical instrument parameters. The IM (or drift) cell filled with nitrogen gas, was maintained at 0.45 mbar. Transmission of the ions through the IM cell was accomplished using a series of 10 V travelling potentials, travelling across the cell at a speed of 300 m/s^{-1}, along with RF radial confinement. Travelling wave parameters for the 'trap' and 'transfer' regions were: Trap 300 m/s, 0.5 V and transfer 248 m/s, 6 V. Other instrument parameters were as described in Table 2.8. Mobility data was viewed using drift scope, within the MassLynx software, which presents *m/z* plotted against drift time and intensity (see *chapter 4*). The data is exported to MassLynx, from which the drift time for each ion (t_D) can be extracted. To construct a calibration plot, the drift time of the calibrants must first be corrected by solving

Equation 2.5, to account for the determine drift time through the mobility cell only (and not the transfer/ TOF region), yielding (t_D').

Equation 2.5: $$t_D' = t_D - \left[\frac{L_{trans}}{v_w}\right] - \left[\frac{c\sqrt{m/z}}{1000}\right]$$

Where L_{trans} is the length of the transfer, v_w is the transfer wave velocity and C is the EDC (enhanced duty cycle) delay coefficient. Collisional cross sections (Ω) obtained from the Bush database for each of the calibrants must also be corrected, using Equation 2.6, in terms of ion charge (z) and reduced mass (μ) to yield a corrected cross section (Ω').

Equation 2.6: $$\Omega' = \frac{\Omega}{\left[z\sqrt{\frac{1}{\mu}}\right]}$$

Where the reduced mass (μ) is defined as follows in Equation 2.7.

Equation 2.7: $$\frac{1}{\mu} = \left[\frac{1}{M_r} + \frac{1}{M_g}\right]$$

Where M_r is the molecular weight of the calibrant (or analyte) and M_g is the molecular weight of the IM buffer gas (N_2, 28). Using Excel, an initial calibration graph is then plotted to determine the exponential correction factor (**m**) between the calibrant Ω' and t_D', by plotting ln t_D' against ln Ω' to yield a linear relationship. Equation 2.8 is then solved to yield td'', which is plotted directly against CCS to provide a mobility conversion plot, from which experimental CCS values of protein ions can be determined.

Equation 2.8: $$t_D'' = t_D'^{\mathbf{m}} . z . \sqrt{\frac{1}{\mu}}$$

2.5 Molecular dynamics

MD simulations utilise atomic co-ordinates, derived from high resolution, typically X-ray or NMR protein structures, to probe the accessible conformational landscape of a protein, within designated parameters.

2.5.1 Determining theoretical CCS

The output of MD simulations can be utilised to determine theoretical CCS, which in turn can be compared to experimental CCS values derived from IM-MS. MD simulations to determine theoretical CCS of USP5 were completed by Dr Neil Oldham. The X-ray structure of USP5 (PDB entry 3IHP) was used as a starting point for this modelling. The 151 missing residues from the USP5 structure were first built-in using iTASSER [233] and hydrogens added with MolProbity [234]. Protein molecular modelling was then performed using the Amber99SB force field [235] on the European eNMR grid [236]. The protein was modelled using its net charge state under physiological conditions. Prior to MD simulations, the structure was relaxed by 20,000 steps of energy minimisation in the gas phase at 300 K (Amber 11 Sander parameters: imin=1 (energy minimisation mode), maxcyc=20000 (number of steps), ncyc= 500 (number of steepest descent algorithm steps), ntb=0 (period boundary off), cut=12.0 (non-bonded cut-off in Å), igb=0 (no generalised Born solvent, i.e. gas-phase); all other parameters Sander defaults). This removed the vast majority of clashes from the original iTASSER structure (from an unacceptable clash score of 131/1000 atoms to an acceptable 0.7/1000 atoms, as assessed by MolProbity). A 2 ns gas-phase MD simulation was then run on the energy minimised structure at a constant temperature of 300 K (Amber 11 Sander parameters: imin=0 (minimisation off, i.e. MD run), ntb=0 (period boundary off), cut=12.0 (non-bonded cut-off in Å), igb=0 (no generalised Born solvent, i.e. gas-phase), ntt=3 (Langevin thermostat), gamma ln=1 (collision frequency), temp0=300 (temperature in K), tempi=300 (initial temperature in K),

nstlim=2000000 (number of 1 fs steps), ntpr=100 (rate of information written to the output file in steps), ntwx=2000 (rate of coordinates written to the output file in steps, i.e. 1000 sets of coordinates over the 2 ns simulation), ig=73277 (random seed); all other parameters Sander defaults). The 1000 output trajectories from this MD simulation were visualised in VMD, and every 5th structure exported to a PDB file (to keep the file size manageable). To estimate the theoretical CCS of the more extended conformation of USP5, corresponding to the higher charge states, the energy minimised USP5 structure (as above) was manually extended in PYMOL to mimic that seen in the SAXS analysis [237]. The resulting structure was then relaxed and submitted to gas-phase MD (2 ns) as described above.

The outputs of MD simulations were utilised to determine theoretical CCS of USP5 by using the PA method, CCS of each structure was calculated using the CCS calculator (DriftScope 2.1 CCS Calc program, Waters). The PA method of calculating CCSs, which has the advantage of computational speed, is known to underestimate the size of large biomolecules by a constant factor. To convert PA CCS values to 'Trajectory Method-like' values they were multiplied by the empirical scaling factor, 1.14 [39].

2.5.2 Designing the t-UBD

In *chapter 5* a synthetic ubiquitin-binding protein is engineered to afford the selective purification of unanchored Lys-48 linked polyubiquitin. This involved generating a tandem UBD containing construct, including the ZnF-UBP domain of USP16 and the UBA2 of hHR23A. MD simulations, completed by Dr Thomas Garner, were utilised to generate an energetically plausible model of this tandem UBD containing construct, and also to determine the linker length required between the respective domains to ensure no steric hindrances. The initial model of the ZnF-UBP domain in complex with monoubiquitin was generated using MOLMOL [238] by overlaying the NMR structure of the ZnF-UBP domain of USP16 (PDB 2I50) [239] with the related ZnF-UBP domain

of USP5 in complex with monoubiquitin (PDB 2G45) [84]. The resulting model of the complex was then overlaid with the proximal ubiquitin of Lys48-linked diubiquitin in complex with the UBA2 domain of hHR23A (PDB 1ZO6) [70] to generate a ternary complex model of the two UBDs in complex with Lys48-linked diubiquitin. The ternary complex model was optimised thermodynamically *via* energy minimization using an AMBER10 force field in MOE (2014.10 Chemical Computing Group Inc., Montreal, Canada). The subsequent model was analysed to assess the optimal linkage strategy for the two UBDs based on the distance between the N and C-terminal residues of the ZnF-UBP and UBA domain respectively, and the average distance spanned by disordered loops found in the PDB database. Linkers of 5-7 residues were modelled using the general loop building procedure of the MOE Homology Modeller and minimized in MOE [240]. A 5-residue linker caused significant steric clashes while the 6 and 7 residue linkers were well tolerated.

Chapter 3 – Purification of endogenous unanchored polyubiquitin chains

In recent years the functional significance of endogenous unanchored (substrate-free) polyubiquitin chains has emerged [11]. However, until recently, a direct and detailed molecular description of the unanchored polyubiquitin pool had remained elusive.

Previous investigations have sought to exploit the intrinsic specificity of the DUB, USP5, which selectively disassembles unanchored polyubiquitin [87], to indirectly infer the substrate-free nature of unanchored polyubiquitin chains, and there specific biological function. Accordingly, USP5 [204], or homologues in numerous species, including yeast [241], *Drosophila* [242], *Dictyostelium* [243] and plants [244] has been subject to genetic manipulation (deletion/knock-down). This approach has proven critical in demonstrating both the widespread abundance, and the biological significance of unanchored polyubiquitin chains, with USP5 knockdown characteristically resulting in the impairment of ubiquitin-dependent proteolysis.

Similarly, the USP5-null phenotype can be reproduced by ectopic expression of a ubiquitin C-terminal mutant, wherein Gly-75 and 76 are replaced with Ala [204] or Val [200] (studied in yeast/mammalian cell models) to 'cap' the chain, further supporting the notion that unanchored polyubiquitin chains are physiologically relevant. Finally as well as USP5, additional regulators of the unanchored polyubiquitin pool have emerged and have accordingly been the subject of manipulation, these include yeast ubiquitylation machinery such as the DUB Rfu1 [199], as well as the E_3-ligases UFD4 and HUL5 [200].

One of the first strategies towards the direct purification of unanchored polyubiquitin chains was developed to investigate the role of RIG-I and unanchored polyubiquitin in innate immune signalling [219]. RIG-I is responsible for the detection of invading viral RNA, and subsequently activating of NF-κB and IRF3 signalling cascades leading to an anti-viral response. Critically investigations by Zeng et al [219], revealed that RIG-I binds specifically to unanchored Lys-63 linked chains to mediate the activation of this pathway.

In order to demonstrate the necessity for unanchored polyubiquitin with respect to RIG-I activation, Zeng *et al* [219] developed a strategy to enrich RIG-I interacting partners. Briefly mammalian cells were first lysed in a buffer supplemented with NEM to alkylate and inhibit cysteine proteases, including DUBs, thus preventing the disassembly of unanchored polyubiquitin chains. The lysate was subsequently enriched on GST-RIG-I and then heated, resulting in the denaturation and precipitation of bait protein and the purification of heat stable (poly)ubiquitin, i.e. selective purification of unanchored polyubiquitin. To validate the unanchored nature of the purified polyubiquitin, a USP5 deubiquitylation assay was completed. Furthermore, the presence of a free ubiquitin C-terminus was confirmed by assessing the ability of the purified polyubiquitin to form a thioester intermediate with E_1 enzymes *in vitro*. The unanchored polyubiquitin chains enriched on GST-RIG-I were also subsequently determined to be Lys-63 linked, by completing a deubiquitylation assay with the Lys-63 chain selective DUB, CYLD [245], by blotting with linkage selective antibodies and finally by utilising the mammalian ubiquitin replacement strategy, discussed in *section 1.2.3*. The GST-RIG-I purified unanchored polyubiquitin could subsequently activate RIG-I signalling *in vitro*, highlighting the role of unanchored polyubiquitin chains in the innate immune response [219].

Consequently, prior to the start of the investigations in this thesis, the Layfield group sought to establish a means to extend and refine this approach to permit the selective

purification of unanchored polyubiquitin chains. As discussed in *section 1.2.4*, UBDs recognise distinct surface patches on ubiquitin to facilitate recognition and signal transduction, and can be utilised as tools to enrich the ubiquitin-modified proteome. The ZnF-UBP family of UBDs, found within numerous proteins including; selected DUBs (USP3, USP5, USP13, USP44, USP45, USP49, and USP51), the E_3-ligase UBP8, and HDAC6 [84], each share a conserved mode of interaction with ubiquitin, binding to the extreme C-terminal residues in their unconjugated form.

Strachan *et al* [107] reasoned that by utilising a ZnF-UBP domain family member in an affinity chromatography approach, the purification of unanchored polyubiquitin would be favoured over substrate-conjugated ubiquitin [107]. Of the ZnF-UBP domains characterised, that of USP5 is reported to have the highest affinity for monoubiquitin (K_d ~2.8 μM), accordingly this domain was utilised, and purified recombinant protein was coupled to Sepharose 'beads', as highlighted in Figure 3.1. This free-ubiquitin binding entity (FUBE) presented, for the first time, a single-step affinity purification strategy, which permits the selective capture of endogenous unanchored polyubiquitin. Initial insights by Strachan *et al* [107] revealed the purification of unanchored polyubiquitin chains from rat skeletal muscle tissue, with chains of >10 moieties purified. Limited MS/MS analyses confirmed the presence of ubiquitin signature peptides corresponding to Lys-11 and 48 linked polyubiquitin.

In this chapter we present the FUBE enrichment strategy, which is the reference for the remainder of the investigations in this thesis. We seek to further characterise the polyubiquitin linkages present in the unanchored polyubiquitin pool as purified from, for the first time, HEK293T mammalian cell extracts. Subsequently we then explore a further iteration of the FUBE to permit the affinity enrichment of polyubiquitin over abundant monoubiquitin.

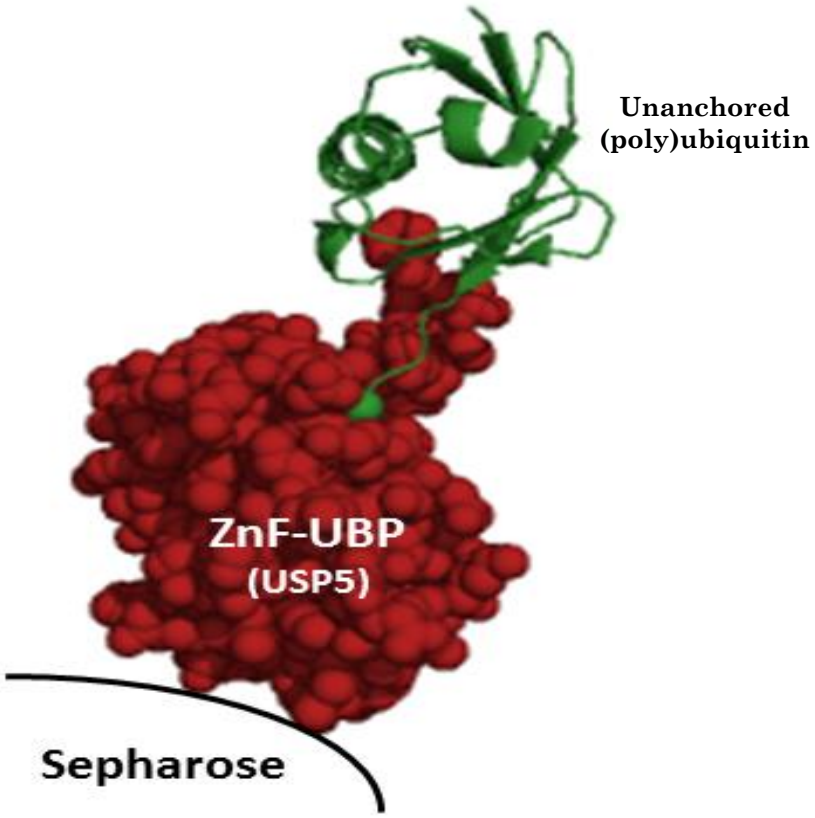


Figure 3.1 Schematic of the FUBE. The crystal structure of the ZnF-UBP domain: monoubiquitin complex (PDB 2G45) highlights the recognition of the C-terminus of unanchored monoubiquitin by the ZnF-UBP domain. The ZnF-UBP domain (isolated from USP5) is covalently conjugated to CNBr-activated Sepharose to generate the FUBE.

3.1 Purification of endogenous unanchored polyubiquitin chains from HEK293T cells utilising the FUBE

3.1.1 Preparation of unanchored polyubiquitin-enriched cell lysate

Investigations by Strachan *et al* [107] established a strategy for processing rat skeletal muscle tissue, prior to enrichment of the unanchored ubiquitome on the FUBE [107]. Subsequently, the broad utility of this approach was realised, owing to the highly conserved nature of the ubiquitin sequence in eukaryotes, to successfully purify unanchored polyubiquitin from mammalian (U20S and HeLa lines), plant and yeast extracts, albeit with limited molecular characterisation [246]. The optimised workflow for this strategy has accordingly been determined, and is illustrated in Figure 3.2. The investigations in this chapter (and thesis) explore the purification of unanchored polyubiquitin from an additional mammalian cell line, namely HEK293T, and provide more detailed MS-based insight into molecular composition of the endogenous unanchored polyubiquitin pool (e.g. different polyubiquitin linkages present).

Cell extracts are first lysed, as per the purification of Zeng *et al* [219], in homogenising buffer containing NEM to inhibit DUBs (NEM is later quenched with DTT prior to incubation with the cysteine containing ZnF-UBP domain). Additionally, the optimised homogenising buffer is supplemented with the metalloprotease inhibitor 1, 10-phenanthroline, MG-132, to transiently inhibit the proteasome and protease inhibitors, to prevent the adventitious proteolysis of polyubiquitin. A critical stage in the isolation of unanchored polyubiquitin, is the incorporation of sample heating step at 75 °C for 20 minutes, followed by centrifugation. This results in the denaturation and precipitation of the majority of cellular proteins, resulting in a lysate containing principally unanchored (poly)ubiquitin, as the majority of conjugated polyubiquitin is ‘removed’ along with denatured substrate proteins. In the experiments throughout this thesis unanchored polyubiquitin was extracted from HEK293T cells.

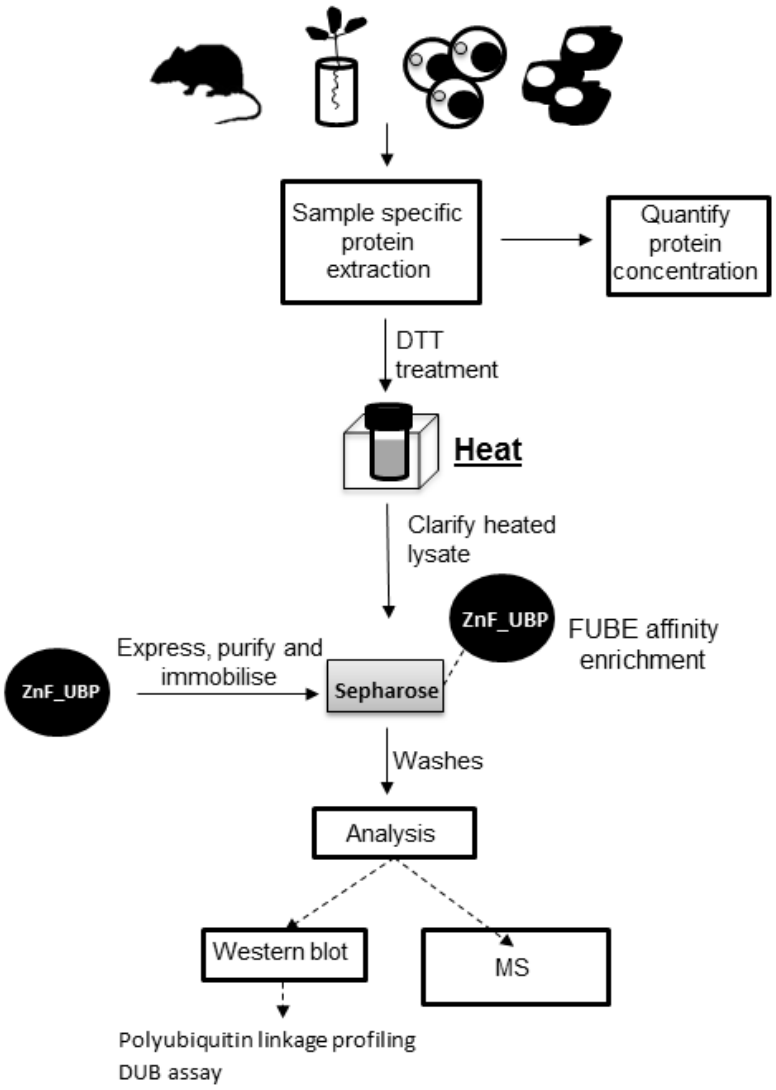


Figure 3.2 The work flow for the FUBE purification of endogenous unanchored polyubiquitin chains from a variety of sources.

3.1.2 Expression and purification of the ZnF-UBP domain of USP5

To complete the purification of unanchored polyubiquitin from mammalian cell extracts (prepared as above), recombinant ZnF-UBP (USP5) domain was expressed, purified and coupled to CNBr-activated Sepharose 4B. Briefly, the cDNA sequence encoding the 129 amino residue ZnF-UBP domain (residues 163-291 of human USP5) was previously cloned in the Layfield lab into the pGEX-4T-1 bacterial expression plasmid,

and was provided at the outset of this study (verified by DNA sequencing). The ZnF-UBP domain was expressed in *E. coli* as a GST-fusion and enriched on glutathione Sepharose, prior to incubation with thrombin, to cleave the GST-tag and yield the ZnF-UBP domain with a Gly-Ser remnant at the N-terminus. The ZnF-UBP domain was then further purified by anionic-exchange chromatography and subject to SDS-PAGE analysis, which resulted in a single band indicated in Figure 3.3 (A). 1 L of culture typically yielded 5 mg of purified protein. The mass of the purified ZnF-UBP protein was confirmed by MS analysis on the Synapt HDMS, to be 14,716 Da (expected 14,717 Da) as illustrated in Figure 3.3.

The ZnF-UBP domain was subsequently immobilised to CNBr-activated Sepharose and utilised to affinity enrich unanchored polyubiquitin from the processed HEK293T cell extracts. Typically the ZnF-UBP domain was coupled to Sepharose at 10 mg/ml, 100 µL of this Sepharose was then used to enrich 15 mg of cell lysate (concentration of the lysate was determined prior to heating, the appropriate volume of lysate (=15 mg of total protein) was added to ZnF-UBP Sepharose post-clarification).

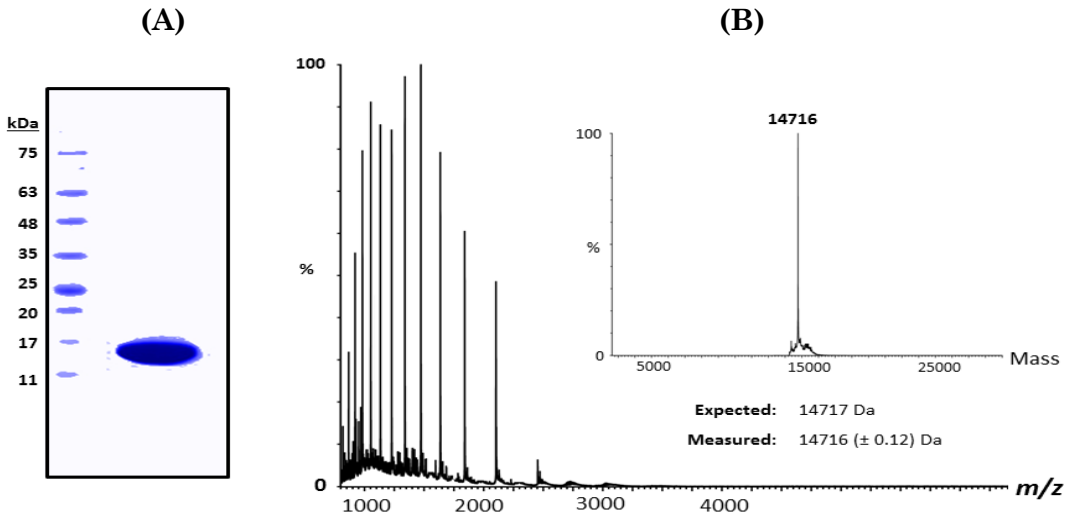


Figure 3.3. Expression and purification of the ZnF-UBP domain of USP5 (FUBE). (A) A representative Coomassie stained 5-20% gradient SDS-polyacrylamide gel following the purification of the ZnF-UBP domain (20 µg loaded). (B) Full and deconvoluted mass spectra of denatured ZnF-UBP sprayed from 80 % MeCN, 0.1 % formic acid (analysis completed on the Synapt HDMS).

3.1.3 Analysis of FUBE enriched unanchored polyubiquitin from HEK293T cell extracts

The validity of the FUBE purification strategy with regards to the purification of unanchored polyubiquitin from HEK293T cells was first confirmed by directly eluting purified material into SDS-PAGE gel application buffer (GAB) and western blotting versus ubiquitin antibodies. Figure 3.4A confirms that ubiquitin is enriched from HEK293T cell extracts by the FUBE Sepharose (and not on control Sepharose). By comparing the SDS-PAGE migration profile of the FUBE-enriched material from HEK293T cell extracts to commercial polyubiquitin ladders (Lys-48 or 63 linked ubiquitin 1-7) (and molecular weight markers) we can deduce that unanchored polyubiquitin chains >10 moieties in length are purified, which supports similar findings in alternate mammalian cell extracts by Strachan *et al* [246].

We then confirmed the sensitivity of the purified polyubiquitin to recombinant WT USP5 in a deubiquitylation assay, to confirm that FUBE-enriched polyubiquitin is unanchored in nature, as evidenced by processing to monoubiquitin. Indeed this was the case, as incubation with WT USP5 resulted in the immunoreactive ladders of polyubiquitin being processed to monoubiquitin, as assessed by Western blotting (Figure 3.4A). The composition of the FUBE-enriched unanchored polyubiquitin pool with respect to polyubiquitin linkages was then assessed by probing immunoreactivity with available Lys-48 and 63 polyubiquitin linkage selective antibodies, alongside Lys-48 or Lys-63 polyubiquitin standards (to confirm antibody specificity). Blotting FUBE enriched material with the Lys-48 polyubiquitin selective antibody (Figure 3.4B) reproducibly detected Lys-48 linked ubiquitin polymers, with the feint presence of a dimer alongside an abundant Lys-48 linked polyubiquitin ladder. This supports previous findings following purification of unanchored polyubiquitin from rat skeletal muscle, where the presence of Lys-48 linked diubiquitin was noted [107]. Blotting of FUBE-enriched material with the Lys-63 linked polyubiquitin selective antibody

(Figure 3.4 C), revealed clear immunoreactivity corresponding to Lys-63 linked di- and triubiquitin. The detection of short unanchored Lys-63 linked polyubiquitin chains supports the notion, originally alluded to by Zeng *et al* [219], that such species are relevant *in vivo*, who implicated such chains in the potent activation of RIG-I [219]. However in this instance it is interesting to note, with the exception of a single high molecular weight species, the absence of long Lys-63 linked polyubiquitin chains, implicated in NF-κB activation [11].

We speculate that the high molecular weight species detected upon Lys-63 selective blotting (Figure 3.4 C) may represent a mixed-linked polyubiquitin chain. This appears plausible owing to the absence of other long Lys-63 linked ubiquitin polymers, although this remains to be tested. With regards to the detection of Lys-63 linked polyubiquitin following FUBE enrichment, the findings presented here and indeed those of other group members (unpublished), revealed the detection of Lys-63 linked polyubiquitin to be variable, perhaps owing to antibody sensitivity, or indeed the extended nature of Lys-63 linked chains, which may be more readily cleaved or more susceptible to losses upon heating.

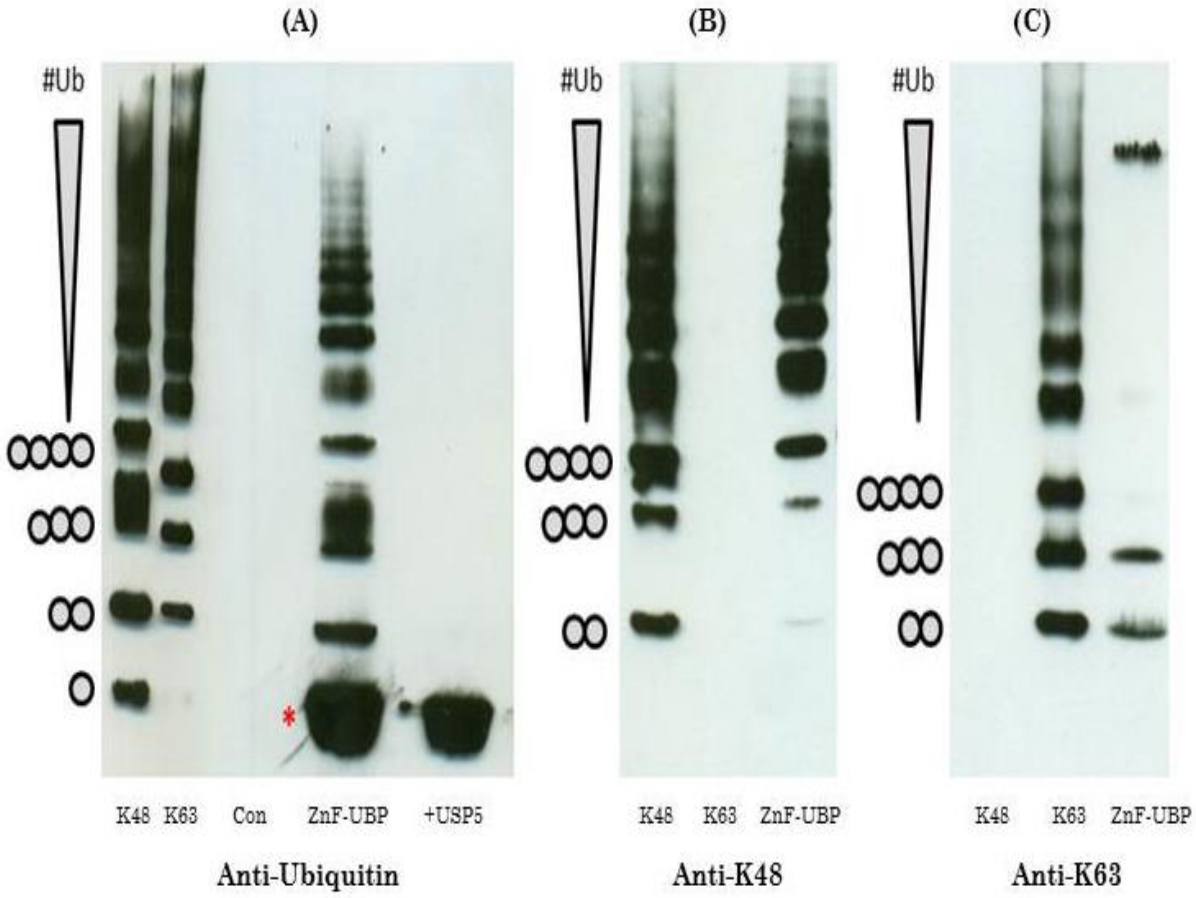


Figure 3.4 Purification of endogenous unanchored polyubiquitin chains from HEK293T cell extracts, using the FUBE purification strategy. Representative anti-ubiquitin western blots of endogenous unanchored polyubiquitin chains, enriched from mammalian HEK293T cell extracts, using the ZnF-UBP Sepharose (10 mg/ml, from USP5). Following enrichment, 50 µL of ZnF-UBP Sepharose was eluted directly into GAB, a further 50 µL of the ZnF-UBP Sepharose was incubated with WT USP5 (2 µg) for 16 hours at 37°C to assess deubiquitylation (+ USP5). Upon completion the reaction was quenched via the addition of 20 µL of GAB. 100% of each sample was separated by SDS-PAGE (on a 5-20% gradient gel), alongside commercial Lys-48 or 63 polyubiquitin standards (100 ng/lane). The gel was transferred to nitrocellulose, prior to immunoblotting versus; (A) anti-ubiquitin (VU-1), (B) a Lys-48 polyubiquitin-specific antibody (apu-2) or (C) a Lys-63 polyubiquitin-specific antibody (hwa4c4). Grey circles indicate ubiquitin chain length (#Ub).* highlights the abundance of the monoubiquitin capture.

3.2 Overcoming the limitations of the FUBE as an enrichment tool for unanchored polyubiquitin

While the FUBE enrichment strategy has permitted the purification of unanchored chains from HEK293T cell extracts (as demonstrated in Figure 3.4) and a range of alternate sources [246], this strategy is inherently limited. Indeed, the ZnF-UBP domain does not favour the purification of polyubiquitin chains over abundant monoubiquitin, highlighted in Figure 3.4, which dominates the purification and impedes downstream proteomic analysis. Accordingly, in the remainder of this chapter (and further extended in *chapter 5*) we sought to investigate a 'second-generation' FUBE to overcome this limitation and favour the selective purification of unanchored polyubiquitin over monoubiquitin.

By considering the recent example of TUBEs, wherein multiple UBDs placed in tandem realise a non-specific avidity effect to favour the non-selective capture of polyubiquitin modified substrates, we rationalised that by fusing the ZnF-UBP to appropriate UBDs, we may be able to generate an avidity effect to afford the selective purification of unanchored polyubiquitin over abundant monoubiquitin, as we evolve towards the linkage selective purification of unanchored chains.

USP5, a model multi-UBD containing DUB, from which the ZnF-UBP domain is isolated, represents a natural extension of the FUBE. USP5 is responsible for the selective disassembly of unanchored polyubiquitin, and can cleave ubiquitin dimers of all eight possible linkages [247]. USP5 achieves the selective recognition of unanchored polyubiquitin chains via avid interactions at four putative UBDs. In addition to the ZnF-UBP domain utilised in the FUBE enrichment strategy, which is responsible for the recognition of the free C-terminus of the proximal ubiquitin, USP5 also contains a catalytic domain and two ubiquitin-associated (UBA) domains, as illustrated in Figure 3.5. Hence we theorised that by utilising the full length catalytically inactive USP5

construct as alternative to the ZnF-UBP domain, under bait-limiting conditions, we could favour the avid interaction of polyubiquitin chains at the tandem UBDs of USP5, and limit the purification of unanchored monoubiquitin.

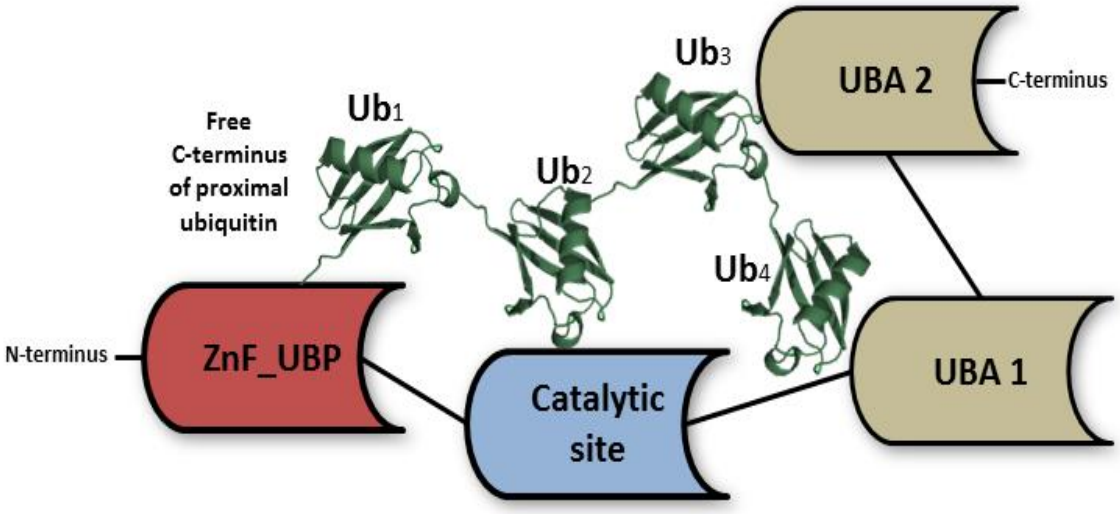


Figure 3.5 Schematic of the proposed mechanism of unanchored tetraubiquitin recognition by USP5. The proximal ubiquitin, with a free C-terminus, is recognised by the ZnF-UBP domain of USP5. Additionally USP5 has three further ubiquitin-binding sites including the catalytic domain and two UBA domains. In the model proposed Reyes-Turcu *et al*, the catalytic domain interacts with the second ubiquitin moiety, the UBA2 (second UBA in order of primary sequence) recognises the third ubiquitin moiety and the UBA1 recognises the fourth ubiquitin moiety.

3.2.1 Expression and purification of USP5

Subsequently we sought to express and purify recombinant full length USP5, to explore the utility as a second generation FUBE. The cDNA sequence encoding the 858 residue USP5 (short isoform) construct was previously cloned in the Layfield lab into the bacterial expression plasmid pGEX-4T-1, and was provided at the outset of this study. USP5 was expressed in *E. coli* as a GST-fusion. Owing to the size of USP5 (~94 kDa), extensive optimisation of overexpression conditions was required, including optimising the incubation time and temperature, following IPTG induction. USP5 was

subsequently purified on glutathione Sepharose, prior to thrombin cleavage, which released USP5 with an N-terminal Gly-Ser remnant. A second purification step, utilising gel filtration, was required which resulted in a single band as assessed by SDS-PAGE (see Figure 3.6A for representative purification). 1 L of culture typically yielded 1 mg of purified USP5. A denatured mass could not be obtained for USP5, however DNA sequencing confirmed construct integrity, and native MS (see *chapter 4*) indicated the mass to be ~94 kDa.

In order to apply USP5 to the purification of unanchored polyubiquitin, we sought to abrogate USP5 catalytic activity. In this regard, previous studies have identified mutation of the catalytic Cys-335 residue of USP5 to Ala as sufficient [87]. Hence this mutation was introduced utilising a modified Quikchange protocol and confirmed by DNA sequencing (see Figure 3.6B). C335A-USP5 was expressed and purified in the same manner as the WT construct. To confirm inactivity, an *in vitro* deubiquitylation assay was completed, comparing the processing of Lys-63 linked tetraubiquitin by WT USP5 to recombinant C335A-USP5 (see Figure 3.6C). This assay demonstrates that under the experimental conditions tested, purified WT USP5 processes tetraubiquitin to monoubiquitin, whilst the catalytic mutant C335A-USP5 does not. Accordingly we can conclude that C335A-USP5 is suitable as an affinity enrichment reagent. In the remainder of the investigations in this thesis C335A-USP5 is utilised, with the exception of DUB assays, where WT USP5 is utilised.

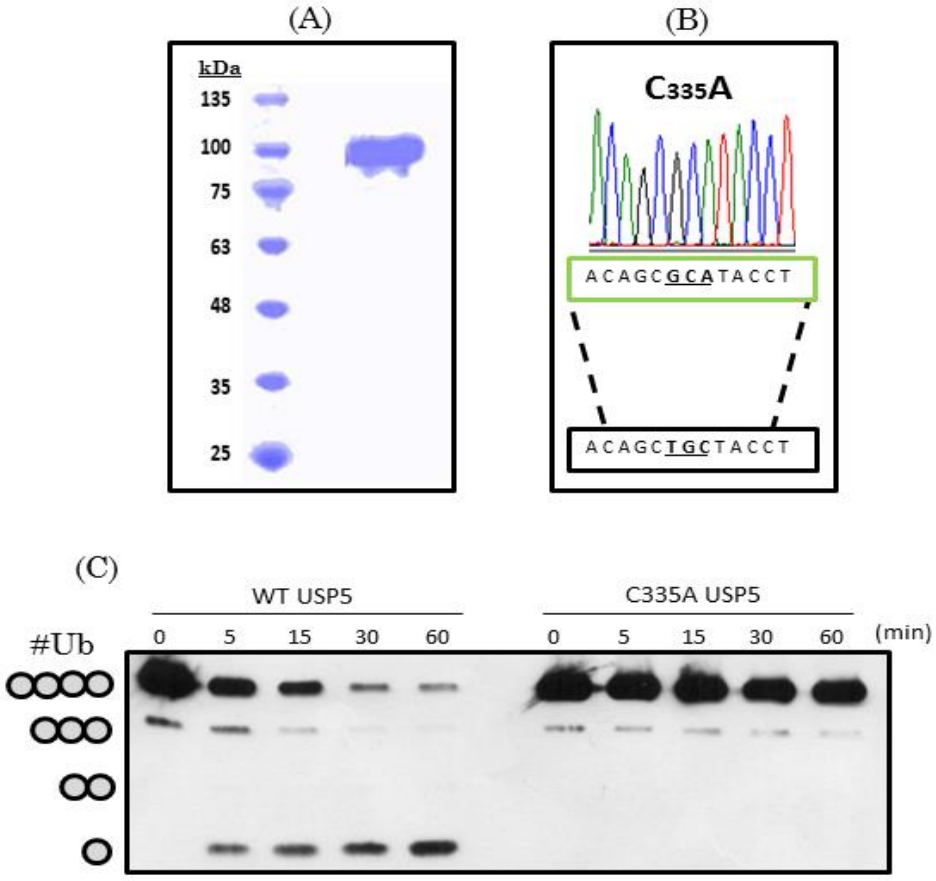


Figure 3.6 Confirming the integrity of the USP5 construct. (A) A representative Coomassie stained 5-20% gradient SDS-polyacrylamide gel following the purification of C335A-USP5 (20 µg loaded). (B) Electropherogram highlighting the mutation of the catalytic Cys-335 of USP5 to Ala. (C) A deubiquitylation assay comparing the deubiquitylation of Lys-63 linked tetraubiquitin (3 ug, Boston Biochem) by WT or C335A-USP5 construct (1:4 enzyme:substrate molar ratio) in a total volume of 30 µL. Deubiquitylation was monitored over a time course of 0-60 minutes at 37 °C, with 6 µL removed and quenched with GAB at the indicated time point. Samples (50 % of aliquot) were separated by SDS-PAGE on a 5-20% gradient gel, and transferred onto nitrocellulose membrane, prior to immunoblotting versus an anti-ubiquitin antibody (VU-1). Grey circles indicate ubiquitin chain length (#Ub).

3.2.2 USP5-ubiquitin binding properties of recombinant USP5

To confirm C335A-USP5 could indeed be utilised to capture polyubiquitin, we first completed *in vitro* ubiquitin-pull-down assays with commercial diubiquitin of defined linkage, as indicated in Figure 3.7. Purified C335A-USP5 was coupled to CNBr-Sepharose (10 mg/ml) and 50 µL was incubated with 250 ng of either *in house* purified Met-1 linked, or commercial Lys-48 or Lys-63 linked diubiquitin, which are readily available and inexpensive. Following overnight incubation, C335A-USP5 Sepharose incubated with each dimer, washed, and the enriched material eluted directly into GAB prior to immunoblotting versus an anti-ubiquitin antibody.

The results, illustrated in Figure 3.7, confirm that C335A-USP5 is functional in the binding of the three ubiquitin dimers tested. Furthermore C335A-USP5 displayed no obvious polyubiquitin linkage preference, which is consistent with investigations by Ritorto *et al* [247] who demonstrated that USP5 can cleave all eight ubiquitin polymers.

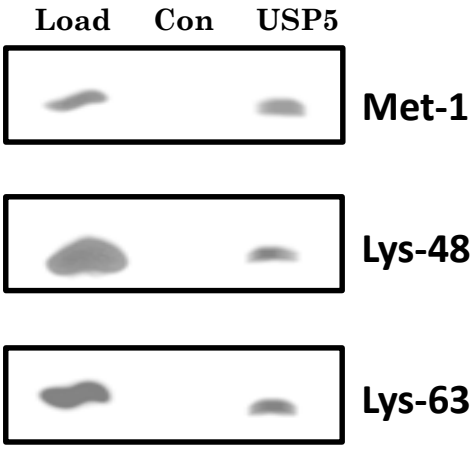


Figure 3.7. Confirming the ubiquitin binding properties of C335A-USP5. 50 µL of control or C335A-USP5 (10 mg/ml) Sepharose was incubated with 250 ng of diubiquitin (Met-1, Lys-48 or 63 linked). Following binding, the Sepharose was washed and the enriched material eluted directly into GAB, prior to separation by SDS-PAGE on a 5-20% gradient gel. The gel was transferred to nitrocellulose membrane and immunoblotted against an anti-ubiquitin antibody (VU-1). Load represents 100% of the polyubiquitin input.

3.2.3 C335A-USP5 as a second generation FUBE

We then applied Sepharose-coupled C335A-USP5 to the purification of unanchored chains from HEK293T cell extracts, adhering to the previously optimised purification strategy, under conditions in which both the ZnF-UBP and C335A-USP5 were assumed to be in excess. Figure 3.8 demonstrates that C335A-USP5 can be utilised to purify unanchored (poly)ubiquitin from HEK293T cell extracts. As with unanchored polyubiquitin purified on the ZnF-UBP domain, unanchored polyubiquitin captured by C335A-USP5 have a free C-terminus, as evidenced by processing by WT USP5 to monoubiquitin in a deubiquitylation assay. Figure 3.8 also highlights that enrichment of unanchored polyubiquitin on either ZnF-UBP or C335A-USP5 Sepharose, results in broadly similar ubiquitin immunoreactive profiles, with the abundant monoubiquitin pool enriched in each case, suggesting no clear advantage of C335A-USP5 Sepharose under conditions in which the bait protein is in excess. Subtle differences between the ZnF-UBP domain and C335A-USP5 with respect to the anti-ubiquitin immunoreactivity profile, and hence the purification of polyubiquitin, indicates that perhaps the bait was not entirely in excess, and alludes to the possibility of favouring longer polyubiquitin chains on C335A-USP5 under conditions in which the bait protein is limiting.

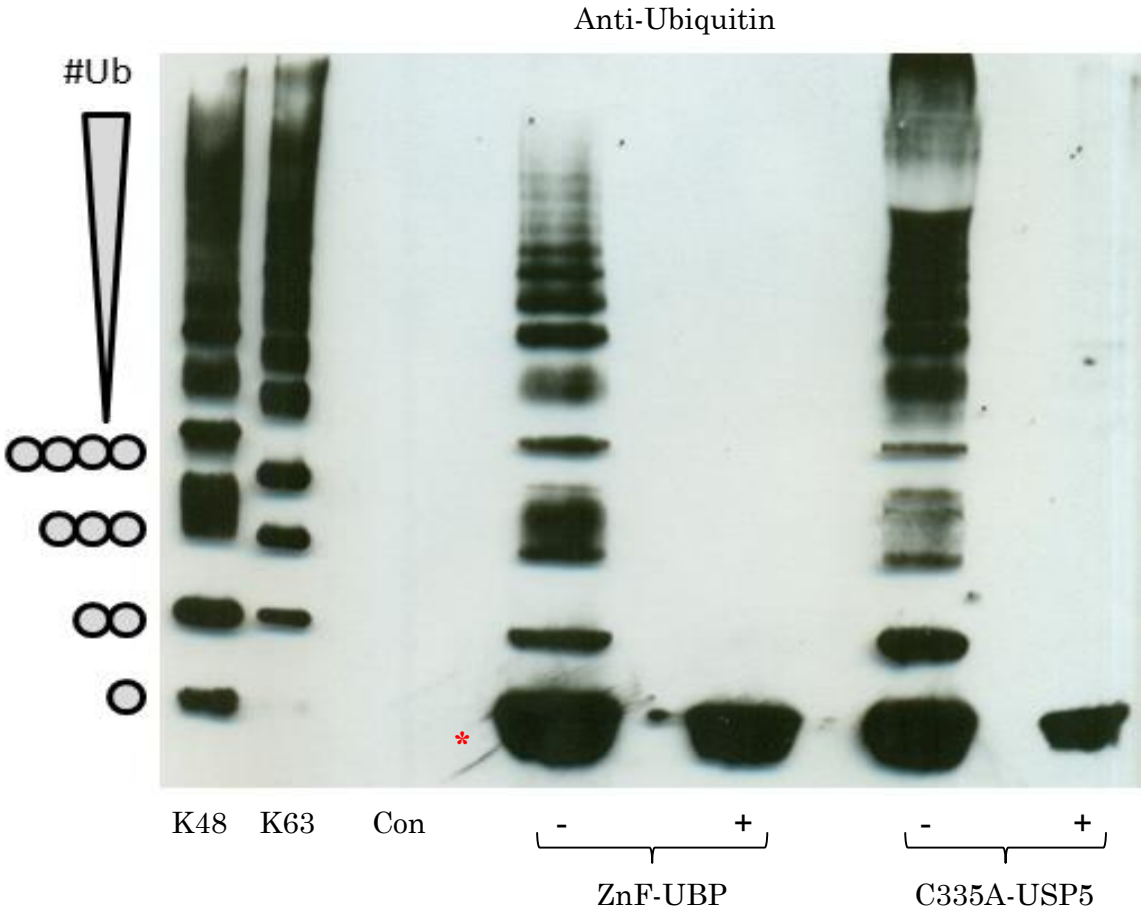


Figure 3.8 USP5 as a second generation FUBE. Representative anti-ubiquitin western blot of endogenous unanchored polyubiquitin chains, enriched from mammalian HEK293T cell extracts, using control, ZnF-UBP (10 mg/ml) or C335A-USP5 (64 mg/ml, equimolar to ZnF-UBP) Sepharose. Following enrichment, 50 µL of each Sepharose was eluted directly into GAB (-), a further 50 µL of the ZnF-UBP and C335A-USP5 Sepharose was incubated with WT USP5 (2 µg) for 16 hours at 37°C to assess deubiquitylation (+). Upon completion the reaction was quenched via the addition of 20 µL of GAB. 100% of each sample was separated by SDS-PAGE (on a 5-20% gradient gel), alongside commercial Lys-48 or 63 polyubiquitin standards (100 ng/lane). The gel was transferred to nitrocellulose, prior to immunoblotting versus anti-ubiquitin (VU-1). Grey circles indicate polyubiquitin chain length (#Ub).* highlights the abundance of the monoubiquitin capture.

Previous investigations by Strachan *et al* [246] established the optimal conditions for FUBE enrichment from alternative mammalian cell extracts, with the findings indicating that 100 μL of 10 mg/ml Znf-UBP Sepharose is in excess when enriching 15 mg of mammalian cell extract (total protein quantified prior to the heating stage). Indeed these conditions were adhered to in the experiment in Figure 3.8. Accordingly, to ensure the bait protein was indeed limiting in subsequent investigations, 100 μL of ZnF-UBP or C335A-USP5 Sepharose was used to enrich 15 mg of HEK293T cell extracts (total protein quantified prior to the heating stage), however the ZnF-UBP domain was coupled at a concentration of 5 mg/ml, (with C335A-USP5 at an eqimolar concentration), instead of 10 mg/ml, reducing bait protein availability by half.

As noted in Figure 3.9 under these conditions C335A-USP5 significantly disfavoured the enrichment of free monoubiquitin, when comparing the ubiquitin immunoreactivity profile to enrichment on the ZnF-UBP domain. Furthermore it was evident that the intensity of the polyubiquitin smear, corresponding to the enrichment of longer unanchored polyubiquitin chains was increased moving from the isolated ZnF-UBP domain to the full length (catalytic mutant) USP5 protein, indicating that C335A-USP5 favours the capture of unanchored polyubiquitin over monoubiquitin, under these bait limiting conditions. To assess the polyubiquitin linkage profile of the purified unanchored polyubiquitin pool, the enriched material was probed with a Lys-48 polyubiquitin linkage selective antibody. This confirmed that enrichment of Lys-48 linked unanchored polyubiquitin is increased on C335A-USP5 Sepharose (under the conditions used), with respect to purification on ZnF-UBP Sepharose.

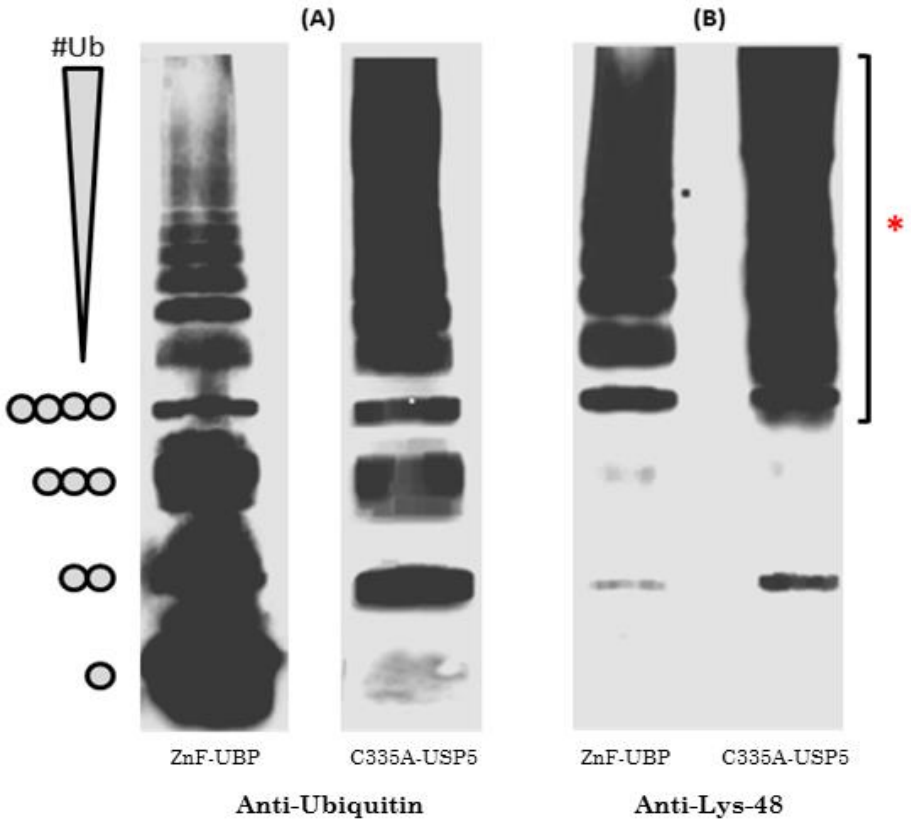


Figure 3.9. Optimising the enrichment of endogenous unanchored polyubiquitin from HEK293T cell extracts with C335A-USP5. Representative anti-ubiquitin western blots of endogenous unanchored polyubiquitin chains, enriched from mammalian HEK293T cell extracts, using ZnF-UBP (5 mg/ml) or C335A-USP5 (32 mg/ml, eqimolar to ZnF-UBP). Following enrichment, 50 µL of ZnF-UBP Sepharose was eluted directly into GAB. 100% of each sample was separated by SDS-PAGE (on a 5-20% gradient gel). The gel was transferred to nitrocellulose, prior to immunoblotting versus; (A) anti-ubiquitin (VU-1), (B) a Lys-48 polyubiquitin selective antibody (apu-2). Grey circles indicate ubiquitin chain length (#Ub). * highlights the differences in the polyubiquitin immunoreactive smear.

3.3 Mass spectrometry analysis of endogenous unanchored polyubiquitin

To validate the immunoblotting results that indicate the presence of Lys-48 linked polyubiquitin chains (and Lys-63 in the case of the ZnF-UBP domain) in the purified unanchored polyubiquitin pool enriched from mammalian (HEK293T) cell extracts, and to further explore the composition of the unanchored polyubiquitin pool, with respect to lysine linkages present, we sought to apply MS-based approaches.

As discussed in *section 1.3.3*, trypsinolysis of polyubiquitin yields a series of signature ubiquitin peptides, each of which contain a Gly-Gly modified Lys residue (at which ubiquitin was conjugated), which is spared from trypsin cleavage. The polyubiquitin signature peptides are indicated in Table 3.1, and form the basis of MS identification of polyubiquitin.

Table 3.1 The unique signature peptides yielded upon trypsin digest of polyubiquitin. (GG) in red denotes di-Gly remnant.

Linkage	Signature peptide	Mass (Da)
K6	MQIFVK(**GG**)TLTGK	1378.6
K11	TLTGK(**GG**)TITLEVEPSDTIENVK	2401.26
K27	TITLEVEPSDTIENVK(**GG**)AK	2100.1
K29	AK(**GG**)IQDK	815.45
K33	IQDK(**GG**)EGIPPDQQR	1636.81
K48	LIFAGK(**GG**)QLEDGR	1459.78
K63	TLSDYNIQK(**GG**)ESTLHLVLR	2243.2

Preceding MS analyses of ZnF-UBP enriched unanchored polyubiquitin from rat skeletal muscle were limited, and indeed only extended to the identification of gel excised diubiquitin bands, which upon trypsin cleavage and MS/MS analysis yielded signature peptides corresponding to Lys-11 and 48 linked chains [107]. Hence to date no MS/MS analyses had been reported detailing the molecular composition of the complete unanchored polyubiquitin pool, purified from mammalian extracts.

Following enrichment on ZnF-UBP or C335A-USP5 Sepharose, as in Figure 3.9, unanchored polyubiquitin isolated from HEK293T cell extracts (typically from 200 µL of Sepharose) was eluted into 0.1% formic acid (pH 2). Due to the compact, stable nature of ubiquitin, complete trypsinolysis proves challenging. Accordingly a modified FASP strategy [227] was developed *in house*, to ensure complete trypsin digestion. Briefly, eluate was buffer exchanged (Amicon ultra, 0.5 ml, 10 kDa MWCO) into digestion buffer (2 M urea, 0.1 M ammonium bicarbonate (pH 8), 5% (v/v) acetonitrile), reduced (50 mM DTT), alkylated (100 mM chloroacetamide [248]) and digested with trypsin (sequencing grade modified porcine trypsin, three additions at four hour intervals) at 37 °C for ~24 hours. Peptides were extracted (70% acetonitrile, 0.1% formic acid) and vacuum centrifuged to dryness.

MS/MS analysis was first completed *in house* on the LTQ FT Ultra in a data-dependent manner. Raw data was searched using Bioworks (Thermo Fisher Scientific) against a 'custom' ubiquitin database in which the ubiquitin sequence (human) was manually inputted. The use of a 'custom' protein database for interpretation of the MS/MS data was required due to the complex ubiquitin gene organisation (mammals have multiple ubiquitin genes, several of which involve tandem repeats) and the nature of identifying PTMs using this software. The data was filtered for peptides modified with Gly-Gly (+114.042927 on lysine) and Leu-Arg-Gly-Gly (+383.228103 on Lysine) in the event of missed trypsin cleavage [249].

We routinely identified 95 % sequence coverage of ubiquitin (72/76 amino acids), with residues at the extreme C-terminus (LRGG) undetected. Furthermore polyubiquitin signature peptides corresponding to chains linked by Lys-11, 48 and 63 were consistently identified in the HEK293T unanchored polyubiquitin pool, enriched on either ZnF-UBP or C335A-USP5 Sepharose, which proved reproducible across numerous biological (including with variances in amounts of starting material purified) and technical replicates. The effectiveness of the database search was confirmed by

manually interrogating the MS/MS data, where no further isopeptide signature peptides were revealed. To more broadly assess the purification strategy, the MS/MS data was retrospectively searched against a human database (Uniprot, 67,911 protein entries), filtering for the same parameters as outlined above. This revealed the presence of a handful of other proteins, including 'bait' proteins from the affinity beads, as well as common contaminants invariably captured by control Sepharose e.g. keratin and serum albumin. Furthermore, Gly-Gly stubs were exclusively associated with ubiquitin and not other ubiquitin-modified proteins, indicating enrichment of unanchored polyubiquitin only.

To validate our *in house* findings, we also completed analysis of material enriched on either ZnF-UBP or C335A-USP5 Sepharose on an LTQ OT Velos instrument (completed in collaboration with Dr Andrew Bottrill at the University of Leicester). Reassuringly ZnF-UBP and C335A-USP5 enriched samples both contained the three polyubiquitin signature peptides noted by *in house* analysis (Lys-11, 48 and 63). Furthermore the remaining four signature peptides, corresponding to isopeptide linked polyubiquitin lined by Lys-6, 27, 29 and 33, were also identified in each sample. Representative MS/MS spectra are illustrated in Figure 3.10.

While the absolute abundance of polyubiquitin chains is a source of debate, the consensus is that Lys-11, 48 and 63 polyubiquitin linkages are the most abundant in HEK293T, and more broadly mammalian cell extracts [108, 186], which supports the initial *in house* findings. The subsequent identification of unanchored polyubiquitin chains linked by residues other than Lys-11, 48 and 63, represents the first direct demonstration that the endogenous unanchored polyubiquitin pool includes these chain types. The development of a quantitative analysis has not yet been realised *in house*, however spectral counting allows us to draw comparisons of relative abundance, between samples purified and analysed in tandem (Figure 3.11). We completed two identical biological replicates of each of the ZnF-UBP and C335A-USP5 enriched

samples, under the purification conditions established in Figure 3.9. Broadly the distributions of linkages (relative number of spectral counts) for each of the bait proteins were similar, with Lys-48 and Lys-63 polyubiquitin signature peptides the most commonly sequenced, supporting the notion that these polyubiquitin chain types are the most abundant. It is interesting to note that Lys-63 linked chains are seemingly enriched on C335A-USP5 sepharose, relative to the ZnF-UBP sepharose, we can speculate that this may result from avid interactions with long, and perhaps mixed-linked unanchored polyubiquitin chains previously indicated by Western blotting of ZnF-UBP enriched material with the Lys-63 selective antibody (see Fig 3.4C).

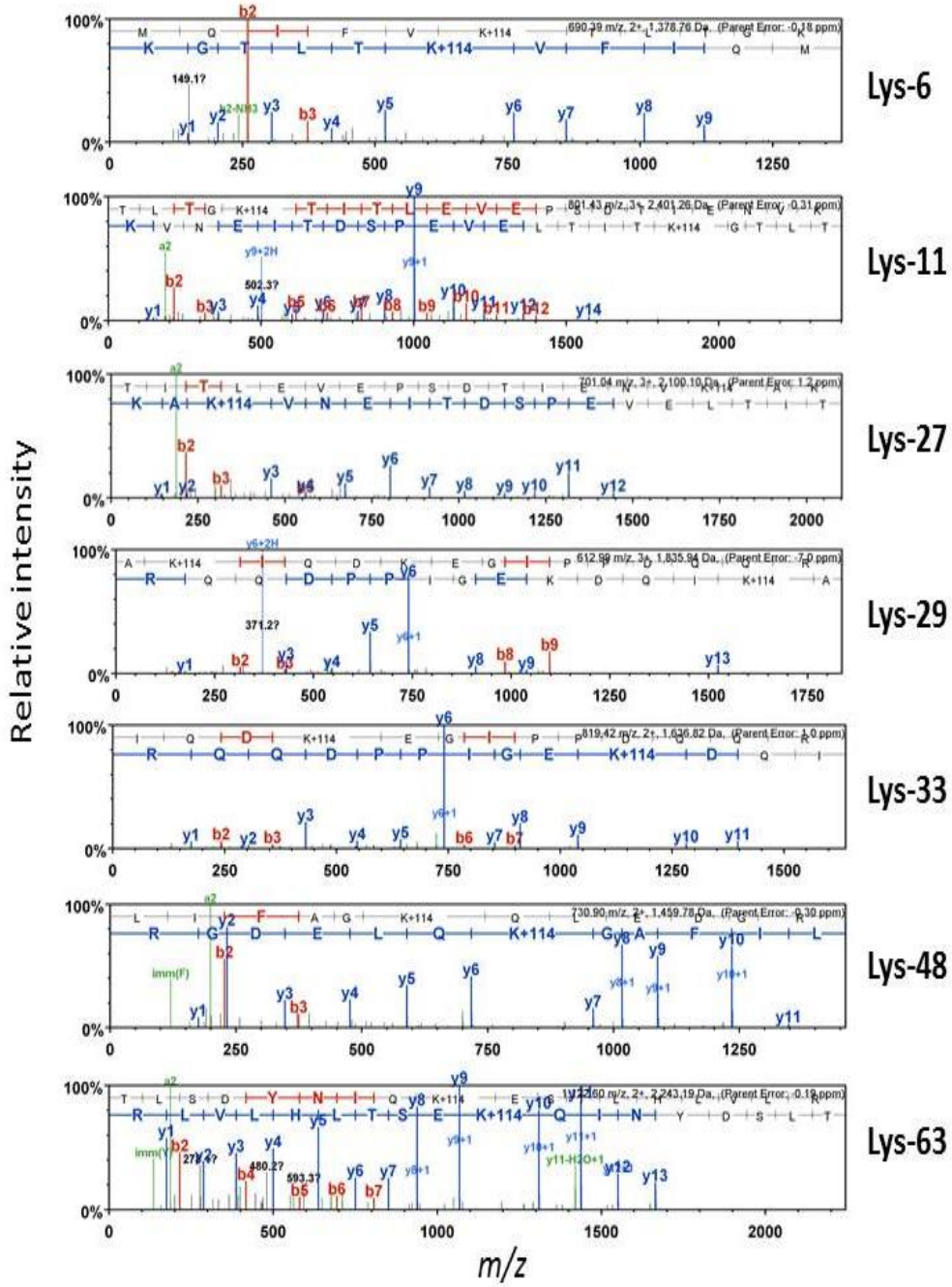


Figure 3.10 Representative MS/MS spectra of identified polyubiquitin signature peptides following analysis of ZnF-UBP or C335A-USP5 enriched unanchored polyubiquitin (from HEK293T lysates) on the LTQ OT Velos. Lysine linkage indicated to the right of spectrum.

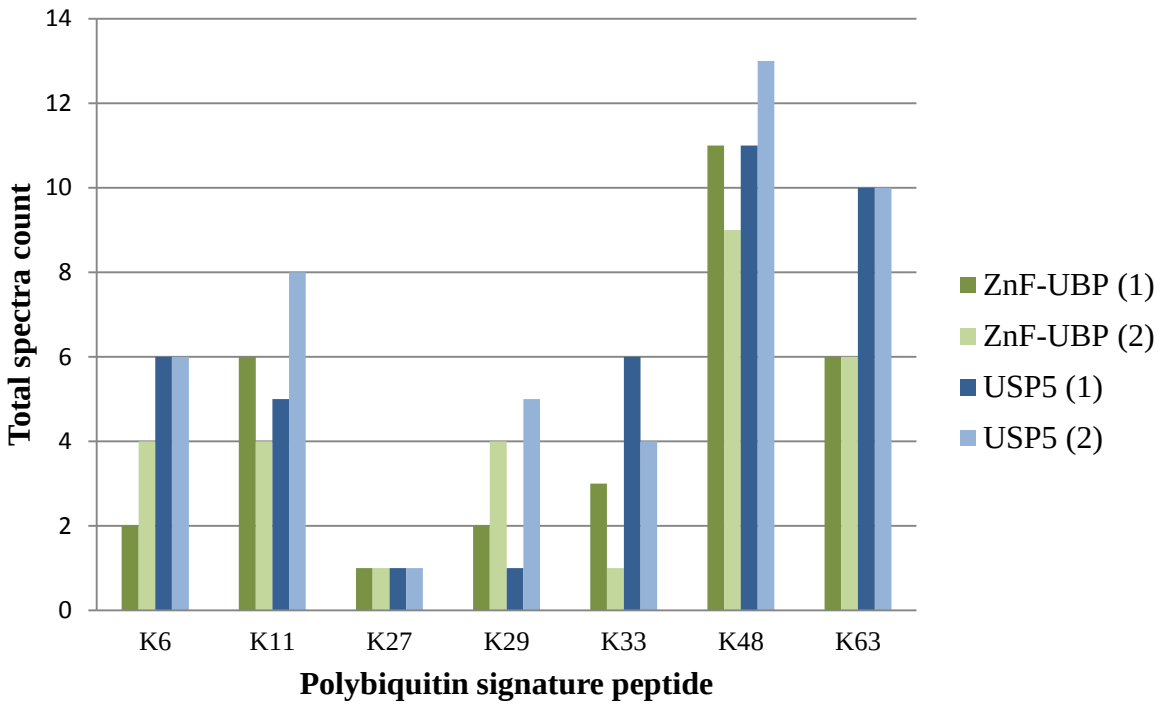


Figure 3.11. MS investigation into the molecular composition of the unanchored polyubiquitin pool, enriched on ZnF-UBP, or C335A-USP5 Sepharose. Two biological replicas of ZnF-UBP and C335A-USP5 Sepharose (200 µL, purified in the same manner as Figure 3.9) were subject to tryptic digestion and LC-MS/MS on the LTQ OT Velos to identify polyubiquitin signature peptides. The bars represent total spectral count of each ubiquitin signature peptide.

3.4 Conclusions

In this chapter we present and apply the FUBE purification strategy, which utilises the ZnF-UBP domain of USP5 to selectively purify unanchored polyubiquitin from, for the first time, mammalian HEK293T cell extracts. We then presented a further iteration of the first generation FUBE, wherein catalytically inactive full-length USP5, containing multiple UBDs in tandem (including the ZnF-UBP domain), is employed to afford the enrichment of unanchored polyubiquitin over abundant monoubiquitin, under bait-limiting conditions. The integrity of purified polyubiquitin (i.e. substrate-free nature) was confirmed via deubiquitylation assays with WT USP5.

The purification of unanchored polyubiquitin presented here, derived from unstimulated HEK293T cell extracts, supports the notion that unanchored polyubiquitin is a relevant and abundant *in vivo* species. The precise source of this unanchored polyubiquitin pool remains unclear from our analyses, however ubiquitylation machinery (E_2 and E_3 enzymes) which synthesize unanchored polyubiquitin *de novo* [11, 200, 209, 215, 219] as well as deubiquitinating enzymes which cleave polyubiquitin chains *en bloc* have been reported [205-207, 210, 211].

Purifications from mammalian cell extracts, utilising both the ZnF-UBP domain and full length USP5, resulted in the capture of long polyubiquitin chains (>10 moieties, based on migration on western blot). This supports previous investigations (in alternative mammalian cell lines [246]), and raises interesting questions regarding the functional roles of such chains. Early studies that probed the proteasome noted that Lys-48 linked polyubiquitin affinity for the proteasome, and thus ability to competitively inhibit the proteasome, increased proportional to polyubiquitin chain length (up to the 8-mer tested) [250]. Accordingly it is interesting to speculate that long unanchored polyubiquitin chains, perhaps generated at the proteasome, could in turn have auto-inhibitory effects via competitive inhibition, to regulate the rate of proteasomal degradation. Short (i.e. dimers) unanchored polyubiquitin chains could also feasibly perform this function, as indeed regulation of other non-proteasome associated DUBs (UCH-L3) by unanchored Lys-48 linked dimers has previously been indicated [251]. However this appears unlikely, as such chains have only a minimal inhibitory effect on proteasome activity [250].

Historically Lys-48 linked tetraubiquitin was considered to be the principal signal for proteasomal degradation, however recent biochemical and biophysical investigations have begun to re-define the requirements for proteasomal degradation, implicating smaller polyubiquitin units [252]. Accordingly the unanchored polyubiquitin pool may act as a store of monoubiquitin, or owing to *en bloc* transfer of pre-formed polyubiquitin

units, may represent a pre-assembled reservoir of polyubiquitin that can be "extracted" from the cytoplasm and transferred to selected substrates to permit rapid protein ubiquitylation upon cellular insult [199, 200].

In addition to chain length dependent roles, Lys linkage specific roles have also emerged for unanchored polyubiquitin. Indeed, Lys-48 linked unanchored ubiquitin has been implicated in the activation of kinases involved in energy metabolism [253], whilst Lys-63 linked chains have been connected to linkage specific roles in innate immune signalling [219]. As alluded to in *section 1.2,* in recent years techniques to probe the polyubiquitin profile of purified samples has evolved. In this chapter we first utilise linkage selective antibodies to confirm the presence of Lys-48 and 63 linked polyubiquitin chains in the endogenous unanchored polyubiquitin pool. However profiling the polyubiquitin profile of FUBE enriched material utilising linkage selective antibodies is inherently limited, owing to inadequate availability of such polyubiquitin linkage selective antibodies. Furthermore investigations in the Layfield group have established detection of unanchored polyubiquitin with the Lys-63 selective antibody to be inconsistent across alternative cell types. Indeed it has been established that Lys-63 linked hexa-ubiquitin represents an abundant species *in vivo*, as purified by Zeng *et al* [219]. However in Figure 3.4, we only detect di- and triubiquitin, with feint bands corresponding to tetra- and pentaubiquitin chains (We also note a high molecular weight band corresponding to a Lys-63 linked polymer of unknown nature). Consequently we sought to utilise MS to probe the linkage profile of the unanchored polyubiquitin pool.

Previous MS analysis of FUBE (isolated Znf-UBP domain) enriched material from rat skeletal muscle revealed the presence of signature peptides corresponding to Lys-11 and 48 [107]. Furthermore qualitative analysis of *in vitro* ubiquitylation assays of TRIM5, an E_3 ligase implicated in the synthesis of unanchored Lys-63 linked polyubiquitin chains involved in innate immune signalling, revealed signature peptides corresponding to this linkage only [254]. While absolute quantities of total

polyubiquitin linkages varies, it has been established that Lys-11, 48 and 63 are the most abundant in HEK293T cells [108]. Our initial *in house* analysis indicates this overall linkage abundance may also be reflected in the unanchored polyubiquitin pool, with the detection of the signature peptides corresponding to Lys-11, 48 and 63 linked polyubiquitin.

Upon validation of in-house findings on the increasingly sensitive LTQ OT Velos instrument, we noted the detection of the polyubiquitin signature peptides, corresponding to all seven isopeptide linked chains in purified samples from HEK293T cell lysates. Whilst we do not discount Met-1 linked chains, and indeed USP5 is demonstrated to capture Met-1 linked diubiquitin *in vitro* (Figure 3.7), manual interrogation of MS/MS data did not reveal the corresponding signature peptide, which could be due to the low abundance or lability of this ubiquitin polymer *in vivo* [108]. The identification of all isopeptide linked ubiquitin polymers raises the possibility that unanchored ubiquitin-mediated processes may be more broadly regulated by polyubiquitin chains of diverse linkages, in addition to the established roles for Lys-48 and 63 linked polymers.

In conclusion we establish the FUBE purification protocol and present the concept of exploiting 'natural' tandem UBDs to favour the purification of unanchored polyubiquitin. Studies thus far have focused on HEK293T cells, cultured under basal conditions, and profiling the global unanchored ubiquitome. However the functional significance of unanchored chain linkage and length and dynamic changes in this sub-ubiquitome are important questions which remain to be addressed. Accordingly a key challenge remains to develop tools which can selectively purify unanchored ubiquitin polymers of defined linkage and length to aid this characterisation.

Chapter 4 - Exploring the mechanisms of polyubiquitin recognition by USP5

USP5 a ~94 kDa DUB, is responsible for selectively processing the biologically significant pool of unanchored polyubiquitin [11]. Additionally USP5 has also been directly implicated in other biological contexts, including in the repair of DNA double-strand breaks [255] and with aberrant splicing linked to glioblastoma tumourigenesis [256]. Wilkinson and co-workers have probed the structure of USP5 using a combination of mutagenesis and ITC, and revealed that at the level of primary structure, USP5 contains four putative ubiquitin-binding sites (domains), linked by flexible loops [87], as well as a cryptic UBP domain, which although does not contribute to ubiquitin binding, appears essential for catalytic function [237].

The ubiquitin-binding sites of USP5, illustrated previously in Figure 3.5, include a ZnF-UBP domain, that binds with high affinity to the free C-terminus of (poly)ubiquitin, and hence underlies USP5's recognition of unanchored polyubiquitin chains [84]. USP5 is also composed of two UBA domains, which as indicated by mutagenesis studies, contribute to avidity effects by binding the third and fourth ubiquitin moieties in an unanchored polyubiquitin chain [87]. Significant ubiquitin interaction also occurs at the papain-type catalytic domain. This was highlighted by the recent crystal structure of USP5, in which USP5 was co-crystallised with a ubiquitin suicide substrate, which modified the USP5 catalytic cysteine covalently [237].

The X-ray structure of USP5, illustrated in Figure 4.1, has offered insight into the relative positioning of each of the domains of USP5 [237]. However as alluded to in

Figure 4.1, the orientation of the ZnF-UBP domain of USP5, with respect to the enzyme active site, appears incompatible with its proposed function in recognition of unanchored polyubiquitin, with the ZnF-UBP domain pointing away from the active site. Interestingly this crystal structure or spatio-temporal 'snap shot' of USP5 is missing ~150 residues (of 835), an indicator of extensive conformational disorder [257]. Furthermore, subsequent biophysical characterisation of the USP5 structure by small angle X-ray scattering (SAXS) (in the same study), a solution-based structural technique, showed the existence of a structural form(s) significantly more extended than that seen by crystallography [237]. These observations indicate considerable flexibility in the structure of USP5. It has been proposed that the inherent flexibility of USP5 permits the recognition of structurally unique ubiquitin polymers. Indeed other USP family members, noted as being non-specific (cleaving all polyubiquitin chain types), are associated with drastic conformational changes upon ubiquitin binding [258, 259]. The results presented in *chapter 3* support this notion, with C335A-USP5 Sepharose enriching seven of the eight polyubiquitin linkages from mammalian cell extracts, whilst previous *in vitro* findings have demonstrated that USP5 can cleave all polyubiquitin linkages [247].

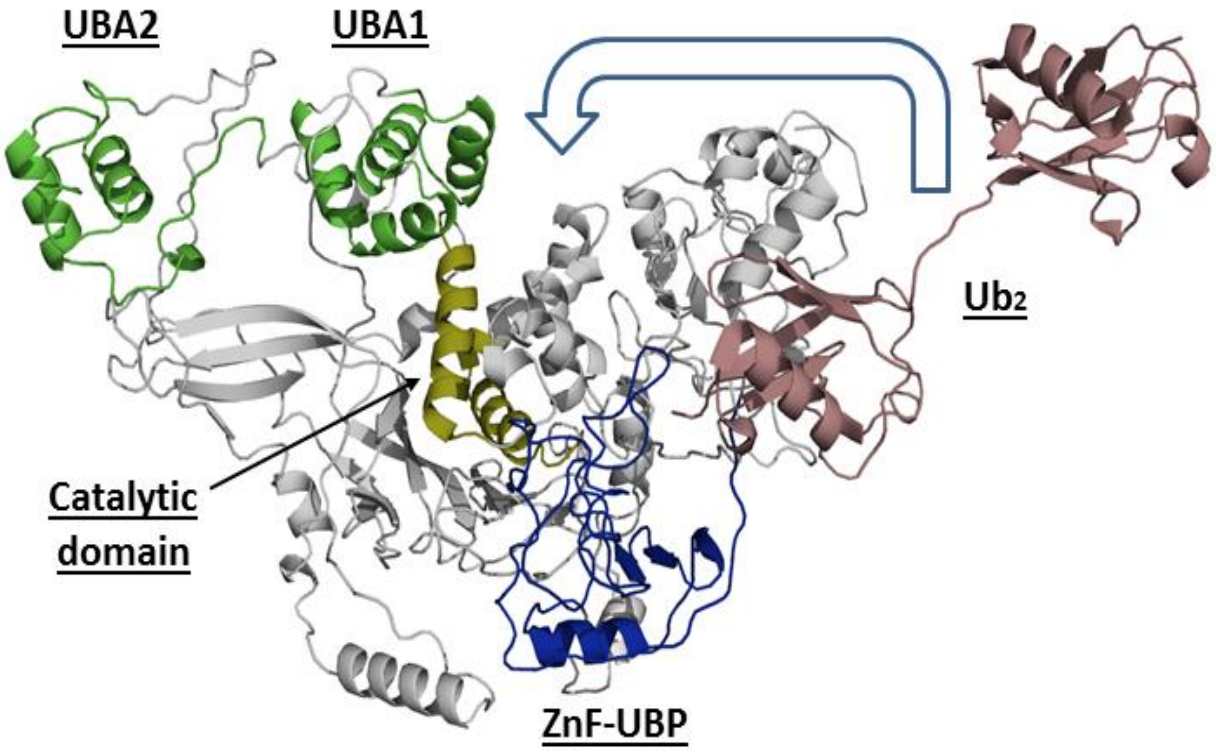


Figure 4.1 Structure of human USP5 (short isoform, PDB 3IHP) with the ZnF-UBP-monoubiquitin complex (PDB 2G45) docked onto the ZnF-UBP domain (blue) to generate an accurate representation of ubiquitin (brown) binding. A further ubiquitin moiety is guided (by the linear diubiquitin structure, PDB 2W9N) to Met-1 of the ZnF-UBP bound proximal ubiquitin moiety. This figure illustrates that the recognition of the distal diubiquitin moiety by the catalytic domain (yellow) or indeed the UBA domains (green) is incompatible with this model. Thus this model indicates a significant structural rearrangement is required, as highlighted by the blue arrow.

In order to study the conformational plasticity of proteins or protein-complexes, techniques must be capable of scrutinising multiple conformations. In this regard, native MS and IMS, particularly when combined with ESI, are powerful approaches, and accordingly have rapidly emerged as important tools in the characterisation of protein structure and interactions [162, 164, 165, 260-262]. It has been established that despite some (protein dependent) collapse upon desolvation, elements of higher protein structure can be maintained in the gas phase to yield biologically relevant structural information [139-141, 263, 264]. Furthermore, numerous studies have demonstrated that non-covalent protein-small molecule and protein-protein interactions can survive the ESI process to be detected and quantified by MS [137, 138, 161, 167, 265]. Investigations by Sokratous *et al* [198]established that following careful optimisation of instrument conditions, native MS could be utilised to quantitatively probe the affinity of UBD-(poly)ubiquitin interactions. Reassuringly, binding affinities estimated by this method, for a range of UBDs with differing polyubiquitin substrates, as indicated in table 4.1, proved to be in excellent agreement with data derived from other biophysical techniques, but proved advantageous with respect to sample consumption and analysis time [198].

Table 4.1 Summary of binding affinities (K_D, μM) for selected UBD-(poly)ubiquitin interactions, as previously derived by ESI-MS by Sokratous et al [198].

UBD	**monoUb**	**Lys-48 Ub_2**	**Lys-63 Ub_2**	**Lys-11 Ub_2**	**Lys-27 Ub_2**	**Lys-48 Ub_2**	**Lys-63 Ub_4**
UBA (UQ1)	22 ± 2	15 ± 2	24 ± 2	21 ± 2	35 ± 3	13 ± 2	15 ± 2
UBA2 (hHR23A)	200 ± 14	18 ± 2	126 ± 18	72 ± 10	81 ± 10	8 ± 1	ND
MIU (Rabex-5)	36 ± 4	11 ± 2	16 ± 2	12 ± 2	ND	ND	ND
ZnF-UBP (USP5)	2.3 ± 0.2	3.6 ± 0.8	2.8 ± 0.3	1.7 ± 0.2	ND	ND	ND

Alongside native MS experiments, additional MS-based strategies including hydrogen-deuterium exchange (HDX), cross-linking and covalent labelling (extensively reviewed in [170]), have been developed to probe protein conformation, dynamics and interactions. While the information afforded by structural MS-based approaches is often of inherently low resolution, this is offset by the increased speed and sensitivity (reduced sample consumption) of the technique in comparison to classic approaches such as X-ray crystallography, NMR spectroscopy and cryo-EM. Furthermore, MS-based approaches are more amenable to reporting on the structural integrity of large, dynamic or heterogeneous protein complexes [167, 262]. Accordingly, structural MS approaches are of significant value to proteins, or protein complexes, which are limiting for reasons of time, complexity or size.

In this chapter we present a structural MS-based workflow, including native IM-MS and carbene footprinting, to probe the structure of the USP5 protein, to clarify X-ray structure inaccuracies and to rationalise the mechanisms of ubiquitin recognition by this model multi-UBD containing construct.

4.1 Probing the conformational dynamics of USP5

4.1.1 Charge-state distribution analysis of USP5

To probe USP5 structure by native MS, full-length C335A-USP5 was expressed and purified as described in *chapters 2* and *3*, and utilised in the structural MS experiments throughout. Prior to native MS analysis, thorough desalting/buffer exchange was completed to prevent the adduction of common purification cations such as potassium and sodium, which upon MS analysis results in suppression of the protonated signal, and add complexity to the spectrum. C335A-USP5 was buffer exchanged into 200 mM ammonium acetate (pH 7) using Vivaspin ultrafiltration devices (10 kDa MWCO). Ammonium acetate is the buffer of choice for native ESI-MS investigations, due to its near neutral pH in water and its volatility, which allow it to be easily removed from analytes during ESI, to produce clean mass spectra.

Purified C335A-USP5 (10 μM, typically 1-10 μL) was loaded into an *in house* manufactured nESI capillary and inserted into the standard waters nESI platform prior to native nESI-MS analysis using a Waters Synapt HDMS operated in positive ion mode. Instrument parameters were carefully optimised to permit efficient desolvation and transmission of ions, whilst minimising the amount of internal energy transferred to the protein. Critically, to preserve non-covalent interactions and avoid gas-phase unfolding, minimal voltages were applied throughout the instrument (see Table 2.8). Furthermore, a high backing pressure (~4-4.5 mBar) was applied to achieve 'collisional cooling' and thus facilitate efficient transmission of large (USP5) ions [266]. Figure 4.2 illustrates the representative native mass spectrum of C335A-USP5, which upon deconvolution, revealed a measured molecular mass of 93,792 Da (expected 93,420 Da, within <1% error associated with measurement of large proteins).

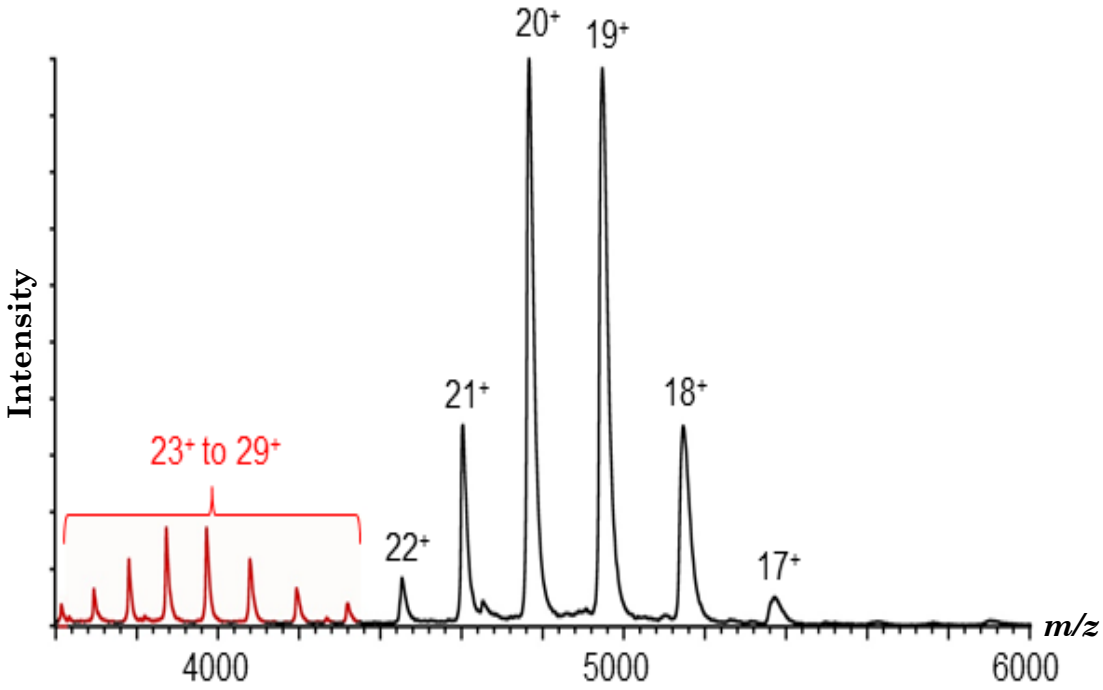


Figure 4.2 Native nESI-MS spectrum of C335A-USP5 (10 μM) sprayed from 200 mM ammonium acetate (pH 7). Each protein charge state is represented by $[M+zH]^+$ (where z is the number of charges), although z+ is indicated on the spectrum. The two charge state distributions of USP5 are clearly visible, with the lower distribution consisting of $[M+17H]^+$ to $[M+22H]^+$ and the higher distribution composed of $[M+23H]^+$ to $[M+29H]^+$ (ions highlighted in red) (Adapted from **[267]**).

As alluded to in *section 1.3.1*, multiple charging of proteins is an intrinsic feature of ESI. The number of charges (protons) attributed to a given protein can be directly correlated to the surface accessible region, and the number of basic side chains residues available for protonation. Accordingly, proteins adopt Gaussian-shaped charge state distributions (CSDs) which reflect the folded state of the protein. Studies by Kaltashov and colleagues have established that the average charge state exhibited by a protein population, sprayed under carefully optimised native conditions (neutral pH), correlates with the surface area of the protein [268, 269]. Thus, an interesting feature of the spectrum in Figure 4.2, is the presence of two distinct (bimodal) CSDs: one around *m/z*

5100, including $[M+17H]^+$ to $[M+22H]^+$, and a second around *m/z* 3700, including $[M+23H]^+$ to $[M+29H]^+$.

Bimodal distribution of CSDs has previously been observed when analysing His-tagged proteins, with removal of the His-tag negating this phenomenon. However in this case recombinant C335A-USP5 was expressed and purified as a GST-fusion, which was subsequently removed by thrombin cleavage prior to analysis. Earlier investigations by Konermann and Douglas [137] into the role of buffer pH on CSD, presented evidence that the appearance of bimodal CSDs indicates that at least two forms of protein molecules co-exist in solution. Accordingly as the instrument operating conditions were carefully optimised, with nESI capillary voltage < 1.3 kV, and the two observed CSDs reproducible, it can be concluded that two significantly different structural forms of USP5 arise from solution.

4.1.2 Ion mobility-mass spectrometry as a tool to examine USP5 conformers

To probe the differing structural forms of USP5 in relation to shape/size, or CCS, IM-MS was employed. The Waters Synapt HDMS, used in the investigations in this thesis, is equipped with a TWIMS cell, accordingly when in mobility mode the instrument reports on the ion drift time (t_d), as well as *m/z* and intensity. Instrument parameters were optimised (see Table 2.8); critically the TWIMS cell was operated at wave-height of 10 V, traveling at 300 m/s throughout IM-MS investigations in this thesis. The data acquired in mobility mode is viewed three dimensionally using the drift scope within the MassLynx software. In the integrated IM-MS spectrum produced, the *m/z* information is plotted against the drift time data, and the intensity is represented on this plot according to colour (as exemplified in Figure 4.3). In the integrated IM-MS spectrum of native C335A-USP5, the two CSD populations of are clearly visible and resolvable on a t_d axis. Accordingly we can conclude that broadly, two distinct USP5

structures or conformations persist in the gas-phase persist, which display distinct shape/size features.

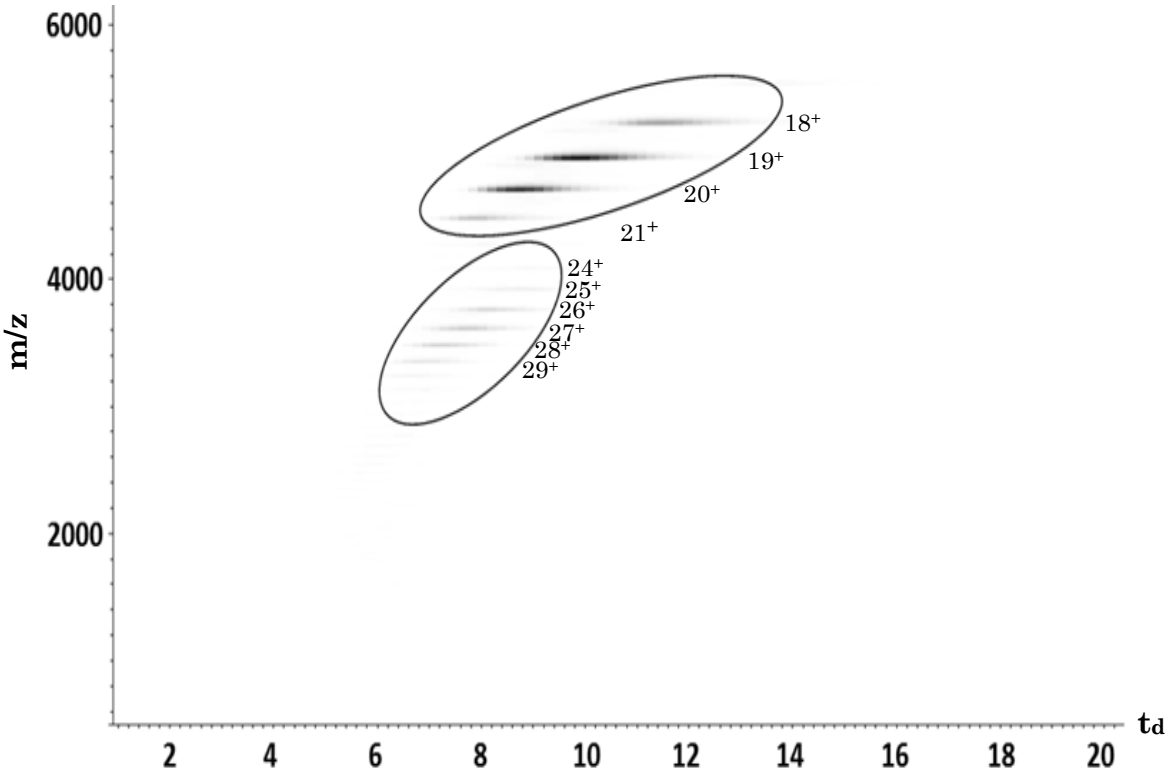


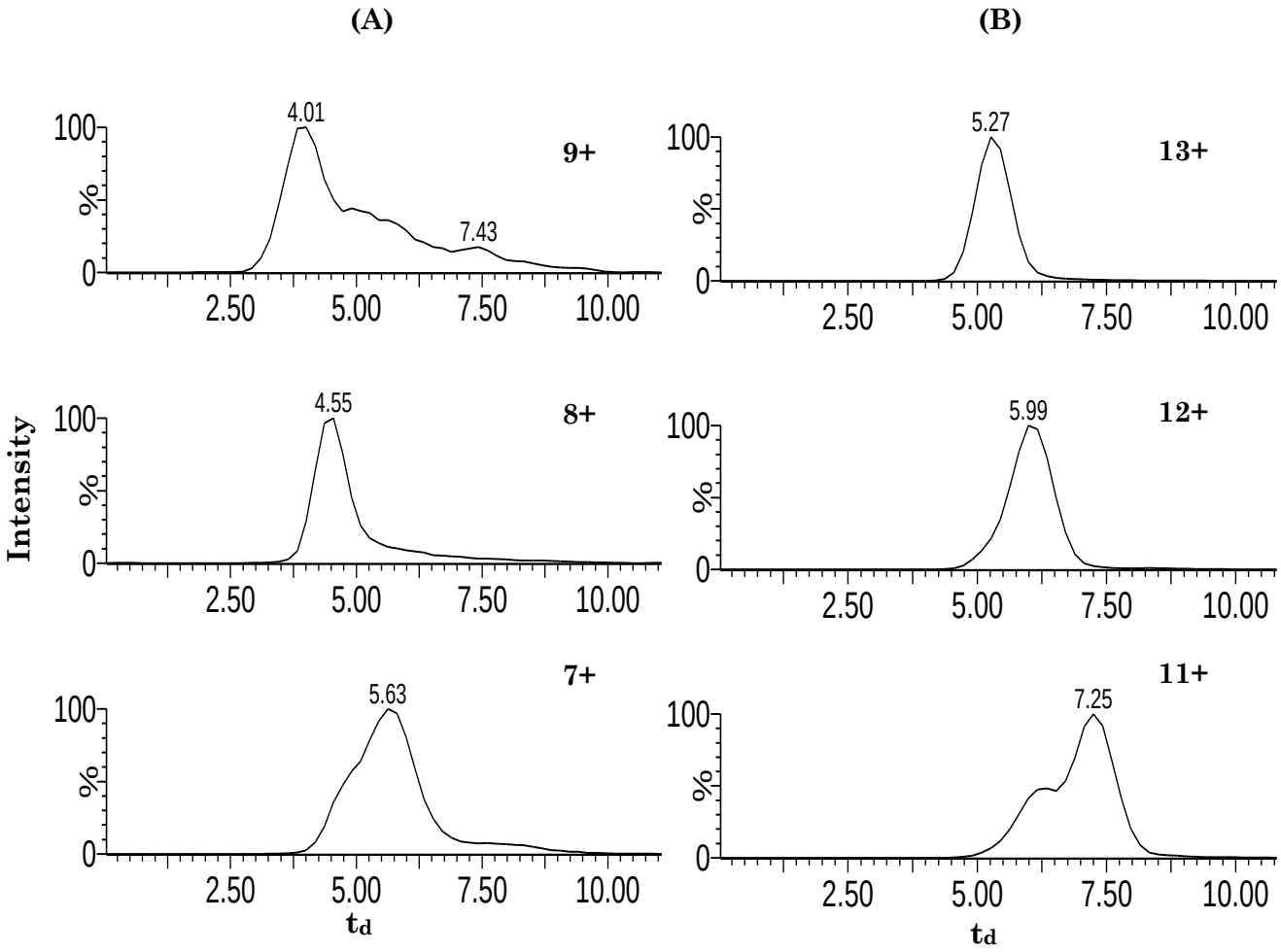
Figure 4.3 IM-MS plot of native C335A-USP5 (10 μM), sprayed from 200 mM ammonium acetate (pH 7). The *m/z* of a given ion (charge state indicated to right) is plotted as a function of drift time (t_d), with the intensity of the ions correlated to the darkness of the plot [267]. Consistent with native ESI-MS, IM-MS analysis indicates two resolvable species on the drift time axis, indicating that broadly two structures persist in the gas-phase (Adapted from **[267])**.

Owing to the complex nature of the (time-varying) electric field employed by the TWIMS device, protein CCS cannot be directly determined from drift time. Instead it is necessary to produce a calibration curve (an approach which has been well described [167]) using a series of protein standards, with a range of masses which bracket the protein of interest. In this case, as C335A-USP5 is ~94 kDa, β-lactoglobulin (18 kDa monomer and 36 kDa dimer), BSA (69 kDa) and ADH (143 kDa) were selected as

calibrants. Identical operating conditions to those used to measure the drift time of C335A-USP5 were utilised to measure the protein standards.

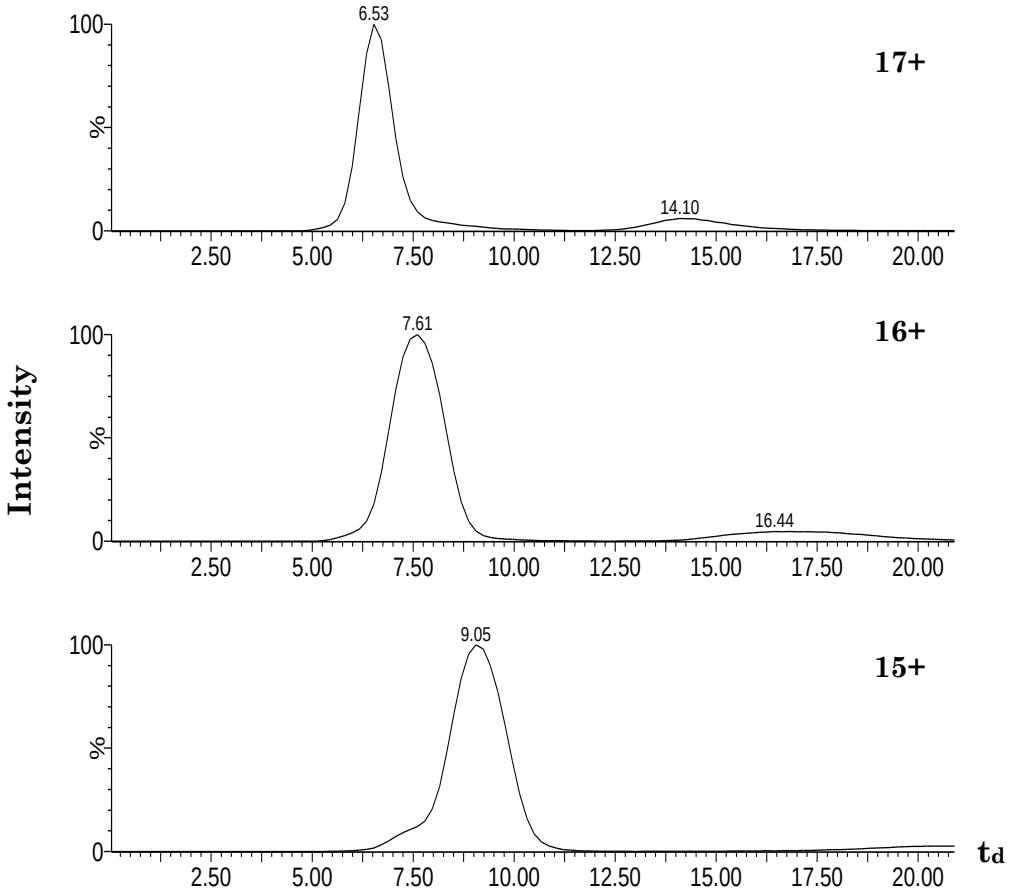
From the IM-MS plots (exemplified in Figure 4.3), all data was transferred and processed in MassLynx software. From the resulting TWIMS drift traces for each charge state (for each calibrant), ion drift time (t_d) can be extracted. The TWIMS drift traces for each charge state, of each calibrant, is highlighted in Figures 4.4-4.6.

The calibrant t_d is converted to t_d' as described in *section 2.4.3.5*, and plotted against the database CCS value indicated by Bush *et al* [232] in a log-log plot, termed the calibration plot (see Figure 4.7). The slope of the line of the calibration plot, in this case 0.2848, subsequently serves as a correction factor, which is applied to correct t_d' to yield a final corrected calibrant drift time (t_d''). This t_d'' value is finally plotted versus literature CCS values for the given ion in question (taken from Bush *et al* [232]) to yield a mobility conversion graph, which permits the conversion of measured t_d of the analyte, C335A-USP5, to CCS (which we will discuss below). The mobility conversion plot obtained (see Figure 4.11) displayed excellent linearity (R^2=0.998).



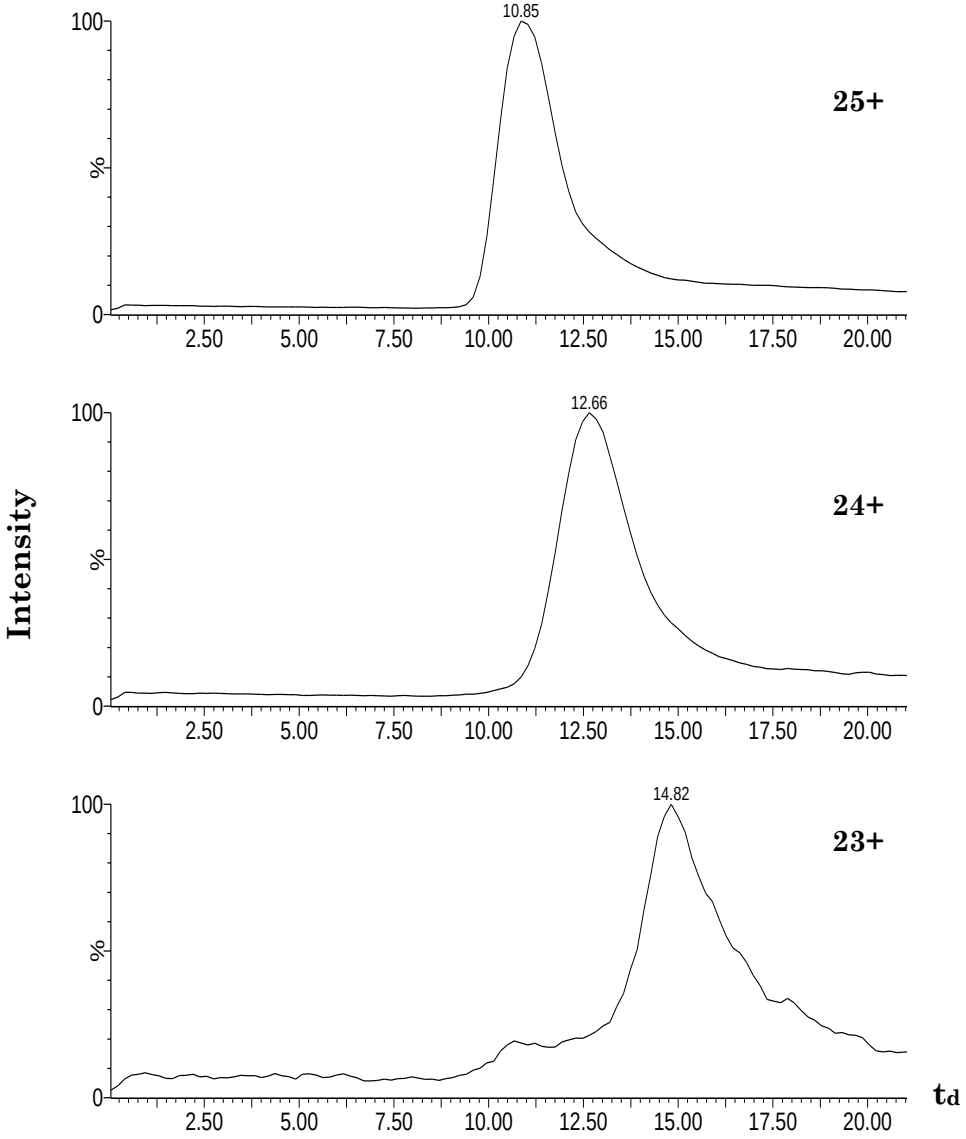
	β-lactoglobulin					
	Monomer			Dimer		
Charge	7	8	9	11	12	13
t_d (ms)	5.63	4.55	4.01	7.25	5.99	5.27
Database cross section (nm^2)	16.6	16.9	17.8	28.5	29	29.6
t_d''	2.100183	2.242225	2.420754	3.569292	3.668921	3.816966

Figure 4.4 TWIMS drift traces for (A) monomeric and (B) dimeric β-lactoglobulin, highlighting the ($[M+7H]^+$ to $[M+9H]^+$ and ($[M+11H]^+$ to $[M+13H]^+$) charge states used in the generation of the calibration plot. The apparent shoulder on the $[M+11H]^+$ drift trace in (B) is due to a high baseline in this region of the spectrum (Adapted from **[267])**.



	BSA		
Charge	15	16	17
t_d (ms)	9.05	7.61	6.53
Database cross section (nm^2)	41	40.6	40.4
t_d''	5.22453	5.283208	5.3517

Figure 4.5 TWIMS drift traces for BSA, highlighting the ($[M+15H]^+$ to $[M+17H]^+$) charge states used in the generation of the calibration plot. The apparent shoulder on the $[M+15H]^+$ drift trace is due to a high baseline in this region of the spectrum (Adapted from **[267]).**



	ADH		
Charge	23	24	25
t_d (ms)	14.82	12.66	10.85
Database cross section (nm^2)	69.4	69.4	68.3
t_D''	9.255017	9.214666	9.164202

Figure 4.6 TWIMS drift traces for ADH highlighting the ($[M+23H]^+$ to $[M+25H]^+$) charge states used in the generation of the calibration plot (Adapted from **[267]).**

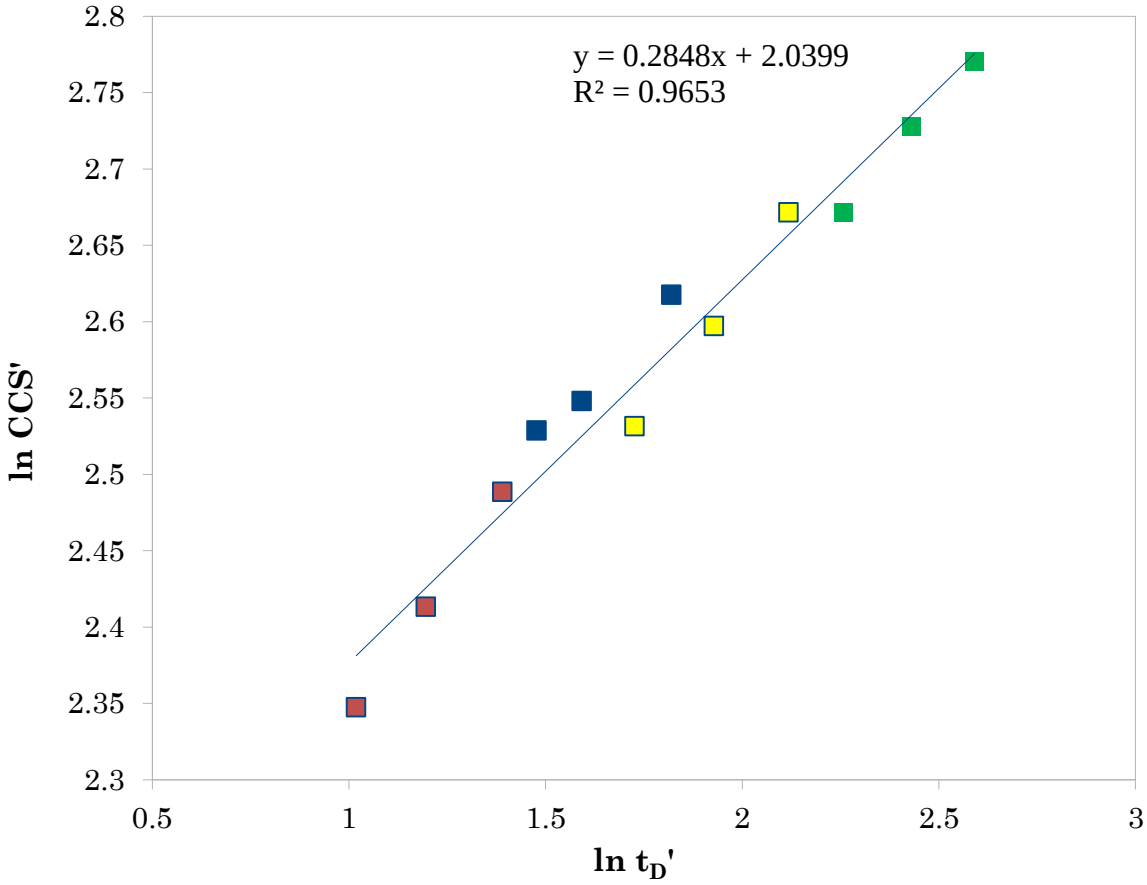
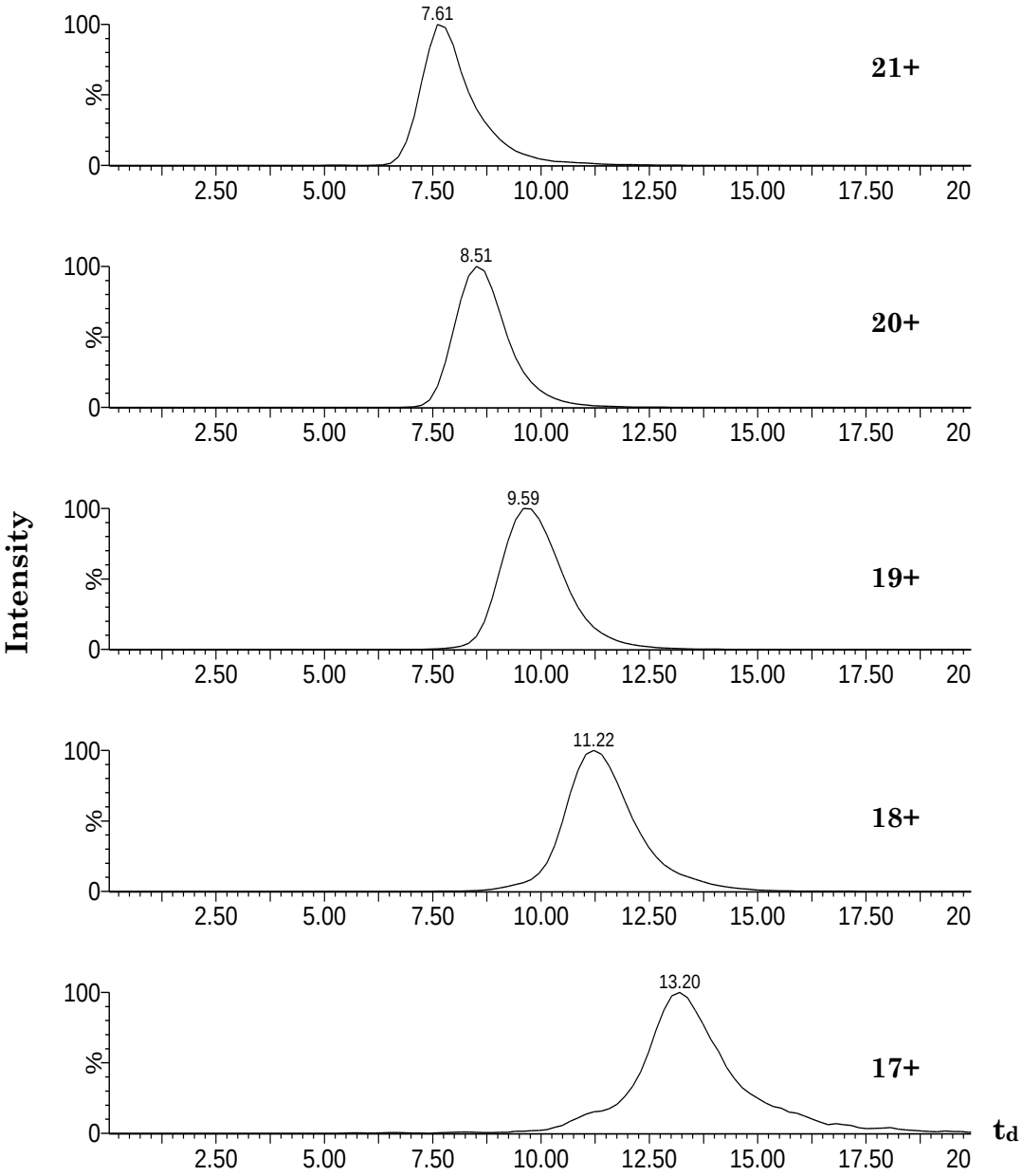


Figure 4.7 IM-MS calibration plot. This initial calibration plot is utilised to determine the correction factor between the corrected CCS and corrected drift time. Each calibrant is indicated with squares; β-lactoglobulin monomer (red), β-lactoglobulin dimer (blue), BSA (yellow), ADH (green) (Adapted from **[267]**).

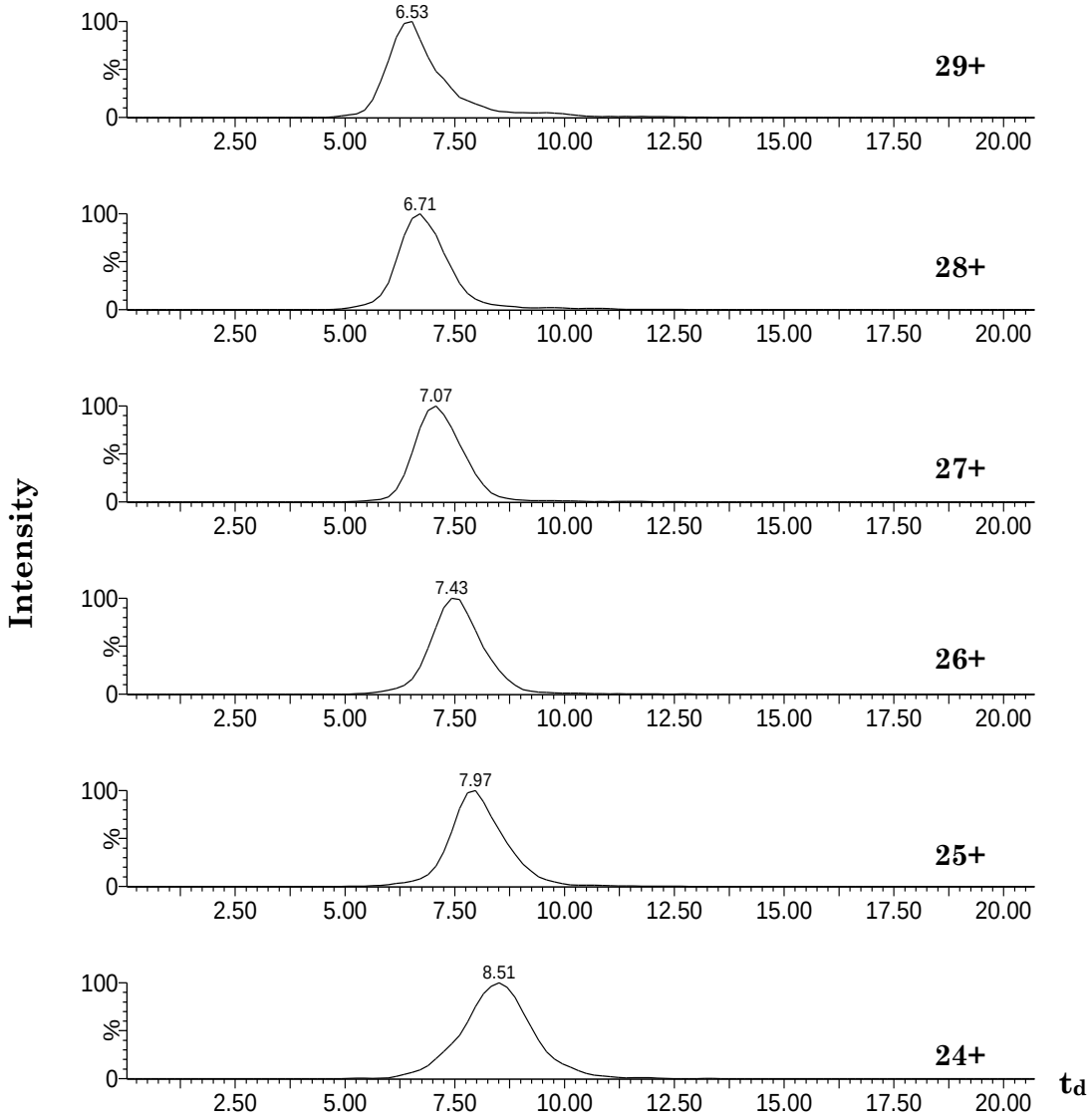
Subsequently, the drift time for each of the charge states in the two CSDs of USP5 ($[M+17H]^+$ to $[M+22H]^+$, and $[M+23H]^+$ to $[M+29H]^+$) was extracted (Figures 4.8-4.9), corrected, and plotted onto the mobility conversion graph, to yield CCS information (Figure 4.10). To permit direct comparison between the two CSDs of USP5, each charge state was plotted on a CCS axis (see Figure 4.11).

The two CSDs noted in Figures 4.2 and 4.3, formed two clearly resolved groups on the CCS axis. The lower CSD composed of $[M+17H]^+$ to $[M+21H]^+$, gave a mean CCS value of 52 ± 4 nm^2, whilst the higher CSD, composed of $[M+24H]^+$ to $[M+29H]^+$, possessed a mean CCS value of 66 ± 6 nm^2. IM-MS analysis thus supports the notion that two significantly different USP5 structures, in relation to average size/shape, are present, this includes compact structure (lower CSD, CCS ~52 nm^2) and extended (higher CSD, CCS ~66 nm^2) structure. The intermediary charge states ($[M+22H]^+$ and $[M+23H]^+$), were poorly resolved on a drift time axis, indicating the presence of both compact and extended conformers at these charge states (data not shown).



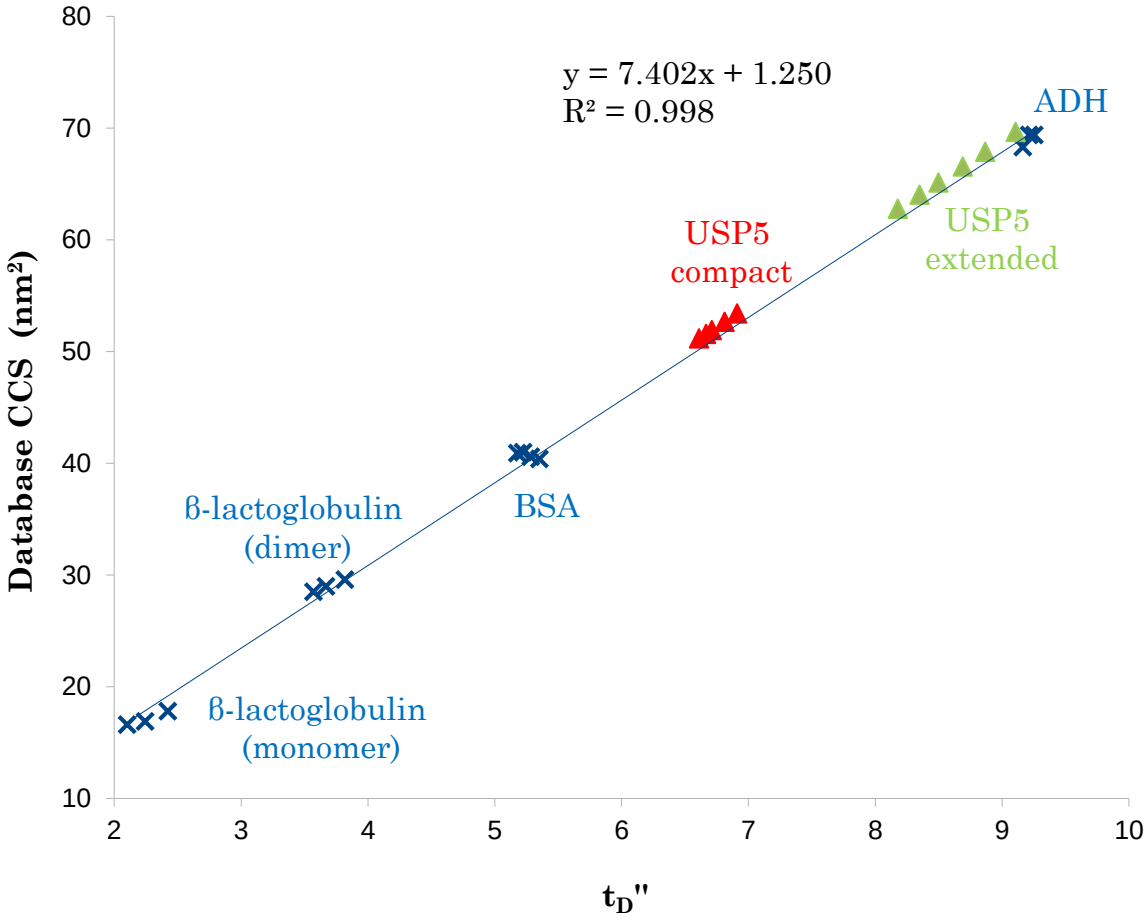
	USP5 (Low charge state population)				
Charge	17	18	19	20	21
t_d (ms)	13.2	11.22	9.59	8.51	7.61

Figure 4.8 TWIMS drift traces for USP5 for the lower charge state ($[M+17H]^+$ to $[M+21H]^+$) distribution, evidenced in Figure 4.3. The measured collisional cross section for this distribution is noted in the table beneath (Adapted from **[267]**).



	USP5 (High charge state population)					
Charge	24	25	26	27	28	29
t_d (ms)	8.51	7.97	7.43	7.07	6.71	6.53

Figure 4.9 TWIMS drift traces for USP5 for the high charge state ($[M+24H]^+$ to $[M+29H]^+$) distribution, evidenced in Figure 4.3. The measured collisional cross section for this distribution is noted in the table beneath (Adapted from **[267]**).



Charge	17	18	19	20	21	
Measured CCS (nm^2)	51.2	51.6	51.9	52.7	53.4	
Charge	**24**	**25**	**26**	**27**	**28**	**29**
Measured CCS (nm^2)	62.8	64.1	65.2	66.6	67.9	69.7

Figure 4.10 Mobility conversion graph. The t_d'' for each calibrant (blue) plotted against the pre-determined database collisional cross sections. The derived t_d'' values for the compact (red) and extended (green) forms of USP5 are plotted, to determine theoretical CCS (R^2 value based on calibrants only). The table beneath indicates the CCS determined for each USP5 species (Adapted from **[267]).**

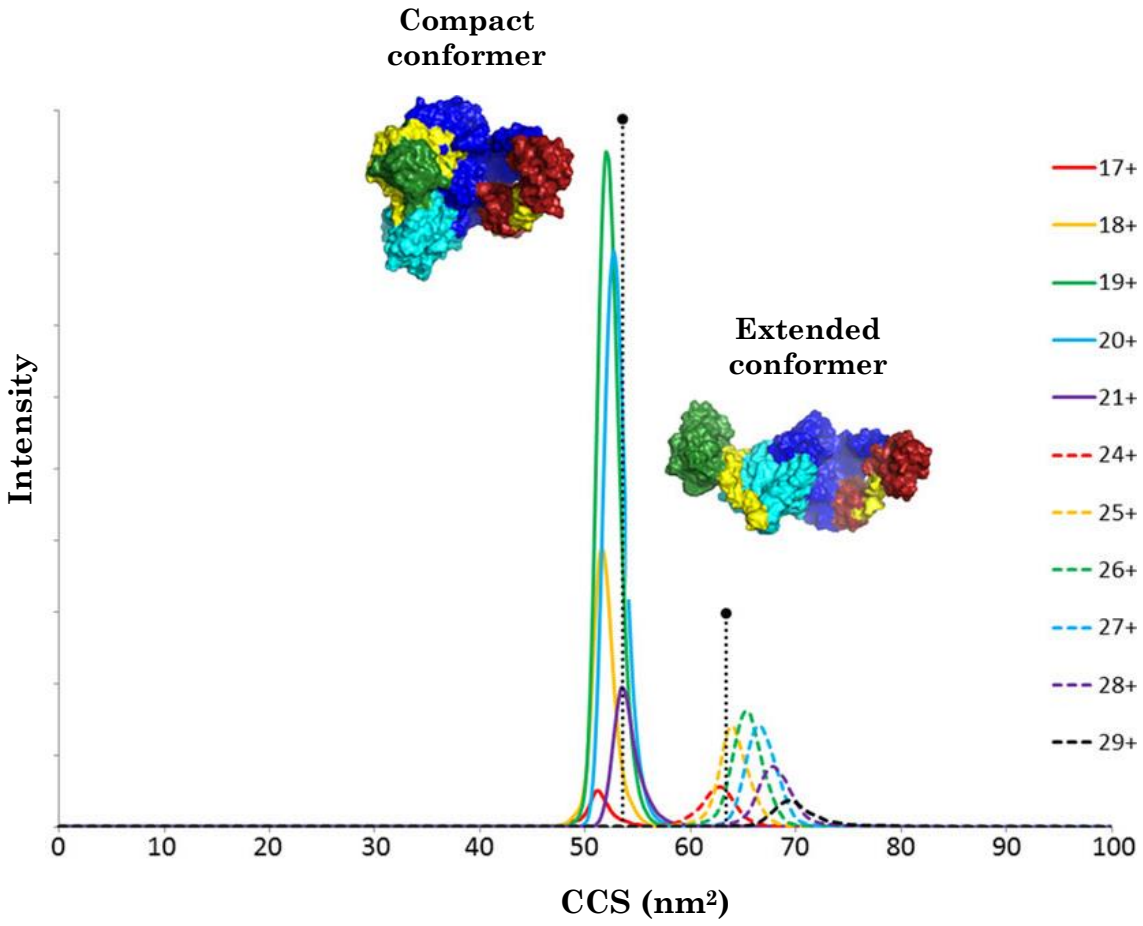


Figure 4.11 IM–MS drift trace for USP5 plotted on a CCS axis, showing the two distinct distributions of conformers by charge state. Markers for 53.5 and 64 nm^2, corresponding to the compact and extended outputs of gas-phase MD simulation (discussed in section 4.1.3), are indicated. The final output structures, corresponding to the compact and extended structure of the 2-ns gas-phase MD simulations are illustrated (ZnF-UBP domain (green), catalytic domain (navy blue), cryptic UBP domain (light blue) and UBA domains (red)) (Adapted from **[267]).**

4.1.3 Computational interrogation of USP5 structures

To validate the experimental IM-MS findings and provide insight into the structures of USP5 in the gas phase, MD simulations were completed using the Amber99B force field (see *section 2.5.1*, submitted by Dr Neil Oldham). As alluded to in *section 1.3.2*, MD simulations utilise high resolution atomic co-ordinates to assess the accessible conformational landscape of a protein. The output of MD simulations, a series of structures, can be utilised to determine CCS, thus permitting comparison between experimental CCS determined by IM-MS, and theoretical CCS derived from high resolution structures.

Starting with the crystal structure of USP5 (PDB code 3IHP), and following modelling-in of 151 missing residues, the energy minimised structure of USP5 was submitted to 2 ns of MD in the gas-phase, at 300 K, using the Amber Sander parameters. 2 ns of MD was completed as this is towards the limit of the available tools with regards to computational expense. The 1000 output trajectories from this MD simulation were visualized in VMD software and every fifth structure (sampling rate of 1 structure every 10 ps) exported to a PDB file (to keep the file size manageable). The PA method was used to calculate the CCS of each structure in the DriftScope 2.1 CCSCalc program. Subsequently the PA CCS was corrected, by multiplying by the empirical scaling factor, 1.14, to yield Trajectory Method-like CCS values [169].

A plot of corrected CCS versus time was generated to assess the changes in structure USP5 over the course of the 2 ns MD run (see

Figure 4.12). A rapid contraction of the structure from an initial CCS of 63.4 nm^2 to approximately 55 nm^2, representing a reduction of ~13 %, was apparent over the first 200 ps of the simulation, which supports the notion that proteins undergo sidechain collapse following desolvation, leading to compaction. For the remaining 1.8 ns of the MD run, a steady decrease to a CCS of approximately 53-54 nm^2 is reported. To draw

comparison with experimental CCS, a marker of 53.5 nm^2 (and the final MD output structure) was plotted onto Figure 4.11. Reassuringly this is close to the range of CCSs exhibited by the low CSD of USP5 (within ca. 5 %). Thus it appears that the low charge states of USP5 undergo only a small degree of additional collapse over that seen theoretically on the 2 ns timescale.

To estimate the theoretical CCS of the more extended conformation of USP5, corresponding to the higher CSD, the energy-minimized structure used in initial MD simulations was manually adjusted (extended) using Pymol, guided by previous SAXS data [237]. The resulting structure was then relaxed and submitted to gas-phase MD (2 ns) as above. The structure was seen to collapse over the 2 ns time frame from 74 nm^2 to 64 nm^2, which reassuringly is again close to the range of CCSs exhibited by the high CSD of USP5 (within ca. 5 %). A marker is again indicated on Figure 4.11, alongside the final output structure.

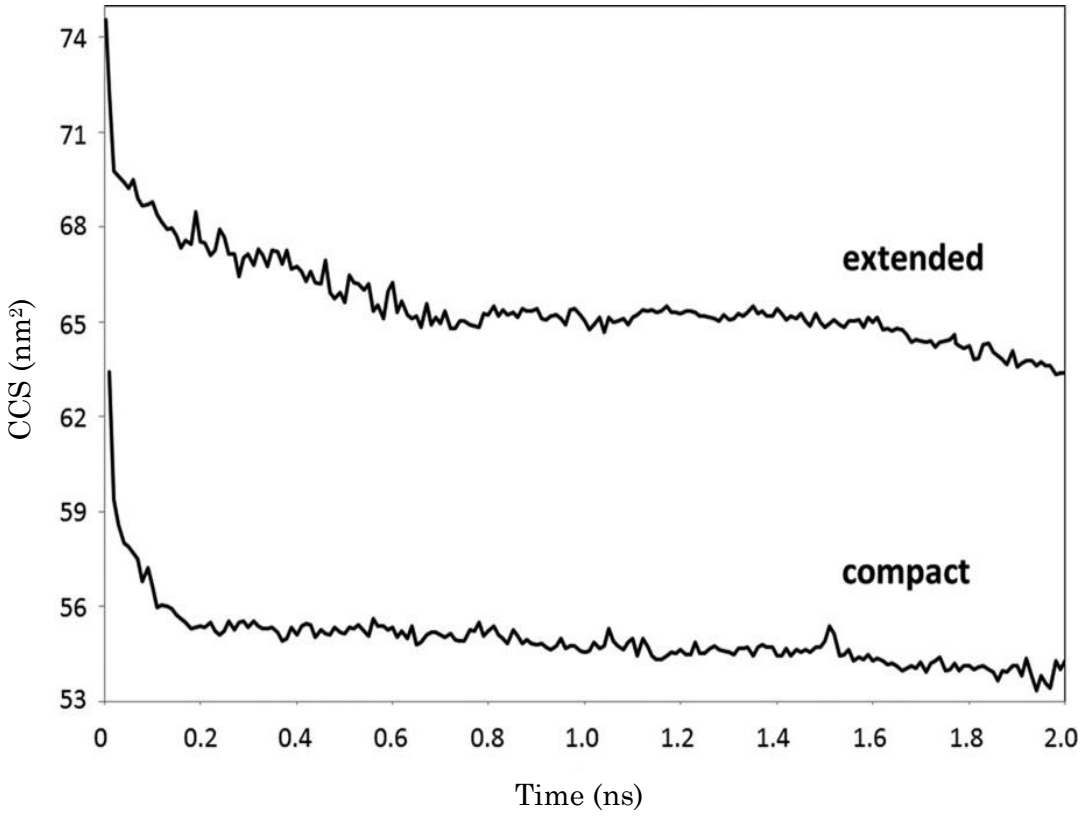


Figure 4.12 Plot of CCS as a function of time, for the gas-phase MD simulation of the compact and extended model structures of USP5, showing the collapse of each species

over 2 ns. Following initial compaction, likely owing desolvation, the structures adopt a steady CCS (Adapted from [267]**).**

4.1.4 Conclusions from native ESI-IM-MS analysis and MD simulations

By utilizing native IM-MS and CSD analysis, which affords the 'visualization' of individual protein conformers, it can be concluded that an ensemble of USP5 conformers persist in the gas phase. Furthermore, by applying MD simulations, the USP5 conformers, indicated experimentally, can be related to the recent USP5 X-ray structure. Indeed, the experimental conformers broadly correspond to a compact conformer (Low CSD, CCS ~52 nm^2), similar in cross-section to the gas-phase model of the X-ray structure (CCS ~53.5 nm^2) which dominates, as well as a minor population of extended conformer (CCS ~66 nm^2). Reassuringly the IM-MS findings demonstrate striking similarities to previous SAXS analysis (performed in solution, under reducing conditions), which alludes to the presence of compact and extended conformers [237]. For example, both IM-MS and SAXS analyses indicate a similar conformer ratio (compact: extended); 3.5:1 by IM–MS (based on the relative intensity of the two distributions, Figure 4.2), and 3:1 by SAXS (based on the three conformers per ensemble representation in [237]). Also the relative sizes (in units of area) of the two conformations are also similar when measured by IM–MS (CCS: 66:52 nm^2 = 1.3) and SAXS (radius of gyration, Rg^2: 20:14 nm^2 = 1.4, again using the data for three conformers per ensemble representation in [237]). While IM-MS does not report a structure that is significantly more compact than the X-ray structure, and thus may permit binding at the ZnF-UBP domain and the catalytic site, this may reflect the inherently low resolution structural information afforded by IM-MS, or alternately such a conformational shift may arise following ligand binding, as is noted by other USP family members [258, 259].

Given the large and flexible nature of the USP5 structure, significant collapse was expected upon desolvation, indeed when considering the theoretical models of USP5

structure in the gas phase, CCS decreases from 63.4 nm^2 to approximately 55 nm^2, representing a reduction of ~13 %. Thus it is noteworthy that despite such collapse, predicted by MD simulations, valuable structural information comparable to the X-ray structure and SAXS analysis, can be obtained from native IM-MS. However questions remain regarding whether such a structure retains native function, which we will now go on to address in *section 4.2*.

4.2 Exploring USP5 structure with respect to (poly)ubiquitin recognition

To probe the significance of the USP5 conformers, reported in *section 4.1*, with respect to ubiquitin recognition, and to confirm the sites of (poly)ubiquitin recognition implied by previous biophysical studies [87, 237], USP5 interactions with various ubiquitin substrates were monitored by native MS interaction studies and carbene footprinting, the results of which, are detailed in *section 4.2*.

4.2.1 Expression and purification of USP5 and ubiquitin mutants

Prior to native MS USP5-ubiquitin interaction studies, point mutations were introduced to either the ZnF-UBP domain (R221A), or the two UBA domains (M643E/M711E) of C335A-USP5. These mutations had previously been identified as key sites for ubiquitin recognition by USP5 [87]. Point mutations were confirmed by DNA sequencing (see Figure 4.13), and mutant proteins were expressed and purified in the same manner as C335A-USP5.

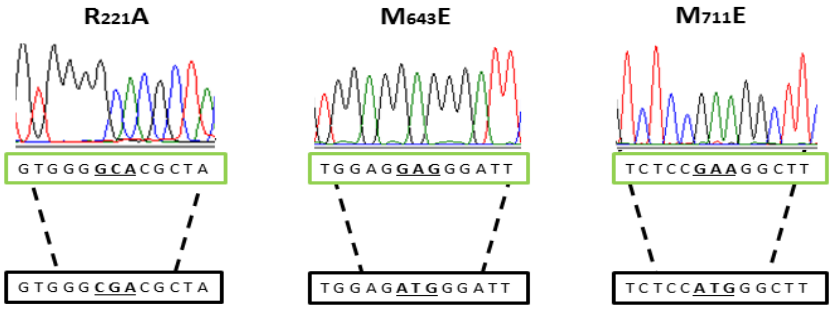


Figure 4.13 DNA sequencing of C335A-USP5 point mutations. Mutation is indicated above the respective electropherogram.

To probe the role of the critical C-terminal region of monoubiquitin in interactions with C335A-USP5, WT and a C-terminal deletion (ΔGly75-Gly76) monoubiquitin protein were expressed and purified as described in *section 2.2*. Briefly the cDNA sequence encoding the 76 residue human monoubiquitin protein was previously cloned into the *Eco*RI-*Hind*III restriction sites of pkk223-3, which permits expression as an untagged protein. Both WT and the mutated ΔGly75-Gly76 constructs were provided at the outset of the studies (verified by DNA sequencing). The plasmid constructs were subsequently transformed and expressed in *E. coli*, specifically using the C41 bacterial strain, suitable for the expression of toxic proteins. Following cell lysis, the majority of impurities were first removed by lowering the pH of the purification buffer to pH 4, which results in the precipitation of the majority of proteins, excluding the highly stable ubiquitin. Subsequently the ubiquitin protein was purified by cationic exchange chromatography, and then gel filtration, prior to being subject to SDS-PAGE analysis, which resulted in a single band indicated in Figure 4.14A. 1 L of culture typically yielded 10 mg of purified protein. The mass of WT and ΔGly75-Gly76 was confirmed by MS analysis on the Synapt HDMS to be 8565 and 8450 Da (expected 8565 and 8450 da), respectively, as illustrated in Figure 4.14B. All USP5 and monoubiquitin constructs were buffer exchanged into 200 mM ammonium acetate (pH 7) as per *section 4.1.1*, prior to native MS interrogation.

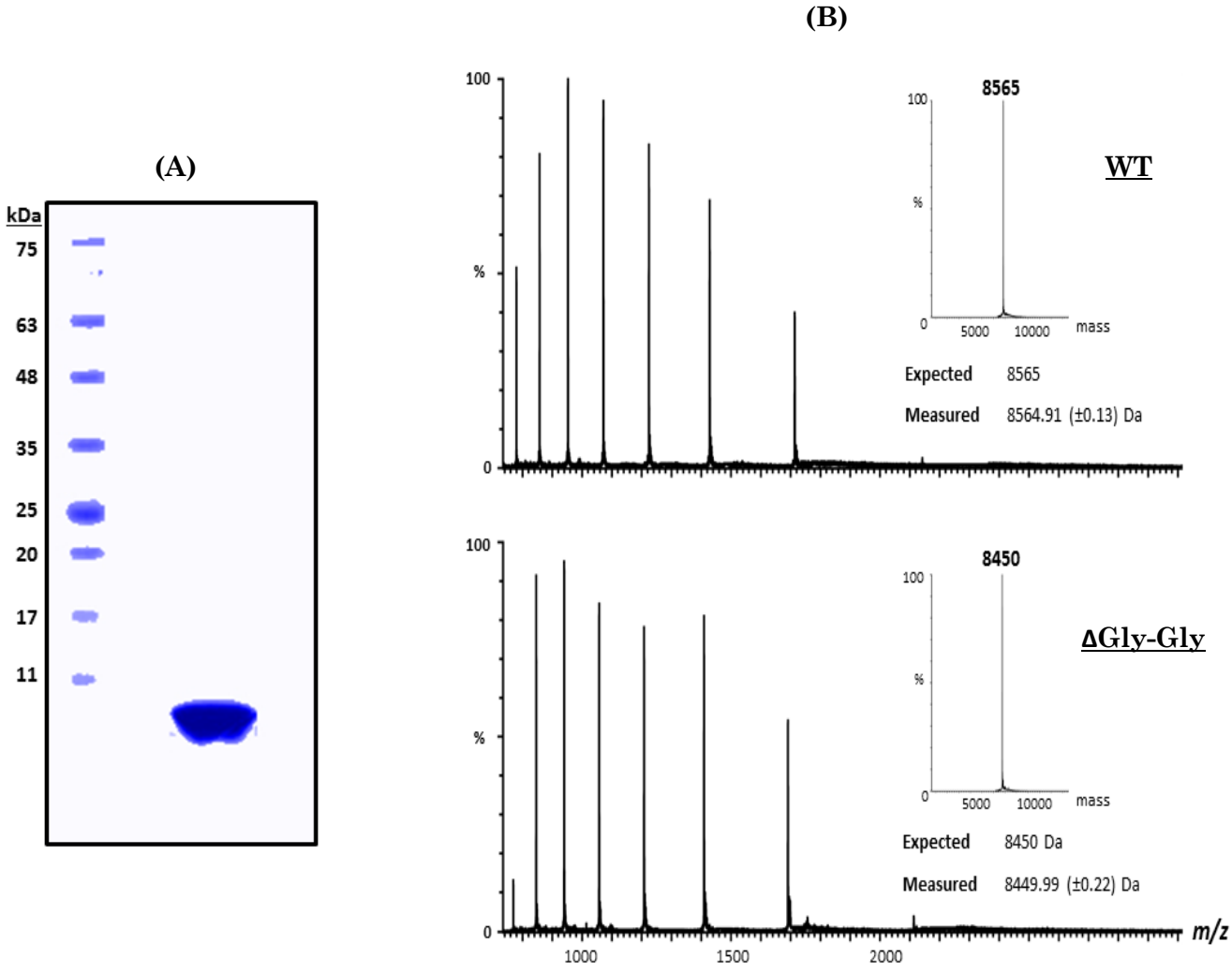


Figure 4.14 Expression and purification of WT and ΔGly75-Gly76 monoubiquitin. (A) A representative Coomassie stained (5-20%) SDS-polyacrylamide gel of purified monoubiquitin (in this case WT). (B) Full and deconvoluted ESI-MS spectrum of denatured monoubiquitin (Top, WT and bottom, ΔGly75-Gly76 mutant monoubiquitin), sprayed from 80 % MeCN, 0.1 % formic acid (analysis completed on the Synapt HDMS).

4.2.2 Probing USP5-monoubiquitin interactions

Native nESI-MS USP5-monoubiquitin interaction studies were completed under the established optimum instrument settings for USP5 transmission (see *section 4.1.1*) on the Waters Synapt HDMS. USP5-ubiquitin titrations were prepared as 1:1 (v/v) mixtures and loaded in the same manner as *section 4.1.1* into nESI capillaries.

4.2.2.1 C335A-USP5 vs WT monoubiquitin

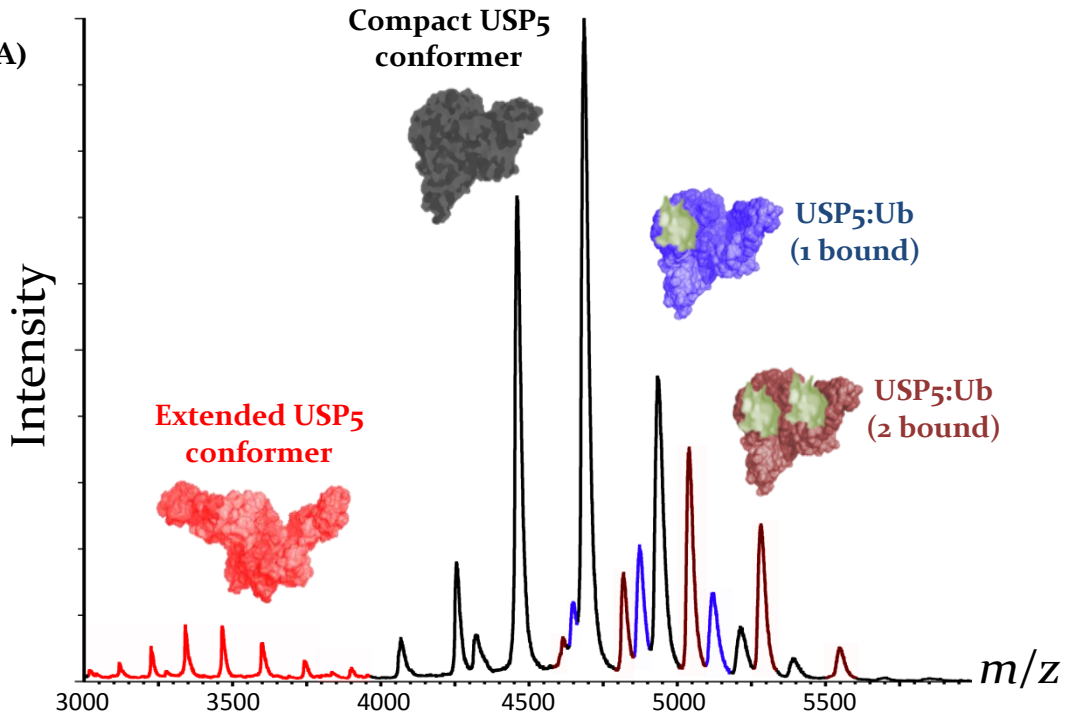
Firstly, C335A-USP5 (10 μM) was mixed with WT monoubiquitin (2 μM) (as a 1:1 v/v mixture) and analysed by ESI-MS. The resulting spectrum, indicated in Figure 4.15, uniquely illustrates the advantages of native ESI-MS and is informative with regards to USP5-monoubiquitin interactions. Firstly, as MS analysis results in the attribution of discrete *m/z* signals to unique species it is possible to identify the stoichiometry of protein-ligand binding. Accordingly in Figure 4.15 we can determine that C335A-USP5 simultaneously binds to one and two ubiquitin monomers. It is also apparent that binding only occurs to the lower CSD of USP5, which from IM-MS and MD simulations in *section 4.1* we know is the compact conformer of USP5, similar in CCS to the gas-phase model of the recent X-ray structure. Furthermore we can conclude this binding does not seem to alter the USP5 conformer equilibrium (based on signal intensities).

It is also evident that the signal intensity of the second ubiquitin binding event is greater than the first (Figure 4.15, signals highlighted in brown vs blue). Crucially, by carefully monitoring such free and bound signal intensities corresponding to USP5 (or indeed ubiquitin) it is possible to quantitate the binding affinity of monoubiquitin for C335A-USP5. Indeed this approach has previously been applied to quantitate the binding affinity for alternative isolated UBD: ubiquitin interactions, in a rapid and sensitive

manner and been found to be in good agreement with binding affinities determined by other biophysical techniques [198]. To ensure reproducibility, a titration of C335A-USP5 versus WT monoubiquitin was completed, where the concentration of C335A-USP5 was kept constant (10 μM), whilst the concentration of WT monoubiquitin was varied (2, 3 and 4 μM). The resulting solutions were then subjected to ESI-MS analysis and the signal intensities from the mass spectra attributed to one ubiquitin bound-, two ubiquitin bound- and unbound-USP5, were recorded. The apparent dissociation constants (K_d) were then calculated as outlined in *section 2.4.3.4.*

An example of the calculation is shown in Figure 4.15. For simplicity the example shows the results obtained for one measurement (to demonstrate reproducibility across multiple concentrations), however a minimum of three replicates were performed per titre range. As highlighted in Figure 4.15, the apparent approximate K_d value for the first binding event of WT monoubiquitin to C335A-USP5 was 2.5 (±0.3) μM, consistent with binding at ZnF-UBP domain [84]. The K_d value for the second binding event was too low to be quantified using the concentrations of binding partners required for detection by MS, indicating a high affinity (nM) interaction for the second ubiquitin moiety. This large increase in affinity between the first and second binding event indicates strong positive cooperativity in the binding of monoubiquitin by USP5.

(A)



(B)

$[USP5]_0$ (µM)	10		
$[Ub]_0$ (µM)	2	3	4
Σ I(Free USP5)	274	246	292
Σ I(USP5.Ub(1))	50	74	120
Ratio (R)	5.5	3.3	2.4
K_d (µM)	2.5	2.3	2.6

Average K_d (µM)	2.5 ± 0.2

Figure 4.15 Probing monoubiquitin recognition by C335A-USP5. (A) nESI-MS spectrum of C335A-USP5 (10 µM) in complex with monoubiquitin (2 µM). Highlighted are the extended USP5 conformer (red), compact USP5 conformer (black), and the compact USP5:monoubiquitin complex (1 ubiquitin moiety bound in blue, 2 ubiquitin moieties bound in brown). (B) A table to illustrate the extracted intensities for sum of unbound USP5 (Σ I(Free USP5)) and USP5-monoubiquitin complex Σ I(USP5.Ub(1))), from which the K_d is determined. The data presented is from one of three replicates per ubiquitin concentration (Adapted from [270]).

4.2.2.3 C335A-USP5 Vs ΔGly75-Gly76 monoubiquitin

To further investigate the positive cooperativity displayed by C335A-USP5, a titration, performed in the same manner as Figure 4.15, albeit versus a ΔGly75-Gly76 mutant monoubiquitin protein, which abrogates binding at the ZnF-UBP domain (as previously indicated [198]). The resulting spectrum is illustrated in Figure 4.16. Only one ubiquitin moiety bound C335A-USP5, with an apparent approximate K_d value of 26.7 (±0.8) μM. As the affinity decreases from ~low nanomolar, when C335A-USP5 is titrated versus WT monoubiquitin (thus the ZnF-UBP domain is occupied), to 26.7 (±0.8) μM when the catalytic domain is occupied in isolation by ΔGly75-Gly76 ubiquitin, we can conclude that the ZnF-UBP domain is required for optimal interaction at the catalytic site.

4.2.2.2 C335A-USP5 mutants vs WT monoubiquitin

To validate the findings with mutant monoubiquitin, mutant C335A-USP5 was subsequently probed versus WT monoubiquitin. Mutation of both of the UBA domains did not alter recognition of two ubiquitin moieties, confirming the role of these domains in the interaction with longer polyubiquitin chains (Figure 4.17A). Mutation of the ZnF-UBP domain of C335A-USP5, resulted in the binding of one monoubiquitin moiety (presumably at the active site) (Figure 4.17B), reproducing the findings of titration with ΔGly75-Gly76 mutant monoubiquitin protein.

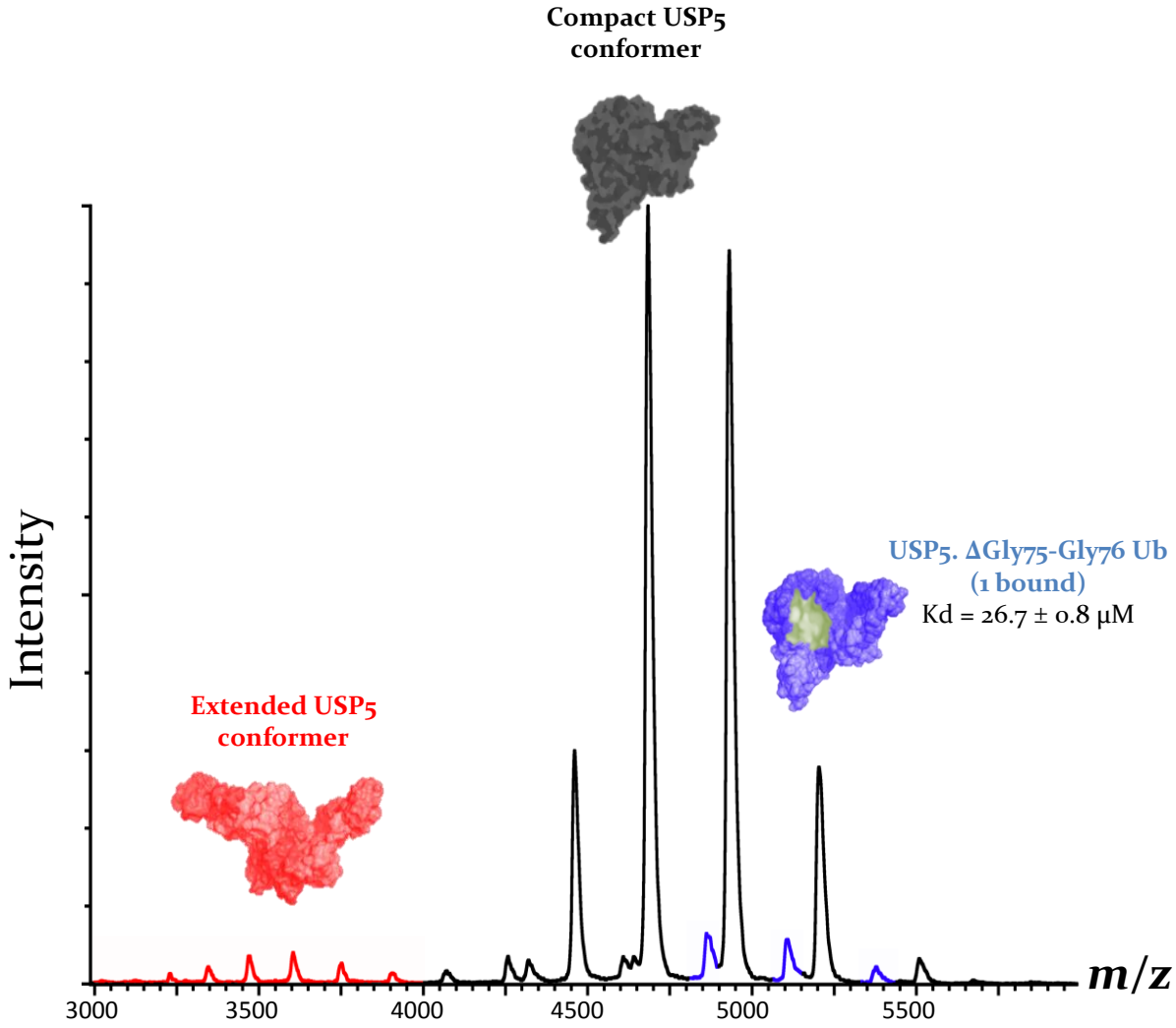


Figure 4.16 nESI-MS spectrum of C335A-USP5 (10 μM) in complex with ΔGly75-Gly76 mutant monoubiquitin (2 μM). Extended USP5 conformer (red), compact USP5 conformer (black) and the USP5:monoubiquitin complex (1 ubiquitin moiety bound in blue). K_d was determined by titrating C335A-USP5 (10 μM) against multiple replicates of 2, 3 and 4 μM ΔGly75-Gly76 mutant monoubiquitin (Adapted from [270]).

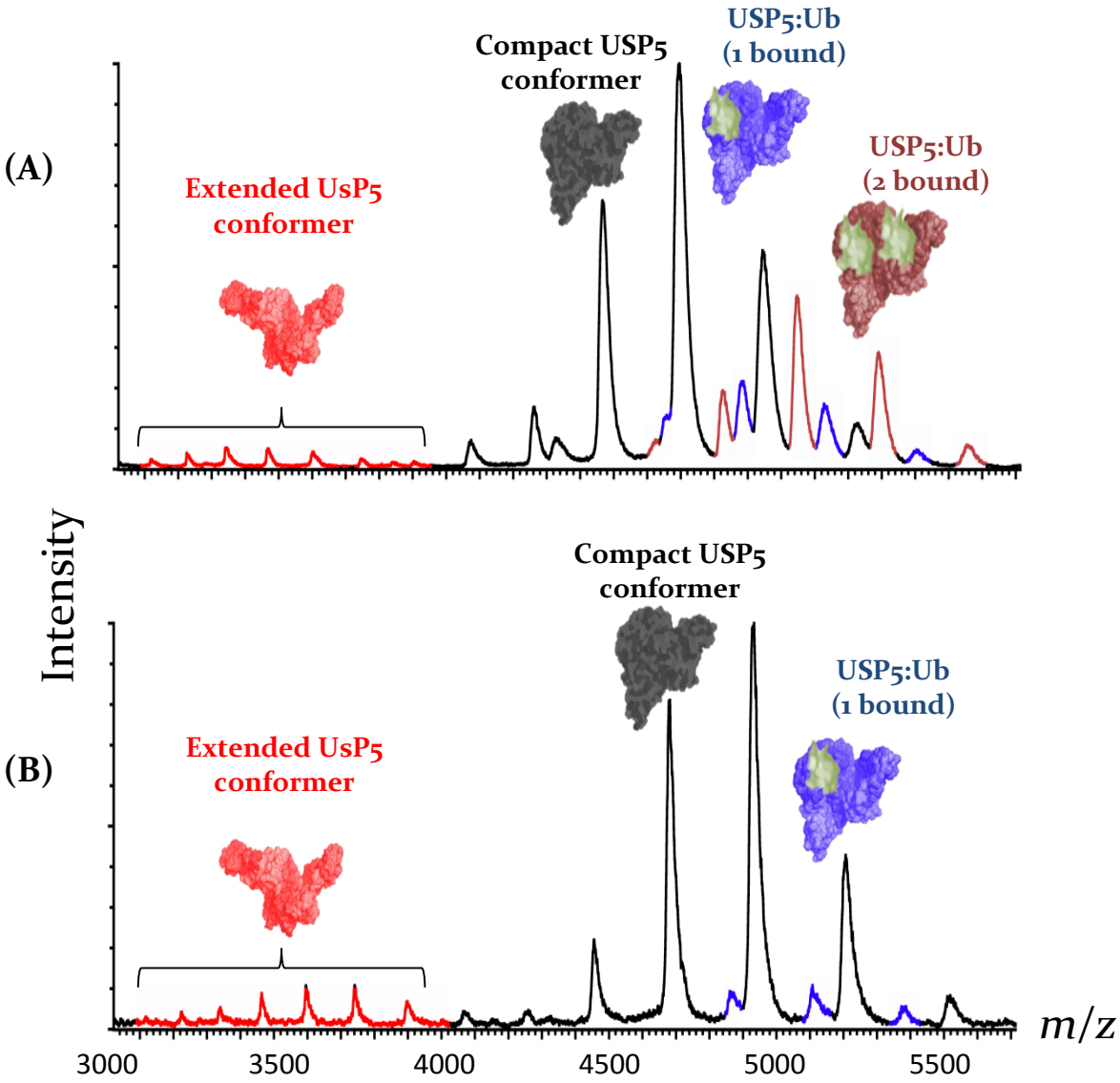


Figure 4.17 nESI-MS spectra of (A) UBA (Met-643-Glu/Met-711-Glu) or (B) ZnF-UBP (Arg-221-Ala) mutant C335A-USP5 (10 μM) in complex with WT monoubiquitin (2 μM). Highlighted are the extended USP5 conformer (red), compact USP5 conformer (black), and the compact USP5-monoubiquitin complex (1 ubiquitin moiety bound in blue, 2 ubiquitin moieties bound in brown) (Adapted from [270]).

4.2.2.5 Conclusions

To probe the significance of USP5 conformational dynamics, native ESI-MS was employed in tandem with protein mutagenesis to monitor USP5-ubiquitin interactions. Titration of C335A-USP5 versus WT monoubiquitin highlighted a clear binding stoichiometry of 1:1 and 1:2. For the first binding event the K_d was determined to be ~2.5 μM. Despite the second binding event proving too strong to quantify by MS, based on previous experience in the Oldham group and MS signal intensities, it is possible to place an upper limit on the associated K_d of 100 nM. This large (>twenty-fold) increase in affinity between the first and second binding event indicates strong positive cooperativity in the binding of monoubiquitin to USP5. Reyes-Turcu *et al* had briefly alluded to this cooperativity utilizing ITC [87], however the results presented here highlight the advantage of the native MS approach, which permits clear 'visualization' of this binding mechanism. Reyes-Turcu *et al* determined K_d values of for the first and second binding events to be approximately 600 nM and 20-30 nM, respectively. However the highest affinity site for monoubiquitin binding to USP5 is believed to be at the ZnF-UBP domain which, as an isolated domain, has been found to possess approximate K_d values for monoubiquitin of 2.3 and 2.8 μM, based on two independent measurements using different biophysical techniques [84, 198]. Accordingly it is noteworthy that these numbers are in excellent agreement with our estimate for the single monoubiquitin binding (~2.5 μM) to full-length USP5.

To further highlight the co-operative mechanism of monoubiquitin binding by USP5, titrations were completed versus a ΔGly75-Gly76 monoubiquitin mutant, which cannot bind at the ZnF-UBP domain. The approximate K_d determined for this single interaction was ~26.7 μM, thus in summary the interaction between the catalytic site of USP5 and monoubiquitin is greatly enhanced by the presence of monoubiquitin at the ZnF-UBP domain. Accordingly despite the ~13% structural collapse predicted by MD simulations

(discussed in section 4.1.4), we can conclude that ubiquitin recognition appears to be unhindered following transfer to the gas phase.

4.2.3 Carbene footprinting

The native ESI-IM-MS investigations of C335A-USP5 structure indicates that broadly two structures persist in the gas-phase, and using MD simulations, we demonstrate that these correspond to a compact structure, similar in cross-section to the recent X-ray structure of USP5, and a more extended structure, similar to that seen in previous SAXS analysis. Native ESI-MS investigations into the interactions of C335A-USP5 with monoubiquitin confirm that the binding only occurs to the compact USP5 conformer. Furthermore by utilising interaction USP5 and ubiquitin surface mutants, we can crudely assign binding sites. However, as alluded to in the *chapter 4* introduction, and indeed in Figure 4.1, the crystal structure of USP5, which in these studies we relate to the compact USP5 conformer (which binds to ubiquitin), does not appear compatible with simultaneous recognition of diubiquitin at the ZnF-UBP and the catalytic site. Accordingly to confirm the sites of ubiquitin recognition by USP5, we sought to map the sites of non-covalent interaction between C335A-USP5 and diubiquitin, the simplest form of polyubiquitin, in high resolution.

As alluded to in *section 1.3.2.2.4*, a number of structural MS based strategies, complementary to native MS, have been developed to permit the high resolution mapping of protein interaction sites. Indeed one the most promising strategies in this regard is diazirine-based protein carbene footprinting. This approach exploits the inherent and indiscriminate reactivity of carbenes, produced upon photoactivation of diazirines, to covalently label amino acids in the surface-accessible region of proteins, resulting in a mass shift (in this case 202.02 Da) detectable by MS. When employed in a differential manner, comparing accessible regions in the presence and absence of

ligand, this approach permits the determination of binding sites at amino acid resolution, as illustrated in Figure 2.2.

Recently investigations from David Schriemer's group [231, 271] have described the application of commercially available photoleucine and 4,4-azipentanoic acid for carbene footprinting, and demonstrated the effectiveness of the former to map the calmodulin-M13 interaction site. However these reagents were required in high concentrations (100 mM) to obtain labeling, and *in house* analyses (data not shown) revealed labeling with photoleucine to be highly protein dependent. Encouraged by these preliminary results, colleagues in the Oldham and Moses groups (at the University of Nottingham) have sought to design bespoke diazirine-based footprinting reagents to provide a platform for high resolution mapping of protein interactions by MS, including aryldiazirine 1 (a trifluoromethylaryl diazirine) indicated in *section 2.4.2*.

4.2.3.1 High resolution mapping of the USP5-diubiquitin interaction site

Accordingly carbene footprinting, utilising aryldiazirine 1, was employed to investigate the sites of Lys-48 linked diubiquitin recognition by C335A-USP5. The detailed experimental protocol is outlined in *section 2.4.2*. Briefly, purified C335A-USP5, in isolation, was first mixed with Aryldiazirine 1 (10 mM) and snap frozen, prior to irradiation at 349 nm using a Nd:YLF laser for 16 seconds. The photo-labelled samples were subject to SDS-PAGE and Coomassie staining. Bands corresponding to USP5 were excised and trypsin digested, prior to LC-MS/MS analysis. Bioworks software was subsequently used to identify C335A-USP5 peptides. A custom FASTA database, that included the C335A-USP5 sequence, was searched and allowed for 202.02 (±0.02) Da modification on all residues, which results from modification of amino acids with aryldiazirine 1.

LC-MS of photo-activated C335A-USP5 reproducibly resulted in 41% (340/837) sequence coverage identification, and pleasingly yielded highly labelled peptides derived from the ZnF-UBP domain (Leu250-Lys287), catalytic site (Asn332-Arg354, Gln435-Arg451) and the second UBA domain, with respect to USP5 primary sequence (Ala734-Arg757). An example of the unlabeled and mono-labelled G1-R19 USP5 peptide spectra and respective extracted ion chromatograms is indicated in Figure 4.18. To quantify the labelling pattern, the fractional modification of peptides was determined according to the method of Jumper *et al* [231], outlined in *section 2.4.2*. Briefly, the peak areas of unlabeled and mono-labelled C335A-USP5 peptides were extracted and used to solve Equation 2.2, yielding the peptide fractional modification. Three biological repeats were completed and the fractional modification value averaged (presented in Figure 4.19). Reassuringly the modification pattern proved highly reproducible across the three measurements.

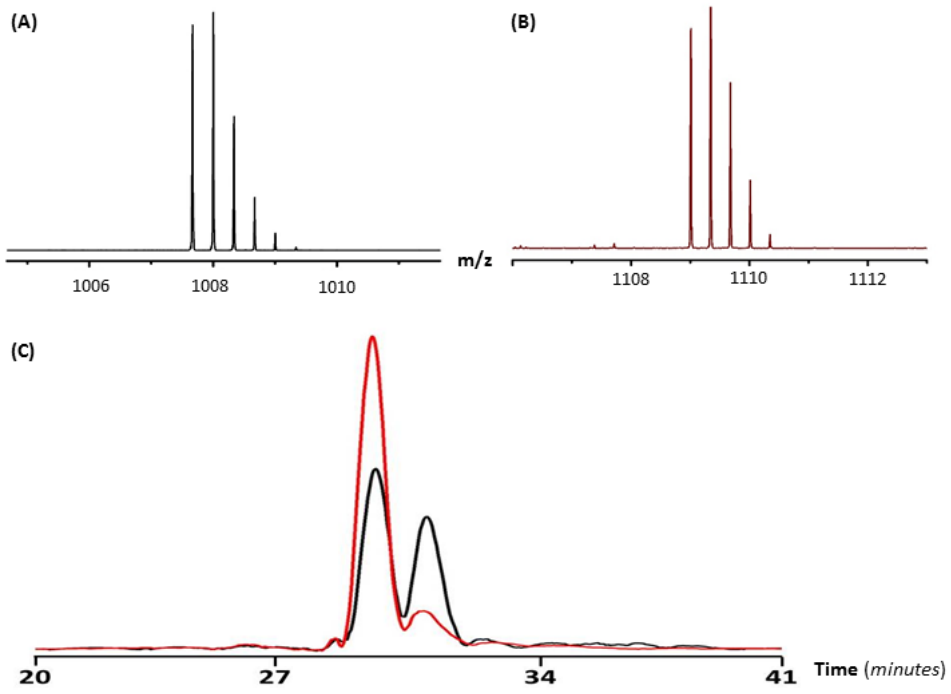


Figure 4.18 Monitoring the labelling of USP5 with aryldiazirine 1. (A) and (B) highlight representative spectra of the unlabeled and mono-labelled Gly-1-Arg-19 USP5 peptide respectively. (C) indicates the overlaid extracted ion chromatograms for the unlabeled (red) and labelled (black) G1-R19 USP5 peptide, which is subsequently utilised to determine peptide fractional modification.

The differential footprint was then determined for C335A-USP5, in the presence and absence of one molar equivalent of Lys-48 linked diubiquitin, following the same protocol as above. Samples were processed and the fractional modification determined. Subsequently the fractional modification values obtained for each peptide, derived from bound and unbound C335A-USP5, were averaged and plotted against each other (see Figure 4.19). Differences in the extent of labelling (or fractional modification) of peptides was considered significant when the p-value obtained from a Students t-test was <0.05. Large (> 10 %) statistically-significant differences in labeling are indicated in Figure 4.19A, with a red dot, furthermore these significant changes are also mapped onto the X-ray structure of USP5 (Figure 4.19B). The carbene footprinting results indicated that in the presence of Lys-48 linked diubiquitin there is a large and significant masking of labelling of peptides found in the ZnF-UBP domain (Leu250-Lys287), the catalytic site (Asn332-Arg354 and Gln435-Arg451) and a flexible loop between the two domains (Ile306-Arg331).

In addition to the large, significant differences in labelling highlighted in Figure 4.19, a very small (5 %), but statistically significant masking of a peptide (G606-K630) corresponding to a flexible loop adjacent to the two UBA domains was noted, which may result from the conformational change associated with substrate binding. Furthermore whilst no significant changes in labelling were associated with the second UBA domain in primary sequence, no peptides corresponding to the first UBA domain were identified.

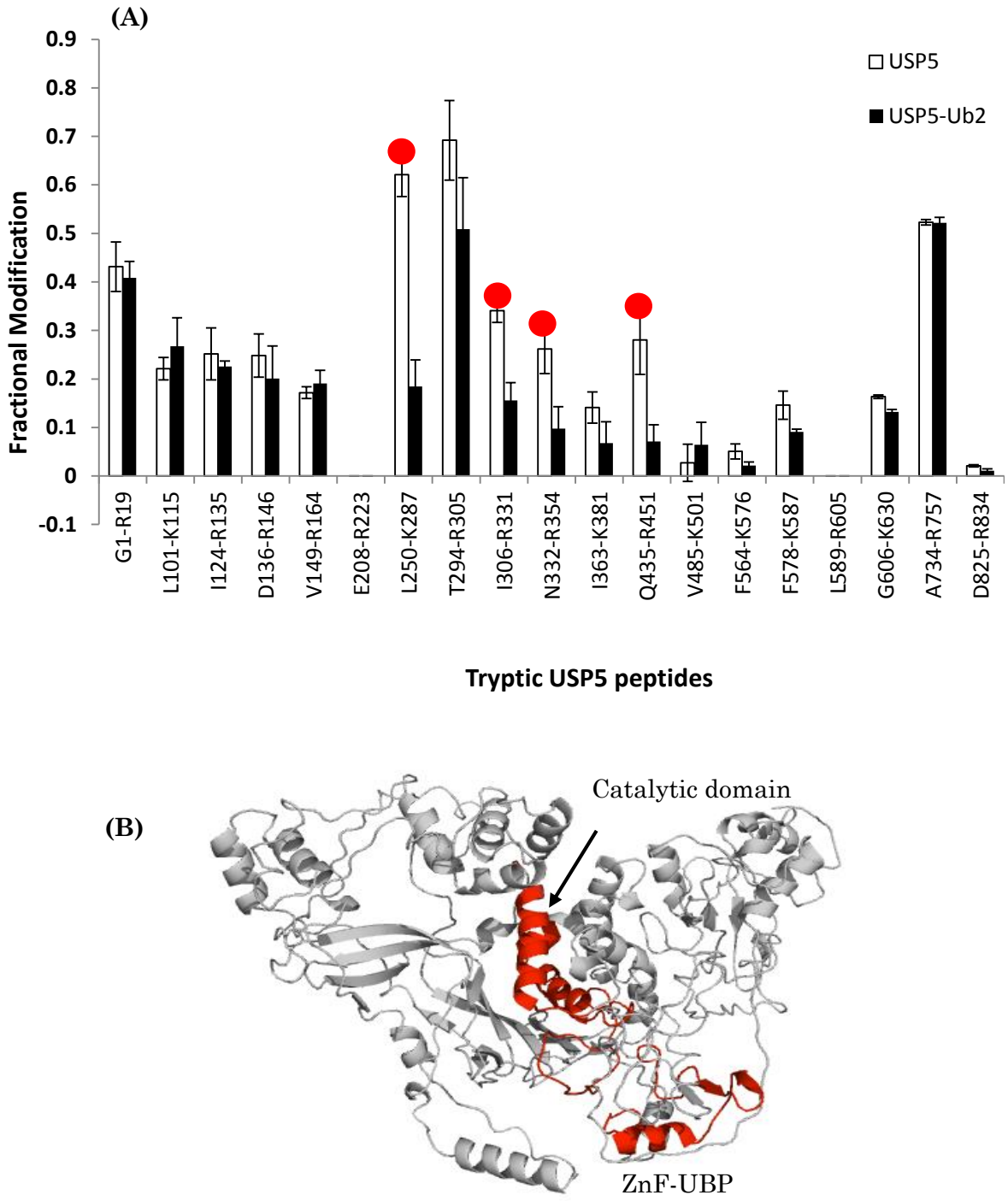


Figure 4.18 (A) The fractional modification of C335A-USP5 peptides with aryldiazirine 1, in the presence (black bars) and absence (white bars) of 1 molar equivalent of Lys-48 linked diubiquitin. Error bars are ± SD (n = 3) and significant differences (Student t-test, $p < 0.05$) are highlighted with a red dot. (B) Model of USP5 (based on PDB 3IHP) highlighting the locations of the four peptides (red) masked from labeling by Lys-48 diubiquitin binding, note that this corresponds to the ZnF-UBP and catalytic domains, there is no significant difference in labelling at either of the UBA domains.

4.2.3.2 Conclusions

Carbene footprinting the C335A-USP5-Lys-48 linked diubiquitin interaction revealed that the ZnF-UBP domain and catalytic site are responsible for interaction with diubiquitin, and since it has previously been established the stoichiometry of USP5-diubiquitin binding is 1:1 [87], this finding provides evidence that a conformation of USP5 exists wherein the ZnF-UBP domain is in close spatial proximity to the catalytic domain to permit recognition of Lys-48 linked diubiquitin at both domains. Thus, the conformation of USP5 revealed in the crystal structure, represents a conformer different to the catalytically active species.

4.2.4 Probing USP5-tetraubiquitin interactions

MS investigations into USP5 structure and interactions have revealed a conformational equilibrium that reflects a major compact conformer ($[M+17H]^+$ to $[M+21H]^+$), which binds to ubiquitin, and a minor extended conformer $[M+24H]^+$ to $[M+29H]^+$. Furthermore carbene footprinting reveals that USP5 undergoes conformational alterations upon ligand binding. We speculate that the extended structure of USP5 enables conformational remodeling by assorted ubiquitin topologies, to permit binding at all four UBDs. Accordingly if such conformers are representative of the solution phase, interconversion should be plausible upon occupation of all ubiquitin-binding sites, however thus far interactions with monoubiquitin did not seemingly alter USP5 conformer equilibrium.

4.2.4.1 C335A-USP5 vs Lys-63 linked tetraubiquitin

To further assess the USP5 conformer equilibrium with respect to ubiquitin binding, C335A-USP5 was titrated against differing concentrations of Lys63-linked tetraubiquitin (which proved to be the only accessible desalted tetramer available) and analysed by native ESI-MS, as previously indicated in *section 4.2.2*.

Addition of tetraubiquitin to C335A-USP5 resulted in a clear, very high affinity, binding event (see Figure 4.20). In the presence of 0.5 molar equivalents of tetraubiquitin, a 1:1 ratio of bound to unbound C335A-USP5 was observed, and when 1 molar equivalent of tetraubiquitin was added, a single set of signals for the C335A-USP5-tetraubiquitin complex was seen, with essentially no un-complexed protein, indicating that ~100 % bound to USP5. As with the second binding event of monoubiquitin (see Figure 4.15), the affinity of the interaction was too high to quantify by MS. Supporting this notion and indeed demonstrating a high affinity for polyubiquitin chains, previous ITC measurements by Reyes-Turcu *et al* revealed the approximate K_d for USP5-tetraubiquitin interaction to be ≤ 2 nM.

It was also clear from the spectra in Figure 4.20, that the binding of tetraubiquitin had a dramatic effect on the CSD of C335A-USP5. Indeed upon increasing additions of tetraubiquitin, the higher C335A-USP5 CSD became depleted to form compact, bound USP5 ions (Figure 4.20B and C). When one molar equivalent of tetraubiquitin was added, no higher USP5 CSD was visible in the mass spectrum. This is a significant result, as it indicates that the high and low CSDs represent species that are capable of inter-conversion in solution, and that tetraubiquitin binding is able to shift this equilibrium distribution in favor of structures that give rise to the low CSD.

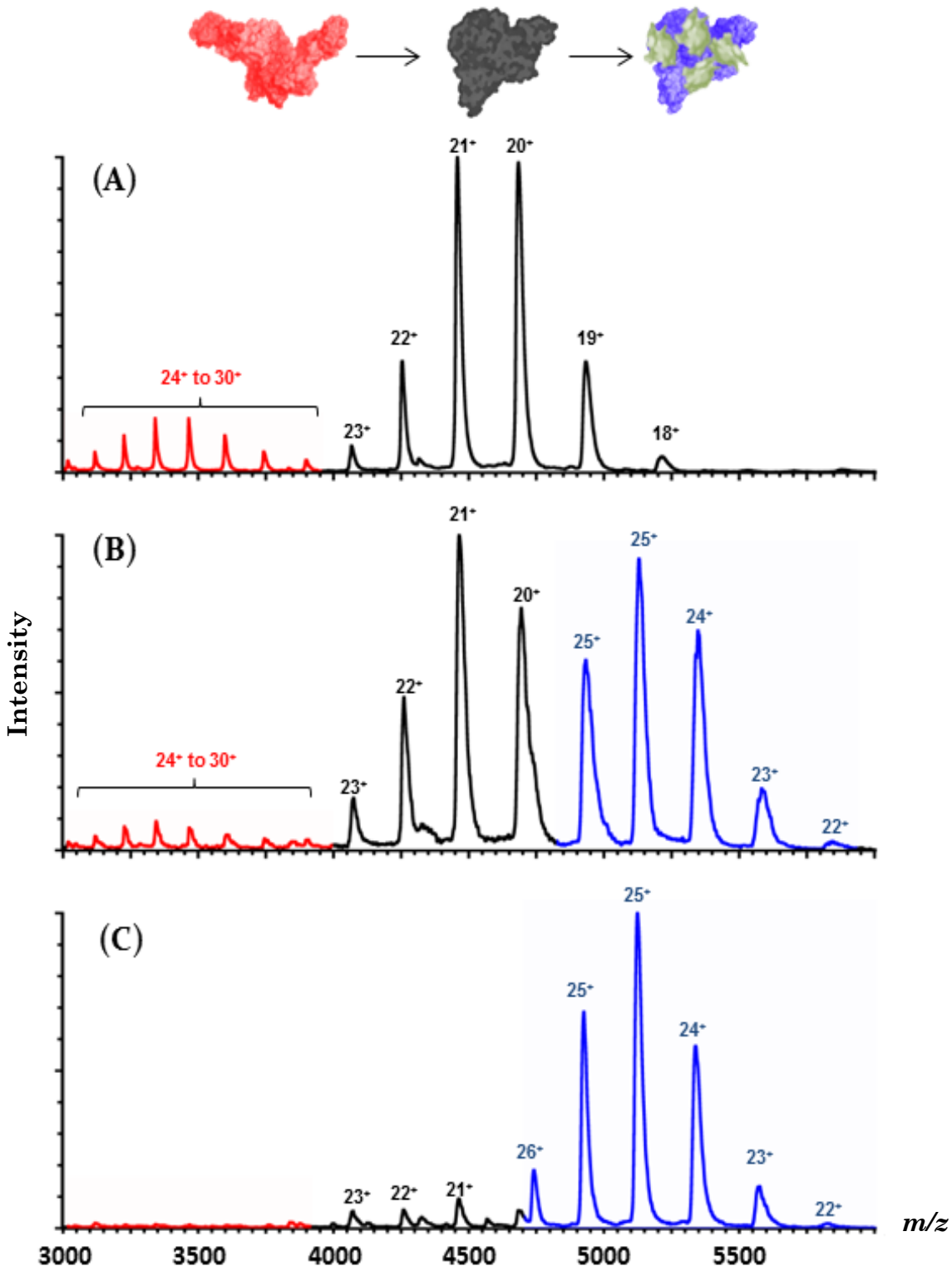


Figure 4.19 nESI-MS spectrum of C335A-USP5 (10 μM) (A) alone, (B) with 5 μM, and (C) with 10 μM Lys-63 linked tetraubiquitin. Highlighted are the extended USP5 conformer (red), the compact USP5 conformer (black) and the USP5-tetraubiquitin complex (blue) (Adapted from [270]).

4.3 Discussion

A recent crystal structure of USP5 has indicated the relative positions of the ubiquitin binding sites, but raised a number of important questions concerning the structure and dynamics of the enzyme; in particular with respect to the orientation of the C-terminal ZnF-UBP domain which, as illustrated in Figure 4.1, seemingly points away from the catalytic domain. This feature is seemingly at odds with the proposed function of USP5 in the recognition and processing of unanchored polyubiquitin [237]. Indeed the authors themselves allude to the possibility of a crystallisation artefact, induced by a non-native disulphide bond in the USP5 structure (between Cys-195 of the ZnF-UBP domain, and Cys-793 at the catalytic site) [33].

In this chapter we have presented a structural MS-based framework to probe USP5 structure and interactions, and thus clarify the crystal structure inaccuracies. First utilising Native ESI-IM-MS, set alongside MD simulations, we reveal two distinct USP5 conformers, a major compact conformer (low CSD, CCS ~52 nm^2), similar in CCS to the model of the recent X-ray structure (CCS ~53.5 nm^2), and a more extended structure (high CSD, CCS ~66 nm^2). USP5 contains a number of highly flexible loop regions between its folded domains, which are not seen by X-ray crystallography, but are indeed noted by SAXS analysis. Thus it is interesting to speculate that the high CSDs seen in the MS are derived from structural states with these loops in a more extended conformation, which gives rise to greater separation between domains, and a resulting higher CSD and CCS.

To probe the functional significance of the USP5 conformational dynamics, we sought to titrate C335A-USP5 against a series of ubiquitin substrates. This approach permitted clear visualization of the (positive) co-operative nature of ubiquitin binding by USP5, and in some cases afforded quantitation of such interactions. Furthermore by titrating against Lys-63 linked tetraubiquitin, we confirm the inter-conversion of the USP5

conformer equilibrium. Upon binding tetraubiquitin the USP5 domains appear to occupy a more compact structure, as they make contact with the various units of the polyubiquitin. Functionally, such inherent structural flexibility may permit the binding of differing polyubiquitin topologies (although it is noteworthy only Lys63-linked tetraubiquitin was explored), and may also be important in the catalytic process itself. In summary, despite (MD) predicted gas-phase collapse of USP5 (~13%) structure, we can conclude that insightful structural information, complementary to ITC and X-ray crystallography, can be obtained by ESI-IM-MS and indeed binding affinities can be determined that are in good agreement with the literature.

Finally in addition to native MS, we also present a carbene footprinting strategy, which utilizes a novel, efficient *in house* synthesized aryldiazirine to map the site of USP5-diubiquitin interaction. In relatively high resolution, this approach confirms that indeed structural rearrangement, relative to the recent X-ray structure, must occur to permit binding of diubiquitin substrates. Owing the large size of both the proteins involved in this interaction (USP5 ~94 kDa and Lys-48 linked diubiquitin ~17 kDa) and the surface area of the interaction, the carbene footprinting data was analyzed at the peptide level only, which proved adequate in identifying binding patches. However it is noteworthy, that by employing LC-MS/MS and sequencing peptides, it is possible to assign fractional modification at the amino acid level. Accordingly carbene footprinting, using aryldiazirine 1, represents a promising strategy to map the sites of protein-ligand interactions in high resolution, with the data generated proving complementary yet unique to that generated by X-ray crystallography.

In summary this chapter presents structural MS strategies to probe protein structure and interactions, in a quantitative manner. The results, where applicable, prove to be in good agreement with previous literature, completed using classic biophysical techniques such as ITC and X-ray crystallography. The MS-based approaches presented here also offer unique insights, unobtainable by these classic approaches, and thus demonstrate utility

as part of a biophysical toolbox to characterise protein structure and interactions in a complementary manner. With regards to mechanisms of ubiquitin recognition, we can conclude that in nature, multiple UBDs separated by flexible linkers permit the avid recognition of polyubiquitin, and presumably permits the recognition of polyubiquitin chains with diverse topologies. Accordingly UBDs separated in a linker-optimised manner may permit the linkage selective recognition of unanchored polyubiquitin.

Chapter 5 - Designing a tandem UBD construct for the selective recognition of unanchored polyubiquitin

Insights into the biological significance of specific polyubiquitin assemblies, and the ubiquitin-mediated processes they regulate have in some cases been afforded by exploiting the inherent specificities of UBDs as *in vitro* tools. Indeed in *chapter 3* the isolated ZnF-UBP domain of USP5, which has a relatively high affinity for the free C-terminus of unanchored polyubiquitin ($K_d \approx 2.3$ μM *vs* monoubiquitin) [272], was employed to permit the purification of unanchored polyubiquitin from mammalian cell extracts. Whilst the isolated ZnF-UBP (FUBE) enrichment strategy excludes the purification of conjugated polyubiquitin and associated covalently modified target proteins, it fails to discriminate between unanchored ubiquitin monomers and polyubiquitin chains, the former of which predominate *in vivo*, as exemplified in Figure 3.4. Furthermore previous quantitative native ESI-MS studies of the isolated ZnF-UBP domain demonstrated that this UBD exhibits no intrinsic selectivity for unanchored ubiquitin dimers or tetramers of different isopeptide linkages [272].

By extending the original FUBE enrichment strategy, to utilise full length multiple-UBD containing USP5, we were able to favour the purification of unanchored polyubiquitin over monoubiquitin, under bait limiting conditions (see *section 3.2*). Alternately, the use of constructs containing tandem repeats of the same UBD, as tools for the affinity enrichment of the ubiquitome, is the premise of the artificial protein sequences, termed Tandem Ubiquitin-Binding Entities (TUBEs) [110] and the recent

tandem hybrid UBD constructs [273]. As with USP5 (see *chapter 3*), whilst avidity effects account for the high affinity binding of polyubiquitin to the multi-UBD containing constructs, typically these artificial polymers do not exhibit linkage selectivity [64].

Towards the goal of selectively purifying defined polyubiquitin topologies, a number of groups have demonstrated the utility of isolated UBDs which display inherent linkage selectivity, to purify or detect ubiquitin-modified proteins *in vitro* [64], and also to probe polyubiquitin modifications in living cells [114]. However, natural polyubiquitin linkage selective domains (or at least those identified to date) are sparse, and such domains have only been identified for Met-1 as well as Lys-29, 48 and 63 linked polyubiquitin chains, as summarised in Table 5.1. An alternative strategy to realise linkage selectivity of ubiquitin-binding proteins, as demonstrated by Sims and co-workers [81], is to place multiple non-specific UBDs in tandem, with the linker region between domains positioning for efficient avid binding across different ubiquitin moieties in a chain. This is typified by linker optimisation of the tandem UIMs of RAP80, which can be fine-tuned to achieve the selective recognition of Lys-63 linked polyubiquitin chains [81].

In summary UBDs can be considered amenable to synthetic biology approaches, wherein tandem UBDs are assembled with engineered recognition properties. Critically, novel reagents for the selective detection and purification of polyubiquitin chains of defined chain linkage/length, and which distinguish unanchored from conjugated polyubiquitin, remain a priority for the field. Accordingly towards this goal, in this chapter we demonstrate the rational design and validation of a tandem UBD (t-UBD) hybrid protein, which for the first time utilises two different UBDs fused in a linker optimised manner, to exploit avidity effects that afford the recognition and purification of Lys48-linked unanchored polyubiquitin chains.

5.1 Designing the t-UBD

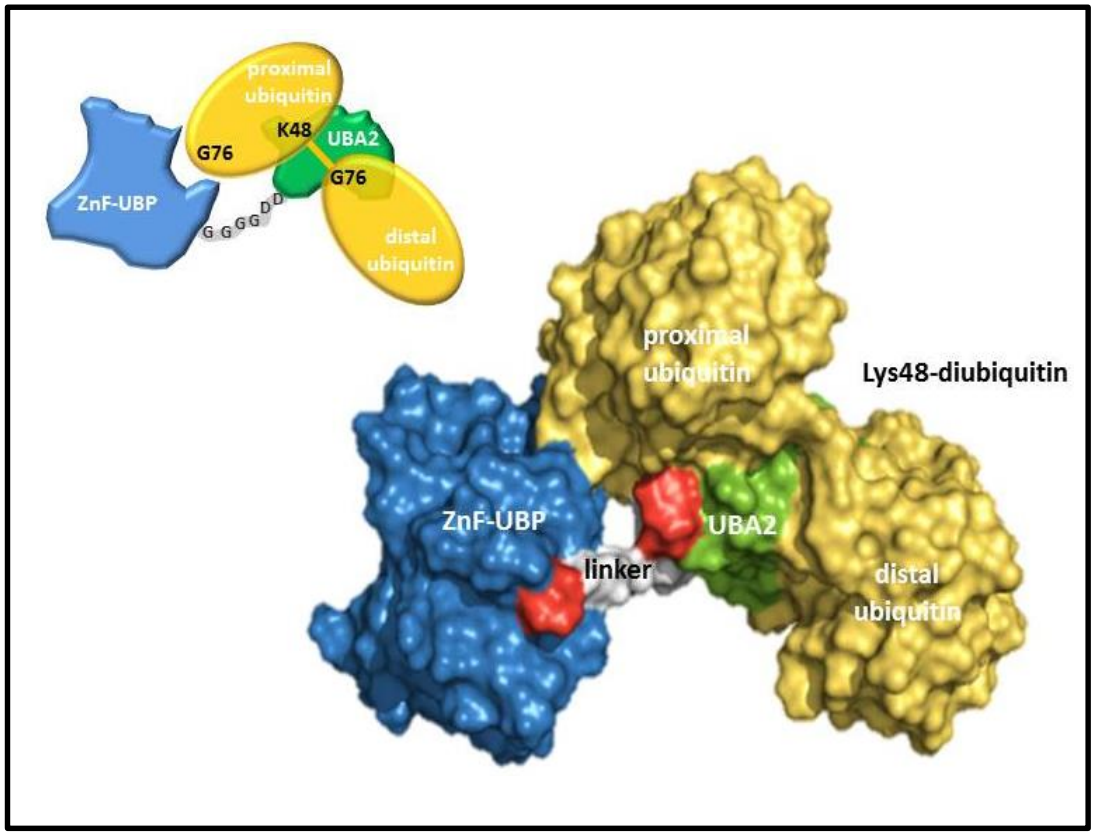
The number of naturally occurring UBDs that display linkage selectivity, summarised in Table 5.1, are limited, although in some cases have found application as *in vivo* biosensors to probe ubiquitin-modified substrates of defined linkages in living cells [114]. Towards the goal of purifying Lys-48 linked unanchored polyubiquitin in a linkage selective manner, logically we sought to exploit the inherent specificity of the UBA2 of hHR23A. As alluded to in chapter 1 (see Figure 1.5), the work of Varadan *et al* [70], has shown that the UBA2 of hHR23A has a natural propensity to recognise Lys-48 linked polyubiquitin chains via exploitation of the compact nature of the Lys-48 linked chain in the simultaneous, two faced binding of the UBA2 domain to Ile-44 sites on both proximal and distal ubiquitin moieties in a dimeric unit.

By linking the UBA2 of hHR23A to a ZnF-UBP domain, it was reasoned that the product would be a reagent which would favour the enrichment of unanchored Lys-48 linked polyubiquitin over other polyubiquitin chains and monoubiquitin. Furthermore, it was rationalised that by utilising a ZnF-UBP domain which displays lower affinity towards monoubiquitin (when compared to the ZnF-UBP of USP5), the reagent would begin to disfavour the purification of unanchored monoubiquitin with respect to polyubiquitin. Thus the target construct included the ZnF-UBP domain (residues 23-143) of the DUB USP16, which displays an affinity of ~6.5 μM versus monoubiquitin (compared to 2.3 μM of the USP5 ZnF-UBP) [239], linked to the Lys48-selective UBA2 domain of hHR23A (residues 315-362) [70].

Table 5.1 Selected individual UBDs which display polyubiquitin linkage specificity

UBD	Specificity	Reference
NZF of HOIL-1	Met-1	[76]
UBAN of NEMO	Met-1	[79]
NZF of TRABID	Lys-29	[56]
UBA2 of hHR23A	Lys-48	[70]
NZF of TAB2/3	Lys-63	[74, 75]
NZF of TRABID	Lys-29	[56]

MD simulations were then used to generate a sterically and energetically plausible model of the complex, and ensure that no direct contact between the two UBDs occurred, and that critical recognition surfaces on each isolated domain were unhindered.The ZnF-UBP domain of USP16 (PDB code 2I50) was first modelled onto the existing USP5 ZnF-UBP: monoubiquitin complex (PDB code 2G45) using MOLMOL [238]. This model was subsequently used to guide the docking of the ZnF-UBP domain on to the protruding free C-terminus of the proximal ubiquitin of the UBA2:Lys48-diubiquitin complex (PDB code 1ZO6). The MD simulations subsequently performed, indicated a requirement for a linker region between the two domains of at least 17 Å (equivalent to six amino acids) between the final structured C-terminal Phe-360 residue of UBA2 and first structured N-terminal Val-23 residue of the ZnF-UBP domain. Accordingly, a linker which accommodates the two native C-terminal aspartates from the UBA2 sequence in addition to four Gly residues was incorporated. Thus, the final arrangement of the t-UBD placed the UBA2 domain N-terminal to the ZnF-UBP domain separated by a six amino acid (Asp_2Gly_4) flexible linker (see Figure 5.1), which was demonstrated by MD simulations to be stable over 1 ns.



GSQEKEAIERLKALGFPESLVIQAYFACEKNENLAANFL
LSQN**F**DDGGGG**V**CRHIRKGLEQGNLKKALVNVEWNICQD
CKTDNKVKDKAEEETEEKPSVWLCLKCGHQGCGRNSQEQ
HALKHYLTPRSEPHAROSECLVLSLDNWSVWCYVCDNEV
QYCSSNQLGQVVDYVRKQAS

Figure 5.1 t-UBD design. (Top) Model (and cartoon) of the t-UBD hybrid protein produced by docking (guided by PDB code 2I50) the ZnF-UBP domain of USP16 (blue) on to the protruding free C-terminus of the proximal ubiquitin in a model of the hHR23A-UBA2:Lys48-diubiquitin complex (green (partly obscured), yellow; PDB code 1ZO6). The distance between C-terminal Phe-360 of UBA2 and N-terminal Val-6 of ZnF-UBP (both red) was estimated as 17.2 Å and indicated an absolute minimum linker length of six residues. We subsequently included a six residue flexible linker, including four glycine residues (white) along with two C-terminal aspartates donated by the UBA2 sequence. (Below) Primary sequence of the final t-UBD construct as follows; Gly-Ser overhang from thrombin cleavage, UBA2 sequence (blue), linker (orange), ZnF-UBP (green). Arg-84 of ZnF-UBP domain indicated in red (Adapted from [274]).

5.2 Synthesising the t-UBD

The codon optimised DNA sequences for the t-UBD construct were assembled and a custom gene was synthesised commercially. The gene was designed with *Bam*HI and *Xho*I restriction sites at the 5' and 3' termini respectively. Subsequently the t-UBD gene sequence was double digested with these enzymes, and ligated into the *Bam*HI and *Xho*I restriction sites in the pGEX-4T-1 bacterial expression vector, which permits expression of the t-UBD with an N-terminal GST-tag. To verify ligation of the t-UBD into the pGEX-4T-1 vector, purified plasmid DNA was subject to enzymatic digest with *Bam*HI and *Xho*I, and the products electrophoresed on a 1% agarose gel (see Figure 5.2A). The construct was further verified by DNA sequencing.

The t-UBD was expressed in E. *coli* as a GST-fusion, which permitted initial purification on glutathione Sepharose, prior to thrombin cleavage to release the t-UBD protein. The t-UBD was subsequently purified by anionic exchange chromatography, and subject to SDS-PAGE analysis, which resulted in a single band indicated in Figure 5.2B (which demonstrates a representative t-UBD purification). 1 L of culture typically yielded 2 mg of purified t-UBD. The mass of the purified t-UBD, as measured by ESI-MS, was reported as 19,747 Da (expected, 19748 Da, Figure 5.2C).

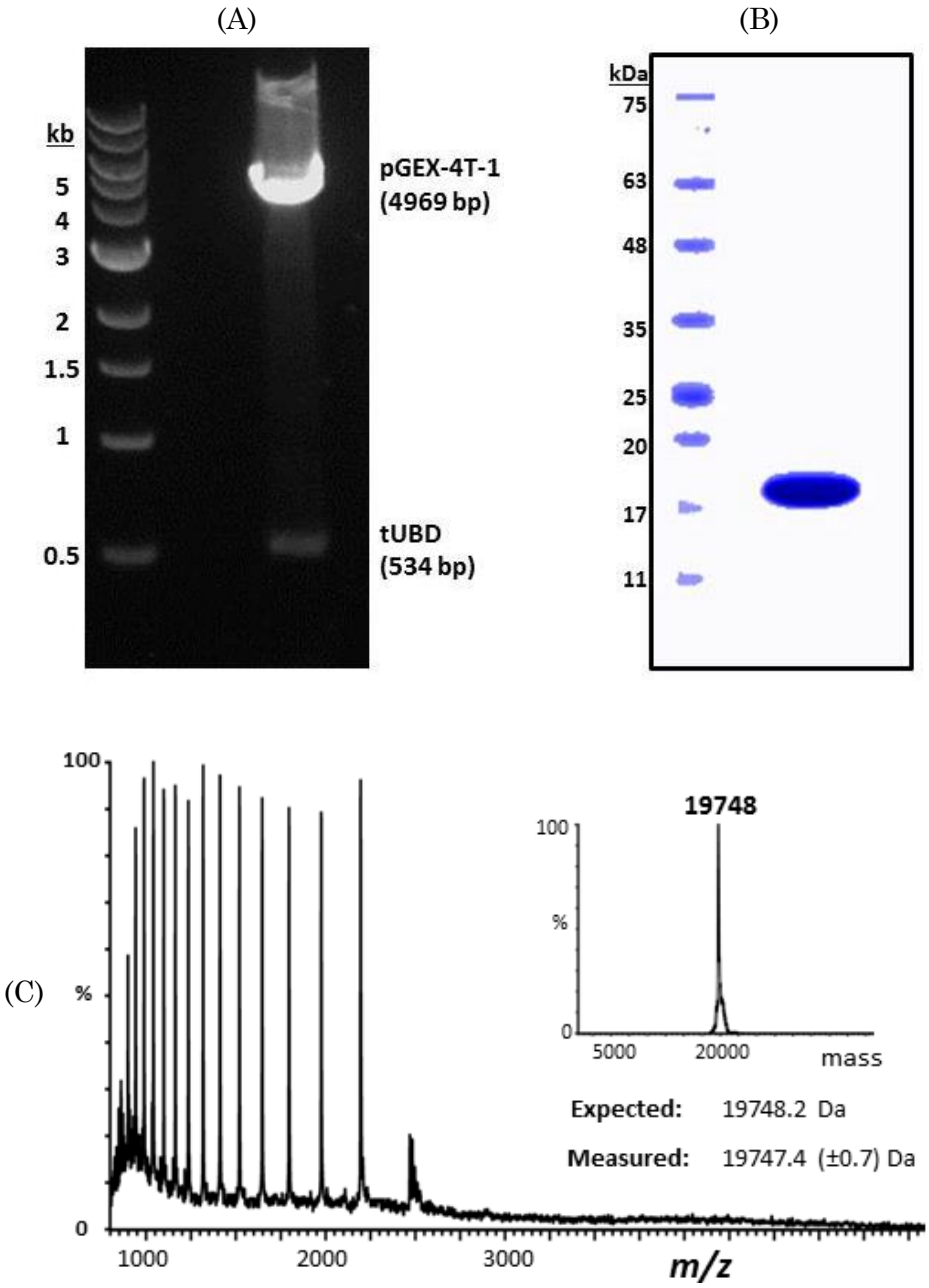


Figure 5.2 Validating the integrity of the t-UBD construct. (A) A 1% agarose gel of the double digest (*Bam*HI/*Xho*I) of the t-UBD-pGEX-4T-1 construct, to confirm construct and insert size. The t-UBD construct was transformed into *E. coli*, overexpressed and purified. (B) A representative Coomassie stained 5-20% gradient gel of the t-UBD protein (20 µg loaded) following final purification. The mass of the purified protein was confirmed by ESI-MS, (C) illustrates the full and deconvoluted mass spectra of denatured t-UBD, sprayed from 80 % MeCN, 0.1 % formic acid (analysis completed on the Synapt HDMS).

5.3 Confirming the functionality of the t-UBD

Following purification, the t-UBD protein was first screened for binding selectivity against Lys-48 and 63 linked diubiquitin in an *in vitro* pulldown assay. Sepharose coupled t-UBD demonstrated preferential capture of unanchored Lys-48 linked diubiquitin compared to Lys-63 linked diubiquitin. Under the same conditions the higher affinity (than USP16) isolated ZnF-UBP domain from USP5 was able to capture both Lys-48 and 63 linked diubiquitin (indicated in Figure 5.3, below).

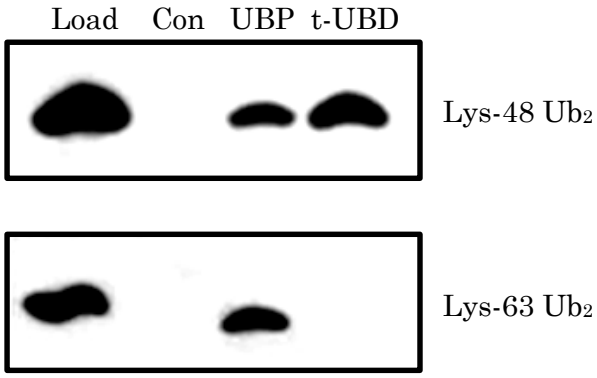


Figure 5.3 Diubiquitin pulldown assays with the t-UBD protein. Representative anti-ubiquitin western blot of t-UBD-diubiquitin pulldowns.50 µL of control, ZnF-UBP (10 mg/ml) or t-UBD (eqimolar to ZnF-UBP) Sepharose was incubated 250 ng of Lys-48 or 63 linked diubiquitin. Following binding, the Sepharose was washed and the enriched material eluted directly into GAB, prior to separation by SDS-PAGE on a 5-20% gradient gel. The gel was transferred to nitrocellulose membrane and immunoblotted against an anti-ubiquitin antibody (VU-1). Load represents 100% of the polyubiquitin input. The t-UBD captures only Lys48-diubiquitin under these conditions. The higher affinity isolated ZnF-UBP domain (USP5) ($K_d \approx 3.6$ µM for Lys48-diubiquitin and 2.8 µM for Lys63-diubiquitin [272]) is capable of capturing both dimers (Adapted from [274]).

To assess whether the desired avidity effects were realised in the synthetic t-UBD protein, we performed native ESI-MS experiments to quantitate interaction of the t-UBD with Lys-48 and 63 linked diubiquitin. Briefly, ubiquitin substrates (1 µM), were titrated against differing concentrations of the t-UBD protein (2, 4, 6 and 8 µM) (with three replicates at each concentration) and analysed by ESI-MS, under optimised

conditions (see *section 2.4.3*). Subsequently, the intensities of peaks corresponding to free and bound ubiquitin were utilised to determine the apparent binding affinity. The findings, which support the pulldown assays, indicate that the t-UBD readily formed complexes with unanchored Lys-48 linked diubiquitin binding with a $K_d \approx 1\mu M$, compared to Lys63-diubiquitin ($K_d \approx 9\mu M$) and monoubiquitin ($K_d \approx 11\mu M$), as illustrated in Figure 5.4 and Figure 5.6. Since the isolated UBA2 domain has a previously reported K_d *vs* Lys48-diubiquitin of ≈ 18 μM and *vs* monoubiquitin of ≈ 200 μM [272], and the ZnF-UBP a K_d *vs* monoubiquitin of ≈ 6.5 μM [239], we rationalised that the higher affinity (≈ 1 μM) interaction of the t-UBD with unanchored Lys-48 linked diubiquitin results from both UBDs simultaneously recognising both ubiquitin units in the Lys-48 linked dimer. Furthermore, binding of the t-UBD to Lys-63 linked diubiquitin with $K_d \approx 9$ μM was entirely consistent with recognition of just the proximal ubiquitin by the ZnF-UBP and the functionality of that UBD. Figure 5.6 illustrates a cartoon schematic of the binding affinities associated with t-UBD-ubiquitin interactions.

To confirm the contribution of both UBDs within the t-UBD to the recognition of unanchored Lys48-diubiquitin, we sought to abrogate ubiquitin binding at the ZnF-UBP domain via introduction of a mutation at Arg-84 of the ZnF-UBP sequence (Arg-110 of t-UBD) to Ala. Whilst this residue had not previously been the target of mutagenesis in published studies, it is conserved throughout all ZnF-UBP domain family members (Arg-221 in USP5, see Figure 5.5A) [239], and can be found in the ubiquitin-binding cleft. This mutant t-UBD was confirmed by DNA sequencing (Figure 5.5B) and expressed and purified in the same manner as the WT t-UBD (spectrum illustrated in Figure 5.5C, see Figure 5.2 for representative SDS-PAGE). The affinity for Lys-48 linked diubiquitin was assessed by native ESI-MS. In this case an apparent K_d of ≈ 23 μM *vs* Lys-48 linked diubiquitin, comparable to that we reported for the UBA2 alone ($K_d \approx 18$ μM), is consistent with abrogating the ZnF-UBP interaction with only the

UBA2 contributing to binding. This was further confirmed following ESI-MS titration of the WT t-UBD construct against the C-terminal monoubiquitin mutant (ΔGly75-Gly76, see *chapter 4*), which is critical for recognition by the ZnF-UBP domain [239]. The apparent affinity for this protein was ≈ 198 μM (Figure 5.6), consistent with only weak binding at the UBA2. Thus, the data suggests that the tandem UBDs are contributing to an avidity effect which underlies high affinity recognition of Lys-48 linked diubiquitin.

We can conclude that our engineered avidity effect contributes to a ≈ 11-fold preference of the t-UBD for unanchored Lys-48 linked diubiquitin over monoubiquitin, and that the t-UBD exhibits selectivity for Lys-48 linked dimers (at least 7-fold over Lys63-linked dimers). It is noteworthy that selectivity of comparable magnitude is reported for the naturally occurring tandem UIMs of RAP80, which show an ≈ 24-fold selectivity for Lys-63 linked diubiquitin over monoubiquitin [81].

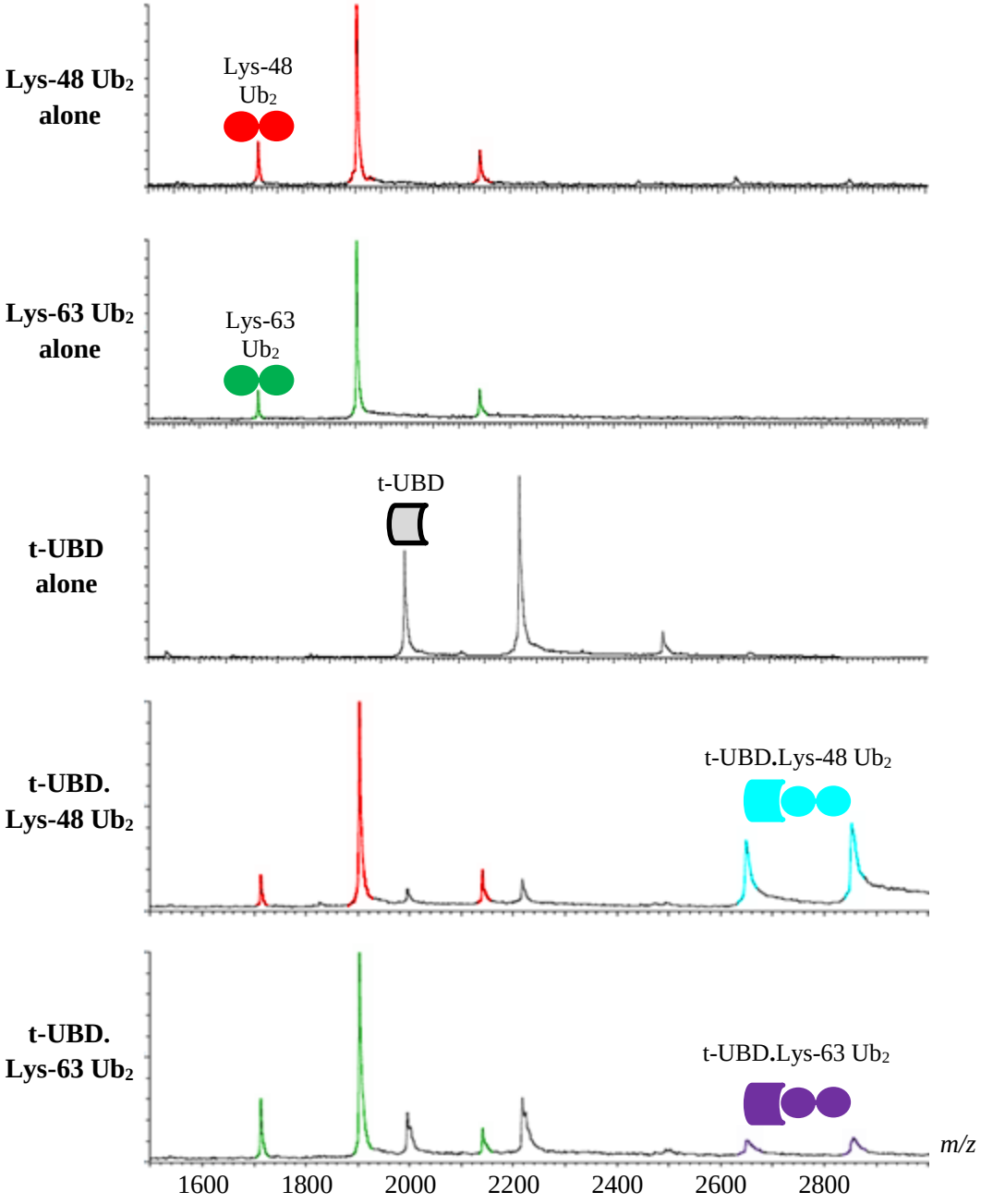
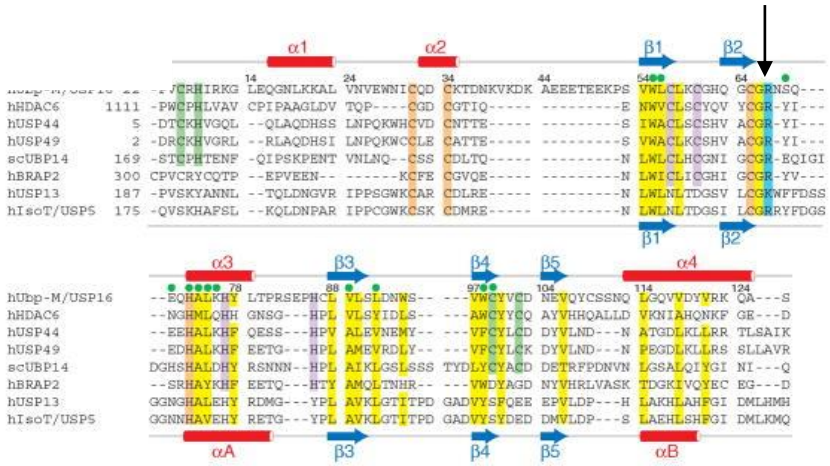
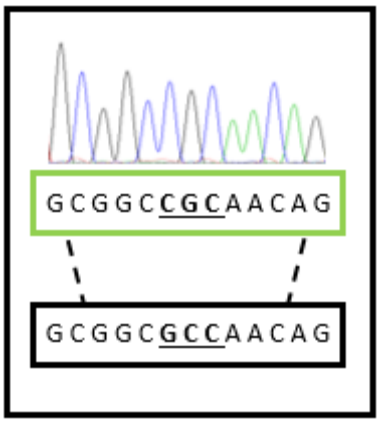


Figure 5.4 Confirming the functionality of the t-UBD with unanchored ubiquitin dimers. Representative ESI-MS spectra of the t-UBD (6 μM) and Lys-48 or Lys-63 linked diubiquitin (1 μM), both in isolation and in complex (as indicated). The t-UBD more readily forms complexes with unanchored Lys-48 diubiquitin than with Lys-63 diubiquitin, as evidenced by the stronger signal for bound complex in the Lys-48 diubiquitin spectrum and for unbound t-UBD in the Lys-63 diubiquitin spectrum. Titration analyses were used to estimate K_d values for the interactions (see Figure 5.6) (Adapted from [274]).

(A)



(B)



(C)

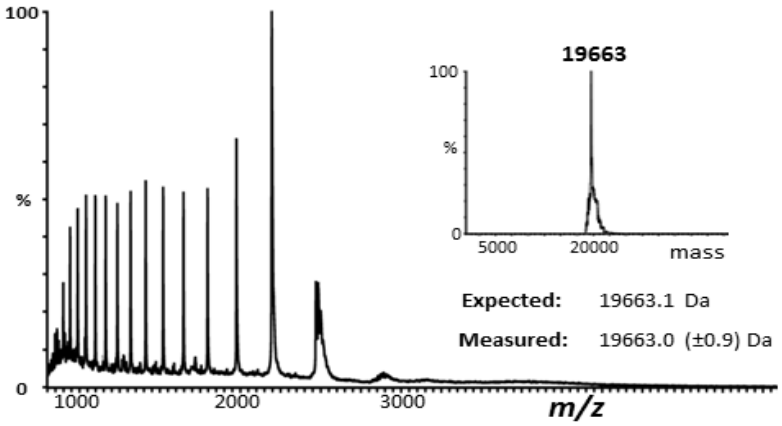
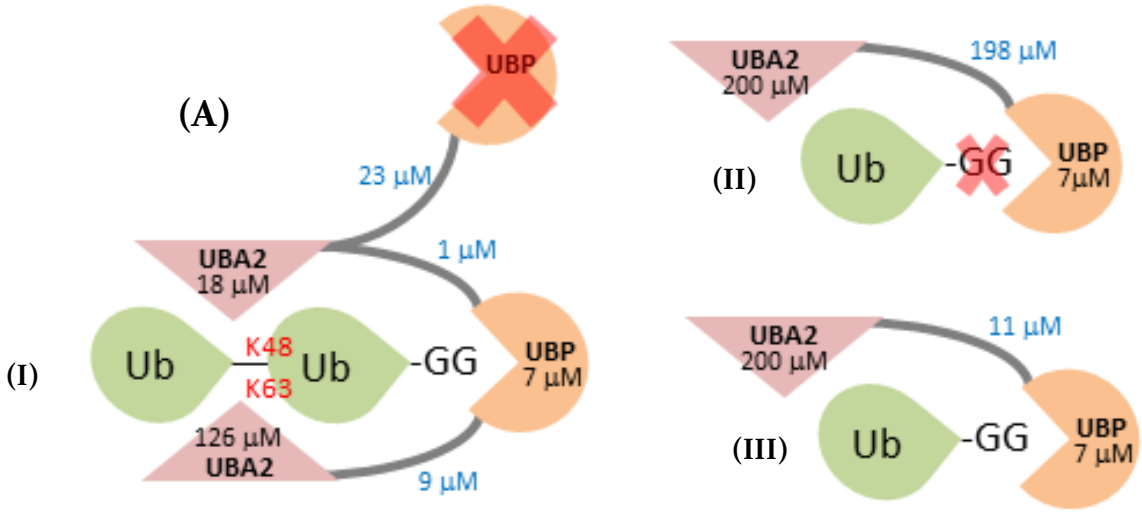


Figure 5.5 Mutating the ZnF-UBP domain of the t-UBD. (A) Sequence alignment of ZnF-UBP family members (adapted from [239]) revealed a conserved Arg residue (Arg-110 of USP16) highlighted in blue (and indicated with an arrow), which was subsequently mutated to Ala by site-directed mutagenesis, and confirmed by Sanger sequencing in (B). Following expression and purification, the mass of the purified ZnF-UBP mutant t-UBD protein was confirmed by ESI-MS. (B) illustrates the full and deconvoluted mass spectra of denatured t-UBD mutant (Arg-110-Ala), sprayed from 80 % MeCN, 0.1 % formic acid (analysis completed on the Synapt HDMS).



(B)

	monoubiquitin	ΔGly75-Gly76 monoubiquitin	Lys-48 diubiquitin	Lys-63 diubiquitin
UBA2[a]	200 ± 14	nd	18 ± 2	126 ± 18
ZnF-UBP[b]	6.5	nd	nd	nd
t-UBD	10.9 ± 0.4	197.5 ± 4.3	1.3 ± 0.3	9.9 ± 1.5
t-UBD (ZnF-UBP) mutant	nd	nd	23 ± 1.8	nd

Figure 5.6 Summary of the affinities displayed by the t-UBD towards ubiquitin substrates. (A) Cartoon schematic of the t-UBD protein in complex with (I) diubiquitin (Lys-48 or 63, as indicated), monoubiquitin ((II) ΔGly75-Gly76 or (III) WT). Binding affinities previously determined for isolated UBDs indicated in black, ESI-MS derived affinities for the t-UBD protein indicated in blue. (B) Table to confirm affinities indicated in (A). The K_d values reported here for the t-UBD (mean ± SD, µM) were averaged from values measured at different concentrations; nd - not determined. Values generated in this study are shaded. [a]Determined by ESI-MS [272]. [b]Determined by ITC [239]. Note that values of the t-UBD versus Lys-63 linked diubiquitin and monoubiquitin are consistent with only the ZnF-UBP making a major contribution to binding in these cases (Adapted from [274]).

5.4 Optimising the t-UBD linker length

To ensure the optimal positioning of UBDs for the selective recognition of unanchored Lys-48 linked diubiquitin had been arrived at by computational modelling, subsequent investigations sought to explore the effect of t-UBD linker length on Lys-48 linked diubiquitin recognition. Previous studies have shown that in the natural tandem UIM construct, RAP80, the linker length and its structural integrity significantly modulates the affinity of the protein for Lys-63 linked polyubiquitin chains. This is achieved by controlling UIM orientation and the relative phasing of binding sites along the helix, determining relative avidity effects [81].

Accordingly site-directed mutagenesis was utilised, in combination with native ESI-MS, to quantitate the binding affinity of the t-UBD for Lys-48 linked diubiquitin upon lengthening, or shortening the existing six residue t-UBD linker. Gly residues were deleted or inserted by site-directed mutagenesis, to yield t-UBD constructs with linkers ranging from 4-8 residues in length. The integrity of the constructs was confirmed by DNA sequencing and they were subsequently expressed and purified in the same manner as the WT t-UBD, prior to native MS analysis to confirm the mass of each linker variant (indicated in Figure 5.7).

Upon titration of the linker variants against Lys-48 linked diubiquitin using native ESI-MS analysis, in the same manner as section 5.3 (see Figure 5.8), we found that the addition or removal of even a single Gly residue from the linker increased the apparent K_d values, consistent with diminishing the avidity effect. The addition or removal of two residues produced apparent K_d values of ≈ 10 μM, close to those reported for the isolated ZnF-UBP *vs* monoubiquitin (K_d ≈ 6.5 μM [239]) suggesting a major contribution of only this domain to the binding event. Whilst initially surprising, presumably weaker binding with shorter and longer linkers reflects clashes between the two UBDs, linker and diubiquitin (shortening) and higher entropic costs associated with

ordering the linker upon binding (lengthening). Accordingly we can conclude that the design strategy is valid, and reassuringly arrived at the optimum linker length.

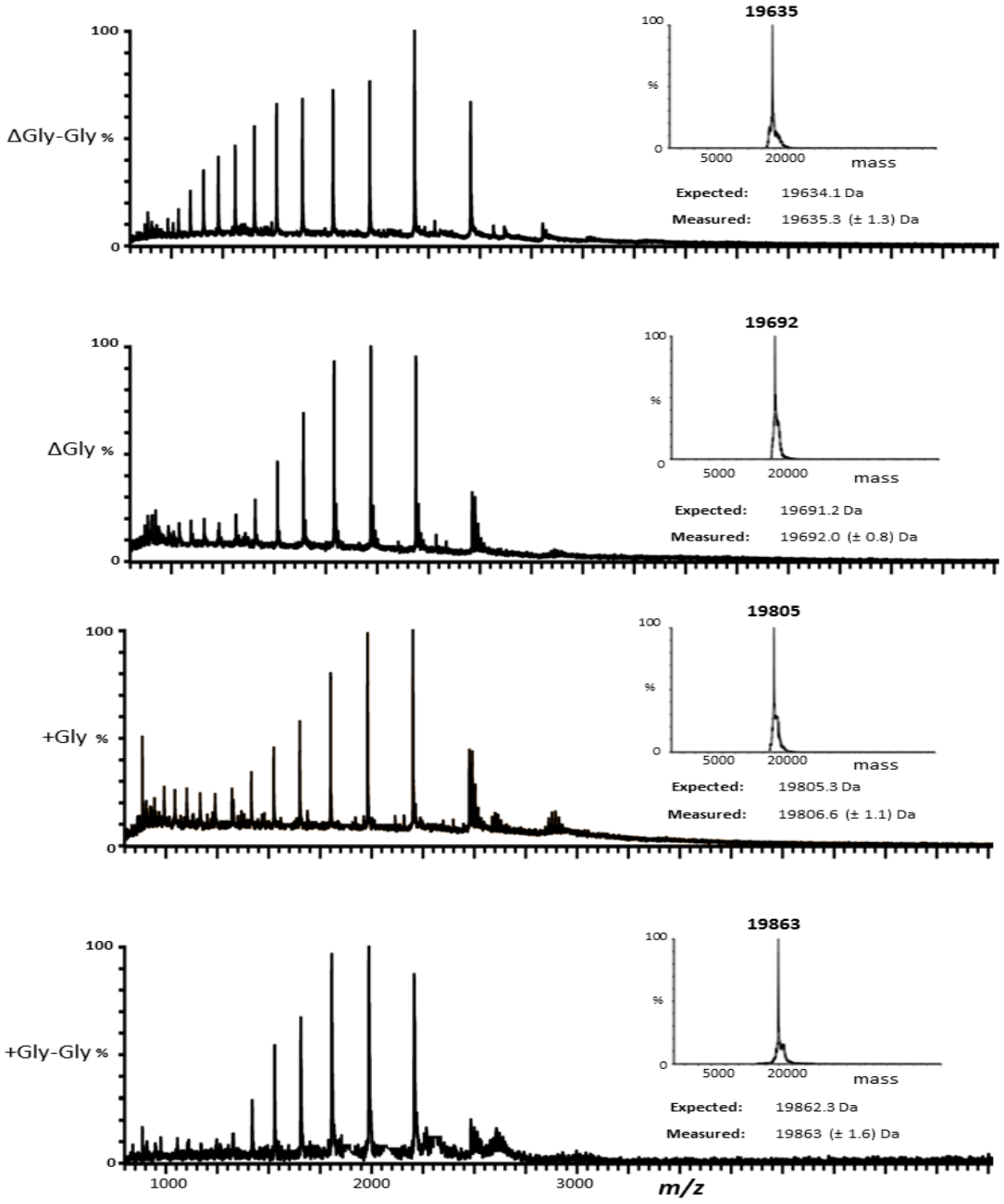


Figure 5.7 Confirming the integrity of t-UBD linker length variants. Full and deconvoluted mass spectra of denatured t-UBD linker variants (variation indicated to left of spectrum) sprayed from 80 % MeCN, 0.1 % formic acid (analysis completed on the Synapt HDMS).

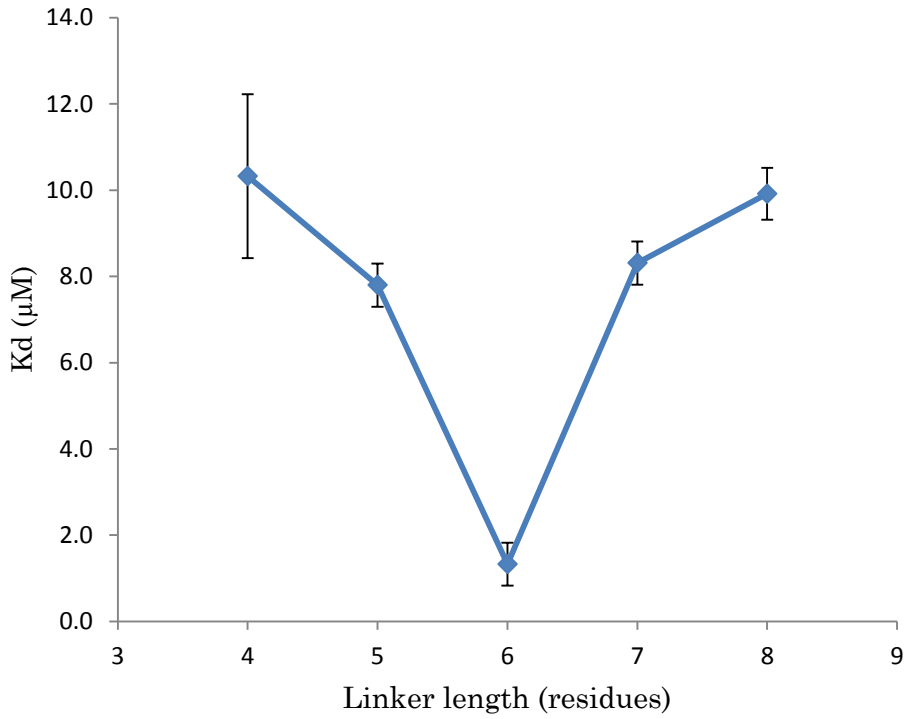


Figure 5.8 Validating the t-UBD design strategy. The effects of shortening or lengthening the six residue t-UBD linker on the recognition of unanchored Lys-48 diubiquitin was investigated. Site-directed mutagenesis was used to insert or delete one or two Gly residues within the t-UBD linker, and ESI-MS applied to estimate K_d values for the interaction with Lys48-diubiquitin (Adapted from [274]).

5.5 Purification of endogenous unanchored polyubiquitin chains by the t-UBD

Final investigations sought to assess the binding properties of the t-UBD with respect to the purification of endogenous unanchored polyubiquitin chains from mammalian cell extracts. HEK293T lysates were prepared according to the unanchored polyubiquitin chain purification protocol, established in *chapter 3*.

Western blot analyses confirmed that in comparison to the isolated ZnF-UBP domain from USP5, under bait-limiting conditions, the t-UBD displayed preferential capture of

longer unanchored polyubiquitin chains from HEK293T cell extracts (Figure 5.9, grey triangle), as detected with both anti-ubiquitin (Figure 5.9A) and Lys-48 linked polyubiquitin specific antibodies (Figure 5.9B). Importantly analysis of the blots revealed a clear advantage of the t-UBD over the original isolated ZnF-UBP domain, with almost no capture of free monoubiquitin associated with the former, reflecting the lower binding affinity of the ZnF-UBP domain of USP16 (Figure 5.9A), in a similar manner to that seen with C335A-USP5 in Figure 3.9. The unanchored nature of the purified polyubiquitin chains was confirmed by completing a deubiquitylation assay, treating samples on beads with catalytically active WT USP5 (Figure 5.9C), which is highly selective for substrate-free polyubiquitin [275], wherein chains were completely disassembled to ubiquitin monomers.

There was no obvious preference of the t-UBD, over the isolated ZnF-UBP domain, for the capture of endogenous Lys-48 linked diubiquitin, revealed through the use of Lys-48 polyubiquitin linkage selective antibodies, although this may represent the relatively high abundance of this linkage in endogenous dimers [107]. The preferential binding of the t-UBD to longer chains over the isolated ZnF-UBP domain supports previous ESI-MS data in which the UBA2 domain displays a higher affinity for tetraubiquitin *vs* diubiquitin (8 μM *vs* 18 μM) [272]. We speculate that as with tandem UIMs [81], the compact nature of long Lys-48 linked chains may lead to increased avidity effects owing to simultaneous interaction of non-adjacent ubiquitin's with the t-UBD protein.

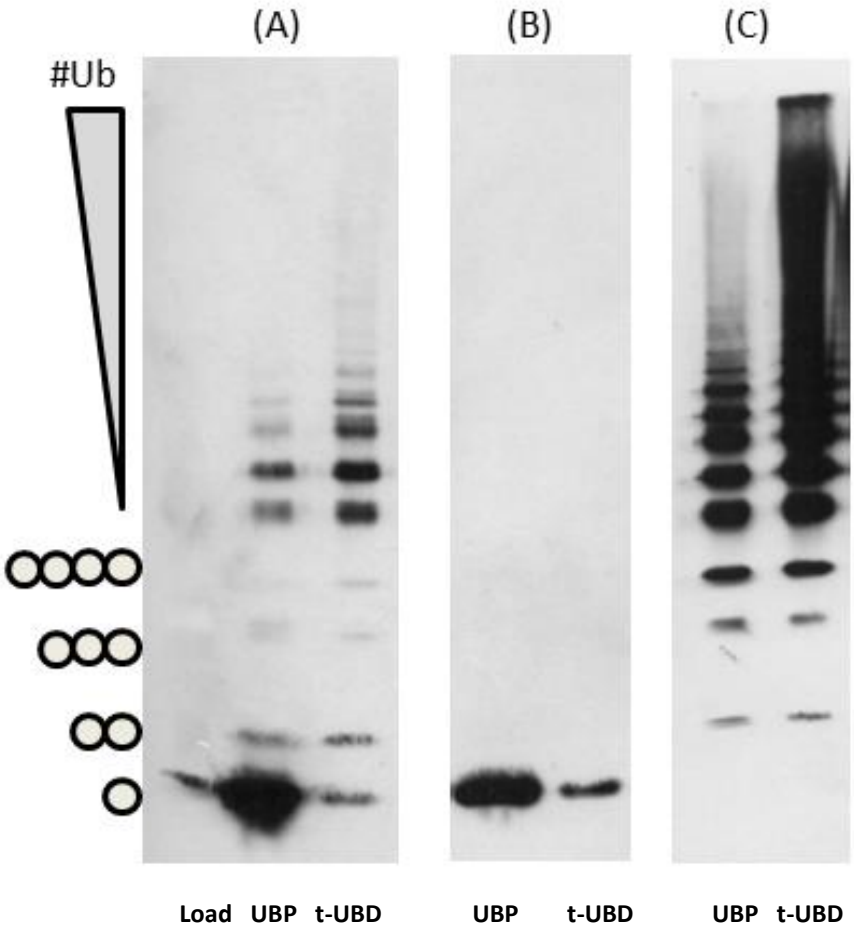


Figure 5.9 Application of the t-UBD to the purification of endogenous unanchored polyubiquitin. Representative anti-ubiquitin western blots showing endogenous unanchored polyubiquitin chains, enriched from mammalian HEK293T cell extracts. Following enrichment, 50 µL of ZnF-UBP (1 mg/ml, indicated as UBP) or t-UBD (eqimolar) Sepharose was eluted directly into GAB. 100% of each sample was separated by SDS-PAGE (on a 5-20% gradient gel). The gel was transferred to nitrocellulose, prior to immunoblotting versus (A) anti-ubiquitin (VU-1) or (C) a Lys48-linked polyubiquitin selective antibody (apu-2). Alongside the direct immunoblotting indicated in (A and C), following enrichment, 50 µL of the each Sepharose was incubated with WT USP5 (2 µg) for 16 hours at 37°C to assess deubiquitylation (B). Upon completion the reaction quenched via the addition of 20 µL of GAB and immunoblotted versus anti-ubiquitin (VU-1). Grey circles indicate numbers of ubiquitins (#Ub) within the chains (Adapted from [274]).

To confirm the identity of polyubiquitin linkages in the purified endogenous unanchored chain mixtures, as per *section 3.3*, following enrichment on ZnF-UBP or t-UBD Sepharose, unanchored polyubiquitin isolated from HEK293T cell extracts (typically from 200 µL of Sepharose) was eluted into 0.1% formic acid (pH 2) and digested using the *in house* developed FASP strategy. LC-MS/MS analyses of samples, captured in parallel on either ZnF-UBP (USP5) or t-UBD Sepharose, was completed and analysed on the LTQ OT Velos (at the University of Leicester). The ZnF-UBP enrichment, which represents a further replicate of the results presented in *section 3.3*, yielded signature peptides indicating the presence of all seven isopeptide linkages of unanchored polyubiquitin chains (Lys-6, 11, 27, 29, 33, 48 and 63), supporting the findings in *chapter 3*. Although we do not provide a quantitative description of relative abundance of signature peptides (and hence polyubiquitin linkage) in the endogenous unanchored polyubiquitin pool, the ~four-fold increase in the spectral count of Lys-48 signature peptides in the samples produced using the t-UBD (see Figure 5.10) strongly indicates linkage selectivity is maintained in these *in vitro* pull-downs. As per the enrichment on C335A-USP5 indicated in Figure 3.11, it is interesting to note that Lys-63 polymers again enriched on a tandem UBD construct when compared to the isolated ZnF-UBP domain. It is interesting to again speculate that this results from the enrichment of long mixed-linked polyubiquitin chains (including a Lys-63 linkage) owing to avidity effects of multi-UBD containing constructs. To conclude, while the Znf-UBP domain, as expected purifies all ubiquitin linkages, the t-UBD displays apparent and desired selectivity for Lys-48 linked polyubiquitin.

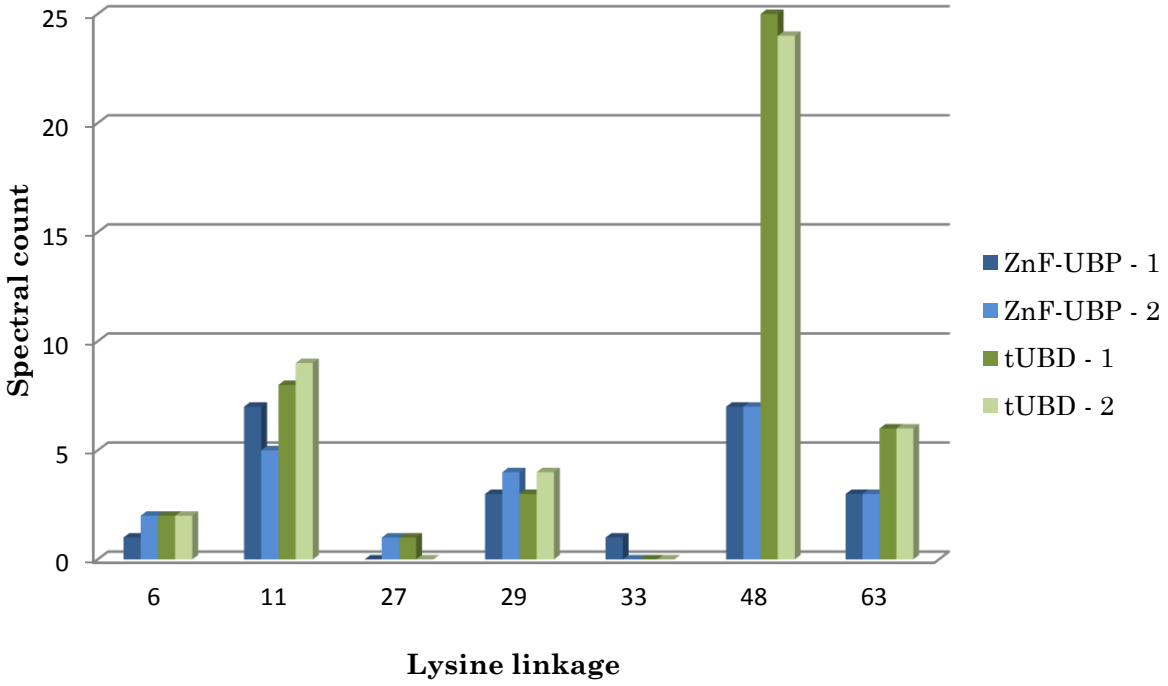


Figure 5.10 Confirming the specificity of t-UBD *in vitro*. LC-MS/MS analyses confirm the identity of the polyubiquitin linkages in the purified endogenous unanchored polyubiquitin pool from HEK293T cells, captured in parallel using the t-UBD or ZnF-UBP domain (USP5) (two biological replicas presented). Purified samples on beads as in Figure 5.9, prepared using equimolar t-UBD or ZnF-UBP, were subject to tryptic digestion and LC-MS/MS analysis to identify polyubiquitin signature peptides. After custom database searching, only peptides corresponding to ubiquitin were identified. The data was manually interrogated for peptides modified with GlyGly (+114.042927 on Lys) and Leu-Arg-Gly-Gly (+383.228103 on Lysine). The bars represent total spectral count of each polyubiquitin signature peptide during the analysis of each sample (Adapted from [274]).

5.6 Discussion

This chapter highlights that the design of UBD-based synthetic reagents, which exploit avidity effects to permit the recognition and purification of unanchored polyubiquitin chains of defined linkage is feasible. Here, by fusing a ZnF-UBP domain to the linkage-selective UBA2 of hHR23A, we favour the enrichment of unanchored Lys-48 linked ubiquitin polymers. The design strategy presented in this chapter builds upon the structural insights that natural linkage-selective tandem UBD interactions have afforded

to date; namely that avidity effects are a key factor in fine tuning polyubiquitin selectivity and can be used to disfavour binding to monomeric ubiquitin. Indeed utilising the lower affinity (than that of USP5) ZnF-UBP domain of USP16 in the t-UBD sequence (K_d *vs* monoubiquitin ≈ 6.5 μM), the enrichment of polyubiquitin is favoured over the capture of monoubiquitin in pull-downs from cell extracts.

Numerous lines of evidence support the notion that within the synthetic t-UBD construct both UBDs are functional individually, and critically work cooperatively to afford higher affinity interactions with unanchored Lys-48 linked diubiquitin. Most persuasively, native ESI-MS investigations demonstrate binding of the t-UBD to unanchored Lys-48 diubiquitin with an apparent K_d of ≈ 1 μM, which is greater than that associated with either of the individual UBDs in isolation. In contrast, binding of t-UBD to unanchored Lys-63 linked diubiquitin was of similar affinity to the isolated ZnF-UBP: monoubiquitin interaction, suggesting in this case simple recognition at the proximal ubiquitin in the dimer by the ZnF-UBP alone. Use of mutant monoubiquitin as ligand and mutagenesis of the ZnF-UBP domain within the t-UBD provided additional confirmation of the functionality of each UBD.

Under the conditions used, when applied to the enrichment of unanchored polyubiquitin from mammalian cell extracts, the t-UBD preferentially captures polyubiquitin over abundant monoubiquitin when compared to enrichment on the ZnF-UBP domain of USP5. Furthermore MS/MS analysis confirms that there is a ~four-fold increase in peptides (MS/MS spectral counts) corresponding to Lys-48 linked polyubiquitin on the t-UBD compared to the ZnF-UBP domain. Whilst we do not explore the utility beyond the initial purification of endogenous unanchored polyubiquitin chains from mammalian extracts, owing the sequence conservation of ubiquitin in eukaryotes it is expected that the t-UBD would be broadly applicable to other organisms. Indeed previous studies highlight the purification of unanchored polyubiquitin chains utilising the isolated ZnF-UBP domain from a range of sources [246]. Furthermore it is possible

to envisage adapting such a tool to monitor dynamic changes in ubiquitylation patterns in live cells or as transient inhibitors of ubiquitin-dependent pathways *in vivo*, as we allude to in Figure 1.7 [64].

Reassuringly, our modelling approach led to the optimum flexible linker length (with longer or shorter linkers associated with weaker binding), although it would be of interest to explore the effects on binding affinity of substituting rigid helical linkers to more tightly constrain the tandem UBD arrangement. As noted previously the magnitude of the avidity effect generated (an ≈ 11-fold preference of the t-UBD for unanchored Lys-48 diubiquitin over monoubiquitin) may appear relatively modest, but is of a similar magnitude to that for the naturally occurring tandem UIMs in RAP80, which display an ≈ 24-fold preference for Lys-63 linked diubiquitin over monoubiquitin [81]. Both the designed (t-UBD) and these natural (RAP80) UBDs do however represent the minimal tandem array, and indeed much higher affinity binding can be achieved by placing multiple UBDs in tandem, as was achieved by Sims and co-workers using three identical UIMs in a single structured α-helix [81]. This similarly explains the high affinity (non-selective) polyubiquitin-binding properties of TUBEs (consisting of a tandem array of flexibly linked UBA domains). Accordingly further avidity effects maybe realised by extending the design of the t-UBD to incorporate further UBDs, promoting interactions which favour the recognition of longer unanchored polyubiquitin chains.

Given that our MS/MS analyses using the linkage-independent isolated ZnF-UBP domain of USP5 alongside the t-UBD revealed that the majority of ubiquitin isopeptide linkages (Lys-6, 11, 27, 29, 33, 48 and 63) participate in unanchored polyubiquitin chain formation in mammalian cells, we anticipate substituting other linkage-selective UBDs for the UBA2, to permit the recognition, detection and purification of unanchored chains of all linkage types, allowing further dissection of the linkage-selective roles of unanchored polyubiquitin in the regulation of cellular physiology. As an exemplar,

Figure 5.11 highlights the substitution of the UBA2 for the NZF of TAB2, to theoretically favour the purification of unanchored Lys-63 linked polyubiquitin chains.

In conclusion this chapter provides the first proof-of-concept that polyubiquitin-selective recognition can be achieved using novel assemblies of differing UBDs in tandem, with optimised linkage-selective avidity effects. Accordingly we conclude that the rational design of UBD hybrids with defined unanchored polyubiquitin selectivity is feasible, and as we allude to in Figure 5.11, such a design strategy is amenable to evolution through further iterations.

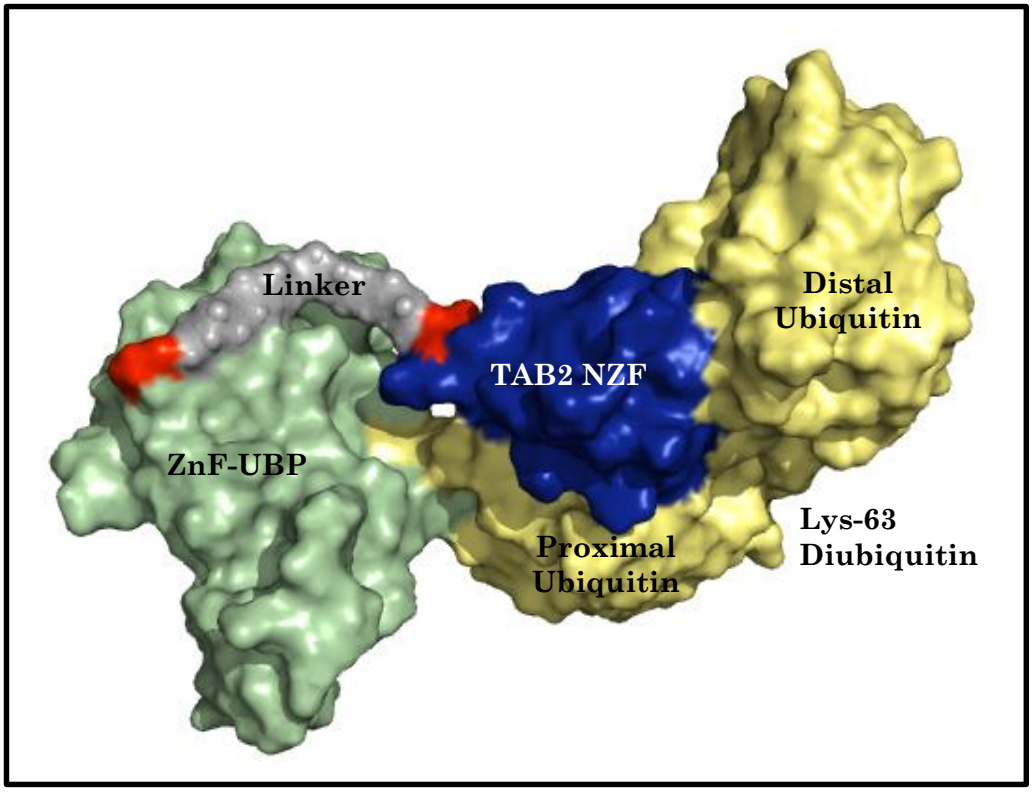


Figure 5.11 Designing a reagent for the selective enrichment of unanchored Lys-63 linked polyubiquitin. The ZnF-UBP of USP16 (pale green) is docked onto the free C-terminus of the proximal ubiquitin of the TAB2 NZF (blue):Lys-63 diubiquitin (yellow) complex. The ZnF-UBP is N-terminal, followed by an 8xGly linker (grey) and then the TAB2 NZF.

Chapter 6 - General discussion

The notion that ubiquitin signalling is mediated solely by the covalent conjugation of (poly)ubiquitin to selected protein substrates has, in recent years, been challenged. Pioneering studies by Zhijian Chen's group revealed unanchored, or substrate-free (poly)ubiquitin, specifically linked by Lys-63, to be potent regulators of the innate immune response, via direct activation of ubiquitin-dependent kinases [11, 219]. Subsequent studies have further established the physiological significance of this unanchored (poly)ubiquitin pool more broadly in innate immune signalling, as well as at the 26S proteasome and in aggresome biogenesis [200, 204, 210-213, 218]. Critically, just as the 'ubiquitin code' emerged for substrate conjugated ubiquitin, wherein Lys-linkage determines fate of the target protein, the notion has been seemingly emerged for the unanchored polyubiquitin pool, with unanchored Lys-48 linked chains linked to roles at the proteasome [250], conversely Lys-63 linked polymers have been linked to innate immune signalling. Furthermore chains linked by alternate Lys residues have been alluded to, and indeed recently mixed linked polyubiquitin chains have been reported [224].

However, tools that permit purification and thus characterisation of this unanchored polyubiquitin were lacking, indeed many studies sought to indirectly infer the unanchored nature of polyubiquitin, by predominantly utilising *in vitro* disassembly assays with the DUB responsible for selectively processing unanchored polyubiquitin, USP5. Consequently prior to the start of the investigations in this thesis, the work of colleagues in the Layfield group sought to establish a strategy to permit the selective purification of endogenous unanchored polyubiquitin. This original free-ubiquitin

binding entity (FUBE) purification strategy exploited the inherent selectivity of the ZnF-UBP domain (isolated from USP5) for the free C-terminus of unanchored ubiquitin, to selectively purify endogenous unanchored (poly)ubiquitin from rat skeletal muscle, and subsequently plant and yeast extracts, as well as a range of mammalian cell lines. Initial insights demonstrated the purification of unanchored polyubiquitin chains, ranging in size, and initial limited MS analyses revealed the diubiquitin pool to be minimally composed of Lys-11 and 48 linked polymers.

The investigations presented in this thesis have sought to extend the initial insights afforded by this strategy, to evolve towards the linkage selective purification of defined polyubiquitin topologies.

6.1 Purifying endogenous unanchored polyubiquitin from HEK293T cell extracts

In *chapter 3*, the original FUBE enrichment strategy was presented and applied to purify endogenous unanchored polyubiquitin chains from, for the first time, mammalian (HEK293T) cell extracts. Initial immunoblotting results supported previous findings (in rat skeletal muscle and (U20S and HeLa) mammalian cell lines) that endogenous unanchored polyubiquitin chains of >10 moieties in length could be purified and detected. Additionally immunoblotting with available linkage selective antibodies confirmed that the HEK293T unanchored polyubiquitin pool includes Lys-48 and 63 linked unanchored polyubiquitin chains. This supports the findings of Strachan *et al* who previously identified Lys-11 and 48 linked polymers [107, 246], and Zeng *et al* who purified unanchored Lys-63 linked polyubiquitin [219].

However the purification of unanchored (poly)ubiquitin, utilising the ZnF-UBP domain, is independent of polyubiquitin chain linkage and length, and thus does not distinguish abundant monoubiquitin from diverse polyubiquitin signals, which competes for binding to the ZnF-UBP domain and impedes downstream proteomic

analyses. The recent introduction of tandem arrays of the same non-selective UBD, such as TUBEs or subtle variations thereof, has revolutionised the bulk affinity purification of ubiquitin-modified substrates, permitting high affinity recognition and purification when compared to the use of isolated UBDs [110]. Accordingly as a natural extension of the FUBE strategy, to favour the purification of polyubiquitin over abundant monoubiquitin, we utilised the multi-UBD containing USP5 protein. In addition to containing a ZnF-UBP domain (used in the FUBE), responsible for the recognition of unanchored free C-termini of ubiquitin, USP5 is also composed of a catalytic domain which displays affinity for ubiquitin, as well as two additional UBA domains, which although in isolation display weak affinity for ubiquitin, contribute to the high affinity recognition of polyubiquitin by avidity effects [87, 237].

Under the conditions established in the initial FUBE enrichment strategy, in which the FUBE is in excess, we demonstrated that purifications utilising both the isolated ZnF-UBP and USP5 (catalytic mutant – C335A) constructs display similar ubiquitin immunoreactive smears (based on western blot migration), again indicating the purification of long polyubiquitin chains (>10 moieties). However under conditions where the bait (USP5) is limiting, we demonstrate that, presumably owing to avidity affects, we favour the purification of unanchored polyubiquitin over abundant monoubiquitin – as indicated in Figure 3.9.

To probe the polyubiquitin linkage profile of the HEK293T cell unanchored polyubiquitin pool an MS/MS analysis was completed *in house* following enrichment of HEK293T cell extracts on ZnF-UBP or C335A-USP5 Sepharose. This revealed polyubiquitin signature tryptic peptides indicative of Lys-11, 48 and 63 linkages. This supports previous, limited MS/MS analyses which identified the presence of ubiquitin signature peptides corresponding to Lys-11 and 48 linked polyubiquitin in rat skeletal muscle [107]. Previous quantitative studies of the broader ubiquitome, have identified Lys-11, 48 and 63 linked polyubiquitin to be the most abundant in mammalian cell

models [34-36], accordingly the *in house* analysis presented here indicates this may also be case for the unanchored polyubiquitin proteome, albeit the caveat is that analysis is purely qualitative, indicating the polyubiquitin signature peptides detected.

To validate these findings, parallel polyubiquitin linkage analysis of ZnF-UBP or C335A-USP5 enriched unanchored polyubiquitin was completed in collaboration with Dr Sharad Mistry and Dr Andrew Bottrill at the University of Leicester. MS/MS analysis of this material revealed, for the first time in Figure 3.11, that the unanchored polyubiquitin pool is composed of all seven isopeptide linked ubiquitin polymers in purified samples from HEK293T. As previously only Lys-11, 48 and 63 had been identified in the endogenous unanchored polyubiquitin pool, identification of all seven isopeptide linked polymers indicates that unanchored polyubiquitin-mediated processes may be more broadly regulated by chains of diverse linkages. Indeed it will be of interest to determine if certain linkages are dedicated to specific chemical processes, as is the case for conjugated polyubiquitin.

6.2 Characterising the mechanism of (poly)ubiquitin recognition by USP5

The MS/MS findings in *chapter 3* demonstrated that USP5 is capable of binding to the seven isopeptide linked ubiquitin polymers present in the HEK293T unanchored ubiquitin pool. This supports previous findings that USP5 can cleave all eight ubiquitin polymers *in vitro* [247]. As alluded to, USP5 contains four ubiquitin binding sites, which underlie USP5 selectivity for unanchored polyubiquitin, however a recent crystal structure of USP5 has raised a number of questions regarding the orientation of the domains of USP5, with respect to polyubiquitin recognition [237], indeed the ZnF-UBP domain appears to point away from the active site.

To probe the mechanism of polyubiquitin recognition by USP5 in comparison to this crystal structure, we applied emerging structural MS approaches including native IM-

MS (alongside MD simulations) and diazirine-based protein carbene footprinting. The application of native IM-MS to the large (~94 kDa) USP5 protein highlighted that USP5 adopts a number of conformers, which broadly correlates to a major compact conformation, which binds to (poly)ubiquitin, and a minor extended conformation, which does not. Critically we validated these gas-phase structures by utilising MD simulations, which draw comparison to the USP5 X-ray structure, and indeed show good agreement.

To probe USP5-ubiquitin interactions we completed a quantitative titration analysis utilising both monoubiquitin and USP5 mutants. Here, in *section 4.2*, we confirmed the binding sites of USP5 with monoubiquitin, and revealed a co-operative mechanism of recognition, wherein binding of proximal ubiquitin at the ZnF-UBP domain permits high affinity recognition of the distal ubiquitin moiety at the active site, while subsequent ubiquitin moieties, in a polyubiquitin chain, are recognised by the UBA domains of USP5. It is noteworthy that binding at the ZnF-UBP by WT monoubiquitin yields an apparent approximate K_d of 2.8 μM, reassuringly this value is in excellent agreement with the ZnF-UBP-monoubiquitin interaction as measured by other biophysical techniques [84]. The most significant finding, with respect to USP5-ubiquitin interactions, indicated in Figure 4.20, was that upon introduction of increasing concentrations of Lys-63 linked tetraubiquitin to USP5, the conformer equilibrium shifted from extended, to compact bound, indicating that USP5 conformer populations are interconvertible.

By utilising carbene footprinting, with *in house* developed reagents, to probe the USP5-diubiquitin footprint, we confirmed that the two sites of diubiquitin recognition correspond to the ZnF-UBP domain and the catalytic site (see Figure 4.19). This supports previous investigations by Reyes-Turcu *et al*, who using a combination of USP5 mutagenesis and ITC, identified and quantified the sites of USP5-ubiquitin recognition [87]. We conclude that the recent USP5 X-ray structure is different to that

which is responsible for binding diubiquitin. Interestingly, the limited structural studies of alternate USP family members, have indicated that alternate USP family members undergo drastic conformational changes upon ubiquitin binding [258, 259]. Accordingly, the inherent conformational flexibility of USP5 noted here, may more broadly reflect the USP-ubiquitin recognition mechanism, however few of the 56 USP family members have been biochemically characterised so this remains to be tested. Taken together, the structural MS findings support the notion that USP5 displays inherent flexibility to permit the recognition of diverse polyubiquitin structures.

The examples presented in *chapter 4* also demonstrate the clear advantages of applying structural MS based approaches to probe protein structure and interactions, and the complementary data obtained by these approaches in comparison to X-ray crystallography, ITC and SAXS analysis. Indeed MS approaches offer rapid and sensitive insights into the structural integrity and binding dynamics of large conformationally dynamic proteins. For example, native IM-MS permits clear visualisation of all discrete species present, in this case including the two USP5 CSDs, which highlight USP5 conformational dynamics, as well as signals corresponding to each ubiquitin binding event, which reveals the positive cooperativity of binding. Furthermore these discrete signals and their intensities can be utilised to quantitate binding affinity, with experimentally determined binding affinities showing good agreement with previous literature values, albeit native ESI-MS proves favourable in relation to time and sample consumption.

The carbene footprinting strategy presented a novel platform to probe USP5-ubiquitin interactions at relatively high resolution, indeed highlighting inaccuracies in the previously reported X-ray structure, that arise owing to the structural snapshot which this provides. We envisage this footprinting approach having broad utility to other protein-ligand systems, particularly in the characterisation of proteins/protein complexes which are large or dynamic, and thus have proved challenging to interrogate

by other biophysical techniques. As an exemplar in recent years DUBs have emerged as attractive therapeutic targets, owing to implication in the pathology of a number of diseases [276]. Accordingly the structural MS strategies presented offer a high-throughput strategy to screen the interactions of DUBs, with protein or small molecule modulators, in a rapid and sensitive manner.

6.3 Designing an artificial tandem UBD construct for the selective purification of Lys-48 linked unanchored polyubiquitin

The findings in *chapter 4* confirmed the notion that the inherent flexibility of USP5, in which the four UBDs are joined by large flexible loops, permits the recognition of differing polyubiquitin topologies. Interestingly Sims *et al* demonstrated that by altering linker length and its structural integrity between tandem UBDs (UIMs of RAP80), the interaction surface can be tuned, controlling UIM orientation and the relative phasing of binding sites along the helix, to favour interactions with defined ubiquitin topologies [81]. Accordingly we reasoned that by utilising UBDs with intrinsic selectivity joined in a linker optimised manner, then specificity for defined ubiquitin topologies could be realised.

The simplest polymer of ubiquitin is a dimer, furthermore with regards to unanchored diubiquitin, we have previously demonstrated the presence of diubiquitin *in vivo* and MS/MS investigations demonstrated these dimers to be linked at least by Lys-11 and 48 [107]. Accordingly, as a proof-of-concept that differing UBDs can be arranged in tandem to yield a reagent that selectively purifies defined ubiquitin topologies, in *chapter 5* we described the design of a linker-optimised UBD hybrid (t-UBD) containing two UBDs. We utilised the existing molecular structures of the ZnF-UBP domain (USP16) and the UBA2 of hHR23A, a Lys-48 linkage-selective UBA domain, combined with MD simulations to generate a construct plausible for the recognition of unanchored Lys-48 linked dimers and longer polyubiquitin chains. Utilising native MS

to quantitatively probe binding affinities we confirm cooperative binding of the UBDs within the synthetic protein, and desired binding specificity for Lys-48 linked ubiquitin dimers over Lys-63 linked diubiquitin (see Figure 5.6). Furthermore MS/MS analyses indicate that the t-UBD, when applied as an affinity enrichment reagent, can be used to favour the purification of endogenous unanchored Lys48-linked polyubiquitin chains from mammalian cell extracts over other linkages. Accordingly we can conclude that this construct, for the first time, utilises two different UBDs fused in a linker optimised manner, to exploit avidity effects that afford the recognition and purification of endogenous Lys48-linked unanchored polyubiquitin chains.

6.4 Future work

As alluded to in *chapter 5*, a future challenge is the development of bio-inspired synthetic reagents for the selective detection and purification of polyubiquitin chains of defined chain linkage/length, and which distinguish unanchored from conjugated polyubiquitin. In *chapter 5*, we presented the t-UBD, which we demonstrate displays selectivity against Lys-48 linked unanchored polyubiquitin. Providing a structure is available, which is the case for linkage selective UBDs structurally characterised to date, then MD simulations can be performed to determine optimal arrangements of tandem reagents, thus permitting substitution of domains. However, whilst this strategy is plausible, only a handful of natural linkage selective UBDs have been identified. Accordingly the design of novel 'tools' to expand the repertoire of ubiquitin recognition motifs with high intrinsic specificity for polyubiquitin linkages in the absence natural linkage selective UBDs remains a priority, and would certainly provide valuable biochemical probes for determining distinct cellular functions.

Chapter 5 indicates that UBDs can act as a tool-box, and are essentially amenable to protein engineering approaches. Furthermore a common feature of many UBDs including UBA, UIM/MIU and A20 motifs is their high helical content and rigid

structures. Thus we propose that by fusing together combinations of natural UBDs through continuous helix extensions and carefully optimising geometrical constraints, the multi-UBD constructs will present distinctive ubiquitin binding surfaces to select unique conformations adopted by different polyubiquitin linkage types. Exemplars of this approach are indicated in Figure 6.1, wherein theoretical tandem UBD arrangements of ZnF-UBP, UBA and MIU/UIM for Lys-6 and a UIM-A20 fusion for Lys-29 linked polyubiquitin selectivity are presented. As alluded to in **Error! Reference source not found.**, in addition to the purification of ubiquitin modifications, such UBD-based tools could have wider application as *in vivo* biosensors or indeed as competitive inhibitors of ubiquitin-mediated processes.

Finally, whilst not addressed in the investigations in this thesis, the use of cultured mammalian cells permits pharmacological or genetic intervention, which alongside the development of an *in house* quantitative MS approach, such as the ubiquitin-AQUA strategy, as well as linkage selective reagents, will permit a thorough analysis of the molecular composition and thus significance of the unanchored polyubiquitin pool, as well as more broadly the ubiquitome.

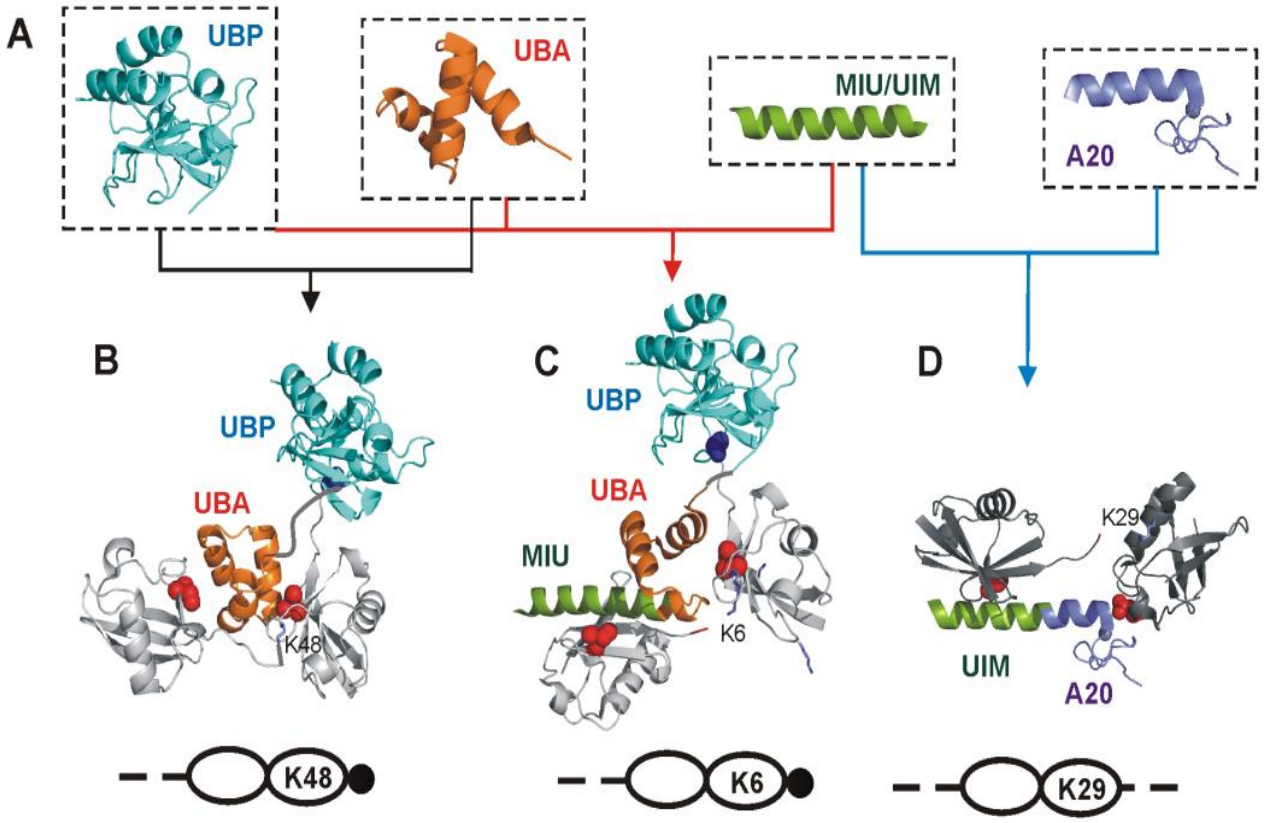


Figure 6.1 Designing multi-UBD constructs with unique polyubiquitin selectivity. The UBD 'tool box' is for example composed of (A) the ZnF-UBP, UBA, MIU/UIM and the ZnF-A20 domains. Structures (B-D) represent arrangements of the UBDs in (A). The structure in (B) is that presented in chapter 5, representing the t-UBD. (C) Represents an arrangement of the ZnF-UBP, UBA and MIU/UIM domains to afford the selective purification of unanchored Lys-6 linked polyubiquitin. (D) Represents a UIM-A20 fusion compatible with Lys-29 polyubiquitin recognition.

Chapter 7 – References

[1] Walsh, C. T., Garneau-Tsodikova, S., Gatto, G. J., Jr., Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew Chem Int Ed Engl* 2005, *44*, 7342-7372.

[2] Seet, B. T., Dikic, I., Zhou, M. M., Pawson, T., Reading protein modifications with interaction domains. *Nat Rev Mol Cell Biol* 2006, *7*, 473-483.

[3] Ciechanover, A., Hod, Y., Hershko, A., A heat-stable polypeptide component of an ATP-dependent proteolytic system from reticulocytes. 1978. *Biochem Biophys Res Commun* 2012, *425*, 565-570.

[4] Ciechanover, A., Heller, H., Elias, S., Haas, A. L., Hershko, A., ATP-dependent conjugation of reticulocyte proteins with the polypeptide required for protein degradation. *Proc Natl Acad Sci U S A* 1980, *77*, 1365-1368.

[5] Ciechanover, A., Proteolysis: from the lysosome to ubiquitin and the proteasome. *Nat Rev Mol Cell Biol* 2005, *6*, 79-87.

[6] Hochstrasser, M., Varshavsky, A., In vivo degradation of a transcriptional regulator: the yeast alpha 2 repressor. *Cell* 1990, *61*, 697-708.

[7] Glotzer, M., Murray, A. W., Kirschner, M. W., Cyclin is degraded by the ubiquitin pathway. *Nature* 1991, *349*, 132-138.

[8] Hoege, C., Pfander, B., Moldovan, G. L., Pyrowolakis, G., Jentsch, S., RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. *Nature* 2002, *419*, 135-141.

[9] Spence, J., Sadis, S., Haas, A. L., Finley, D., A ubiquitin mutant with specific defects in DNA repair and multiubiquitination. *Mol Cell Biol* 1995, *15*, 1265-1273.

[10] Kaiser, P., Flick, K., Wittenberg, C., Reed, S. I., Regulation of transcription by ubiquitination without proteolysis: Cdc34/SCF(Met30)-mediated inactivation of the transcription factor Met4. *Cell* 2000, *102*, 303-314.

[11] Xia, Z. P., Sun, L., Chen, X., Pineda, G.*, et al.*, Direct activation of protein kinases by unanchored polyubiquitin chains. *Nature* 2009, *461*, 114-119.

[12] Scheffner, M., Werness, B. A., Huibregtse, J. M., Levine, A. J., Howley, P. M., The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. *Cell* 1990, *63*, 1129-1136.

[13] Maxwell, P. H., Wiesener, M. S., Chang, G. W., Clifford, S. C.*, et al.*, The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis. *Nature* 1999, *399*, 271-275.

[14] Lam, Y. A., Pickart, C. M., Alban, A., Landon, M.*, et al.*, Inhibition of the ubiquitin-proteasome system in Alzheimer's disease. *Proc Natl Acad Sci U S A* 2000, *97*, 9902-9906.

[15] Alves-Rodrigues, A., Gregori, L., Figueiredo-Pereira, M. E., Ubiquitin, cellular inclusions and their role in neurodegeneration. *Trends Neurosci* 1998, *21*, 516-520.
[16] Komander, D., Rape, M., The ubiquitin code. *Annu Rev Biochem* 2012, *81*, 203-229.
[17] Hershko, A., Heller, H., Elias, S., Ciechanover, A., Components of ubiquitin-protein ligase system. Resolution, affinity purification, and role in protein breakdown. *J Biol Chem* 1983, *258*, 8206-8214.
[18] Pickart, C. M., Mechanisms underlying ubiquitination. *Annu Rev Biochem* 2001, *70*, 503-533.
[19] Morreale, F. E., Walden, H., Types of Ubiquitin Ligases. *Cell* 2016, *165*, 248-248 e241.
[20] Berndsen, C. E., Wolberger, C., New insights into ubiquitin E3 ligase mechanism. *Nat Struct Mol Biol* 2014, *21*, 301-307.
[21] Hoppe, T., Multiubiquitylation by E4 enzymes: 'one size' doesn't fit all. *Trends Biochem Sci* 2005, *30*, 183-187.
[22] Windheim, M., Peggie, M., Cohen, P., Two different classes of E2 ubiquitin-conjugating enzymes are required for the mono-ubiquitination of proteins and elongation by polyubiquitin chains with a specific topology. *Biochem J* 2008, *409*, 723-729.
[23] Terrell, J., Shih, S., Dunn, R., Hicke, L., A function for monoubiquitination in the internalization of a G protein-coupled receptor. *Mol Cell* 1998, *1*, 193-202.
[24] Shih, S. C., Katzmann, D. J., Schnell, J. D., Sutanto, M.*, et al.*, Epsins and Vps27p/Hrs contain ubiquitin-binding domains that function in receptor endocytosis. *Nat Cell Biol* 2002, *4*, 389-393.
[25] Haglund, K., Sigismund, S., Polo, S., Szymkiewicz, I.*, et al.*, Multiple monoubiquitination of RTKs is sufficient for their endocytosis and degradation. *Nat Cell Biol* 2003, *5*, 461-466.
[26] Zhang, Y., Transcriptional regulation by histone ubiquitination and deubiquitination. *Genes Dev* 2003, *17*, 2733-2740.
[27] Hough, R., Pratt, G., Rechsteiner, M., Ubiquitin-lysozyme conjugates. Identification and characterization of an ATP-dependent protease from rabbit reticulocyte lysates. *J Biol Chem* 1986, *261*, 2400-2408.
[28] Chau, V., Tobias, J. W., Bachmair, A., Marriott, D.*, et al.*, A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. *Science* 1989, *243*, 1576-1583.
[29] Thrower, J. S., Hoffman, L., Rechsteiner, M., Pickart, C. M., Recognition of the polyubiquitin proteolytic signal. *EMBO J* 2000, *19*, 94-102.
[30] McCullough, J., Row, P. E., Lorenzo, O., Doherty, M.*, et al.*, Activation of the endosome-associated ubiquitin isopeptidase AMSH by STAM, a component of the multivesicular body-sorting machinery. *Curr Biol* 2006, *16*, 160-165.
[31] Moldovan, G. L., Pfander, B., Jentsch, S., PCNA, the maestro of the replication fork. *Cell* 2007, *129*, 665-679.
[32] Skaug, B., Jiang, X., Chen, Z. J., The role of ubiquitin in NF-kappaB regulatory pathways. *Annu Rev Biochem* 2009, *78*, 769-796.
[33] Tan, J. M., Wong, E. S., Kirkpatrick, D. S., Pletnikova, O.*, et al.*, Lysine 63-linked ubiquitination promotes the formation and autophagic clearance of protein inclusions associated with neurodegenerative diseases. *Hum Mol Genet* 2008, *17*, 431-439.

[34] Peng, J., Schwartz, D., Elias, J. E., Thoreen, C. C., *et al.*, A proteomics approach to understanding protein ubiquitination. *Nat Biotechnol* 2003, *21*, 921-926.
[35] Xu, P., Duong, D. M., Seyfried, N. T., Cheng, D., *et al.*, Quantitative proteomics reveals the function of unconventional ubiquitin chains in proteasomal degradation. *Cell* 2009, *137*, 133-145.
[36] Kirisako, T., Kamei, K., Murata, S., Kato, M., *et al.*, A ubiquitin ligase complex assembles linear polyubiquitin chains. *EMBO J* 2006, *25*, 4877-4887.
[37] Bennett, E. J., Bence, N. F., Jayakumar, R., Kopito, R. R., Global impairment of the ubiquitin-proteasome system by nuclear or cytoplasmic protein aggregates precedes inclusion body formation. *Mol Cell* 2005, *17*, 351-365.
[38] Matsumoto, M. L., Wickliffe, K. E., Dong, K. C., Yu, C., *et al.*, K11-linked polyubiquitination in cell cycle control revealed by a K11 linkage-specific antibody. *Mol Cell* 2010, *39*, 477-484.
[39] Haas, T. L., Emmerich, C. H., Gerlach, B., Schmukle, A. C., *et al.*, Recruitment of the linear ubiquitin chain assembly complex stabilizes the TNF-R1 signaling complex and is required for TNF-mediated gene induction. *Mol Cell* 2009, *36*, 831-844.
[40] Gerlach, B., Cordier, S. M., Schmukle, A. C., Emmerich, C. H., *et al.*, Linear ubiquitination prevents inflammation and regulates immune signalling. *Nature* 2011, *471*, 591-596.
[41] Tokunaga, F., Nakagawa, T., Nakahara, M., Saeki, Y., *et al.*, SHARPIN is a component of the NF-kappaB-activating linear ubiquitin chain assembly complex. *Nature* 2011, *471*, 633-636.
[42] Tokunaga, F., Sakata, S., Saeki, Y., Satomi, Y., *et al.*, Involvement of linear polyubiquitylation of NEMO in NF-kappaB activation. *Nat Cell Biol* 2009, *11*, 123-132.
[43] Wickliffe, K. E., Williamson, A., Meyer, H. J., Kelly, A., Rape, M., K11-linked ubiquitin chains as novel regulators of cell division. *Trends Cell Biol* 2011, *21*, 656-663.
[44] Kulathu, Y., Komander, D., Atypical ubiquitylation - the unexplored world of polyubiquitin beyond Lys48 and Lys63 linkages. *Nat Rev Mol Cell Biol* 2012, *13*, 508-523.
[45] Michel, M. A., Elliott, P. R., Swatek, K. N., Simicek, M., *et al.*, Assembly and specific recognition of k29- and k33-linked polyubiquitin. *Mol Cell* 2015, *58*, 95-109.
[46] Yuan, W. C., Lee, Y. R., Lin, S. Y., Chang, L. Y., *et al.*, K33-Linked Polyubiquitination of Coronin 7 by Cul3-KLHL20 Ubiquitin E3 Ligase Regulates Protein Trafficking. *Mol Cell* 2014, *54*, 586-600.
[47] Durcan, T. M., Tang, M. Y., Perusse, J. R., Dashti, E. A., *et al.*, USP8 regulates mitophagy by removing K6-linked ubiquitin conjugates from parkin. *EMBO J* 2014, *33*, 2473-2491.
[48] Gatti, M., Pinato, S., Maiolica, A., Rocchio, F., *et al.*, RNF168 promotes noncanonical K27 ubiquitination to signal DNA damage. *Cell Rep* 2015, *10*, 226-238.
[49] Morris, J. R., Solomon, E., BRCA1 : BARD1 induces the formation of conjugated ubiquitin structures, dependent on K6 of ubiquitin, in cells during DNA replication and repair. *Hum Mol Genet* 2004, *13*, 807-817.
[50] Walczak, H., Iwai, K., Dikic, I., Generation and physiological roles of linear ubiquitin chains. *BMC Biol* 2012, *10*, 23.

[51] Behrends, C., Harper, J. W., Constructing and decoding unconventional ubiquitin chains. *Nat Struct Mol Biol* 2011, *18*, 520-528.
[52] Swatek, K. N., Komander, D., Ubiquitin modifications. *Cell Res* 2016, *26*, 399-422.
[53] Grice, G. L., Lobb, I. T., Weekes, M. P., Gygi, S. P.*, et al.*, The Proteasome Distinguishes between Heterotypic and Homotypic Lysine-11-Linked Polyubiquitin Chains. *Cell Rep* 2015, *12*, 545-553.
[54] Meyer, H. J., Rape, M., Enhanced protein degradation by branched ubiquitin chains. *Cell* 2014, *157*, 910-921.
[55] Emmerich, C. H., Ordureau, A., Strickson, S., Arthur, J. S.*, et al.*, Activation of the canonical IKK complex by K63/M1-linked hybrid ubiquitin chains. *Proc Natl Acad Sci U S A* 2013, *110*, 15247-15252.
[56] Kristariyanto, Y. A., Abdul Rehman, S. A., Campbell, D. G., Morrice, N. A.*, et al.*, K29-selective ubiquitin binding domain reveals structural basis of specificity and heterotypic nature of k29 polyubiquitin. *Mol Cell* 2015, *58*, 83-94.
[57] Kim, H. T., Kim, K. P., Uchiki, T., Gygi, S. P., Goldberg, A. L., S5a promotes protein degradation by blocking synthesis of nondegradable forked ubiquitin chains. *EMBO J* 2009, *28*, 1867-1877.
[58] Hendriks, I. A., D'Souza, R. C., Yang, B., Verlaan-de Vries, M.*, et al.*, Uncovering global SUMOylation signaling networks in a site-specific manner. *Nat Struct Mol Biol* 2014, *21*, 927-936.
[59] Fan, J. B., Arimoto, K., Motamedchaboki, K., Yan, M.*, et al.*, Identification and characterization of a novel ISG15-ubiquitin mixed chain and its role in regulating protein homeostasis. *Sci Rep* 2015, *5*, 12704.
[60] Kim, W., Bennett, E. J., Huttlin, E. L., Guo, A.*, et al.*, Systematic and quantitative assessment of the ubiquitin-modified proteome. *Mol Cell* 2011, *44*, 325-340.
[61] Ohtake, F., Saeki, Y., Sakamoto, K., Ohtake, K.*, et al.*, Ubiquitin acetylation inhibits polyubiquitin chain elongation. *EMBO Rep* 2015, *16*, 192-201.
[62] Koyano, F., Okatsu, K., Kosako, H., Tamura, Y.*, et al.*, Ubiquitin is phosphorylated by PINK1 to activate parkin. *Nature* 2014.
[63] Husnjak, K., Dikic, I., Ubiquitin-binding proteins: decoders of ubiquitin-mediated cellular functions. *Annu Rev Biochem* 2012, *81*, 291-322.
[64] Scott, D., Oldham, N. J., Strachan, J., Searle, M. S., Layfield, R., Ubiquitin-binding domains: mechanisms of ubiquitin recognition and use as tools to investigate ubiquitin-modified proteomes. *Proteomics* 2015, *15*, 844-861.
[65] Scott, D., Oldham, N. J., Strachan, J., Searle, M. S., Layfield, R., Ubiquitin-binding domains: Mechanisms of ubiquitin recognition and use as tools to investigate ubiquitin-modified proteomes. *Proteomics* 2015, *15*, 844-861.
[66] Garner, T. P., Strachan, J., Shedden, E. C., Long, J. E.*, et al.*, Independent interactions of ubiquitin-binding domains in a ubiquitin-mediated ternary complex. *Biochemistry* 2011, *50*, 9076-9087.
[67] Kristariyanto, Y. A., Choi, S. Y., Rehman, S. A., Ritorto, M. S.*, et al.*, Assembly and structure of Lys33-linked polyubiquitin reveals distinct conformations. *Biochem J* 2015, *467*, 345-352.
[68] Bremm, A., Freund, S. M., Komander, D., Lys11-linked ubiquitin chains adopt compact conformations and are preferentially hydrolyzed by the deubiquitinase Cezanne. *Nat Struct Mol Biol* 2010, *17*, 939-947.

[69] Virdee, S., Ye, Y., Nguyen, D. P., Komander, D., Chin, J. W., Engineered diubiquitin synthesis reveals Lys29-isopeptide specificity of an OTU deubiquitinase. *Nat Chem Biol* 2010, *6*, 750-757.
[70] Varadan, R., Assfalg, M., Raasi, S., Pickart, C., Fushman, D., Structural determinants for selective recognition of a Lys48-linked polyubiquitin chain by a UBA domain. *Mol Cell* 2005, *18*, 687-698.
[71] Komander, D., Reyes-Turcu, F., Licchesi, J. D., Odenwaelder, P.*, et al.*, Molecular discrimination of structurally equivalent Lys 63-linked and linear polyubiquitin chains. *EMBO Rep* 2009, *10*, 466-473.
[72] Varadan, R., Walker, O., Pickart, C., Fushman, D., Structural properties of polyubiquitin chains in solution. *J Mol Biol* 2002, *324*, 637-647.
[73] Ye, Y., Blaser, G., Horrocks, M. H., Ruedas-Rama, M. J.*, et al.*, Ubiquitin chain conformation regulates recognition and activity of interacting proteins. *Nature* 2012, *492*, 266-270.
[74] Kulathu, Y., Akutsu, M., Bremm, A., Hofmann, K., Komander, D., Two-sided ubiquitin binding explains specificity of the TAB2 NZF domain. *Nat Struct Mol Biol* 2009, *16*, 1328-1330.
[75] Sato, Y., Yoshikawa, A., Yamashita, M., Yamagata, A., Fukai, S., Structural basis for specific recognition of Lys 63-linked polyubiquitin chains by NZF domains of TAB2 and TAB3. *EMBO J* 2009, *28*, 3903-3909.
[76] Sato, Y., Fujita, H., Yoshikawa, A., Yamashita, M.*, et al.*, Specific recognition of linear ubiquitin chains by the Npl4 zinc finger (NZF) domain of the HOIL-1L subunit of the linear ubiquitin chain assembly complex. *Proc Natl Acad Sci U S A* 2011, *108*, 20520-20525.
[77] Trempe, J. F., Brown, N. R., Lowe, E. D., Gordon, C.*, et al.*, Mechanism of Lys48-linked polyubiquitin chain recognition by the Mud1 UBA domain. *EMBO J* 2005, *24*, 3178-3189.
[78] Bosanac, I., Wertz, I. E., Pan, B., Yu, C.*, et al.*, Ubiquitin binding to A20 ZnF4 is required for modulation of NF-kappaB signaling. *Mol Cell* 2010, *40*, 548-557.
[79] Rahighi, S., Ikeda, F., Kawasaki, M., Akutsu, M.*, et al.*, Specific recognition of linear ubiquitin chains by NEMO is important for NF-kappaB activation. *Cell* 2009, *136*, 1098-1109.
[80] Coleman, K. A., Greenberg, R. A., The BRCA1-RAP80 complex regulates DNA repair mechanism utilization by restricting end resection. *J Biol Chem* 2011, *286*, 13669-13680.
[81] Sims, J. J., Cohen, R. E., Linkage-specific avidity defines the lysine 63-linked polyubiquitin-binding preference of rap80. *Mol Cell* 2009, *33*, 775-783.
[82] Zhang, N., Wang, Q., Ehlinger, A., Randles, L.*, et al.*, Structure of the s5a:k48-linked diubiquitin complex and its interactions with rpn13. *Mol Cell* 2009, *35*, 280-290.
[83] Rahighi, S., Braunstein, I., Ternette, N., Kessler, B.*, et al.*, Selective Binding of AIRAPL Tandem UIMs to Lys48-Linked Tri-Ubiquitin Chains. *Structure* 2016.
[84] Reyes-Turcu, F. E., Horton, J. R., Mullally, J. E., Heroux, A.*, et al.*, The ubiquitin binding domain ZnF UBP recognizes the C-terminal diglycine motif of unanchored ubiquitin. *Cell* 2006, *124*, 1197-1208.
[85] Kravtsova-Ivantsiv, Y., Ciechanover, A., Non-canonical ubiquitin-based signals for proteasomal degradation. *J Cell Sci* 2012, *125*, 539-548.

[86] Komander, D., Clague, M. J., Urbe, S., Breaking the chains: structure and function of the deubiquitinases. *Nat Rev Mol Cell Biol* 2009, *10*, 550-563.
[87] Reyes-Turcu, F. E., Shanks, J. R., Komander, D., Wilkinson, K. D., Recognition of polyubiquitin isoforms by the multiple ubiquitin binding modules of isopeptidase T. *J Biol Chem* 2008, *283*, 19581-19592.
[88] Volk, S., Wang, M., Pickart, C. M., Chemical and genetic strategies for manipulating polyubiquitin chain structure. *Methods Enzymol* 2005, *399*, 3-20.
[89] Ekkebus, R., Flierman, D., Geurink, P. P., Ovaa, H., Catching a DUB in the act: novel ubiquitin-based active site directed probes. *Curr Opin Chem Biol* 2014, *23*, 63-70.
[90] McGouran, J. F., Gaertner, S. R., Altun, M., Kramer, H. B., Kessler, B. M., Deubiquitinating enzyme specificity for ubiquitin chain topology profiled by di-ubiquitin activity probes. *Chem Biol* 2013, *20*, 1447-1455.
[91] Altun, M., Kramer, H. B., Willems, L. I., McDermott, J. L.*, et al.*, Activity-based chemical proteomics accelerates inhibitor development for deubiquitylating enzymes. *Chem Biol* 2011, *18*, 1401-1412.
[92] Kumar, K. S., Spasser, L., Ohayon, S., Erlich, L. A., Brik, A., Expeditious chemical synthesis of ubiquitinated peptides employing orthogonal protection and native chemical ligation. *Bioconjug Chem* 2011, *22*, 137-143.
[93] Keren-Kaplan, T., Attali, I., Motamedchaboki, K., Davis, B. A.*, et al.*, Synthetic biology approach to reconstituting the ubiquitylation cascade in bacteria. *EMBO J* 2012, *31*, 378-390.
[94] Keren-Kaplan, T., Prag, G., Purification and crystallization of mono-ubiquitylated ubiquitin receptor Rpn10. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 2012, *68*, 1120-1123.
[95] Matsumoto, M. L., Dong, K. C., Yu, C., Phu, L.*, et al.*, Engineering and structural characterization of a linear polyubiquitin-specific antibody. *J Mol Biol* 2012, *418*, 134-144.
[96] Finley, D., Sadis, S., Monia, B. P., Boucher, P.*, et al.*, Inhibition of proteolysis and cell cycle progression in a multiubiquitination-deficient yeast mutant. *Mol Cell Biol* 1994, *14*, 5501-5509.
[97] Deng, L., Wang, C., Spencer, E., Yang, L.*, et al.*, Activation of the IkappaB kinase complex by TRAF6 requires a dimeric ubiquitin-conjugating enzyme complex and a unique polyubiquitin chain. *Cell* 2000, *103*, 351-361.
[98] Shang, F., Deng, G., Liu, Q., Guo, W.*, et al.*, Lys6-modified ubiquitin inhibits ubiquitin-dependent protein degradation. *J Biol Chem* 2005, *280*, 20365-20374.
[99] Boname, J. M., Thomas, M., Stagg, H. R., Xu, P.*, et al.*, Efficient internalization of MHC I requires lysine-11 and lysine-63 mixed linkage polyubiquitin chains. *Traffic* 2010, *11*, 210-220.
[100] Xu, M., Skaug, B., Zeng, W., Chen, Z. J., A ubiquitin replacement strategy in human cells reveals distinct mechanisms of IKK activation by TNFalpha and IL-1beta. *Mol Cell* 2009, *36*, 302-314.
[101] Franco, M., Seyfried, N. T., Brand, A. H., Peng, J., Mayor, U., A novel strategy to isolate ubiquitin conjugates reveals wide role for ubiquitination during neural development. *Mol Cell Proteomics* 2011, *10*, M110 002188.

[102] Layfield, R., Tooth, D., Landon, M., Dawson, S., *et al.*, Purification of poly-ubiquitinated proteins by S5a-affinity chromatography. *Proteomics* 2001, *1*, 773-777.
[103] Weekes, J., Morrison, K., Mullen, A., Wait, R., *et al.*, Hyperubiquitination of proteins in dilated cardiomyopathy. *Proteomics* 2003, *3*, 208-216.
[104] Chapelsky, S., Batty, S., Frost, M., Mogridge, J., Inhibition of anthrax lethal toxin-induced cytolysis of RAW264.7 cells by celastrol. *PLoS One* 2008, *3*, e1421.
[105] Ehlers, M. D., Activity level controls postsynaptic composition and signaling via the ubiquitin-proteasome system. *Nat Neurosci* 2003, *6*, 231-242.
[106] Mayor, T., Deshaies, R. J., Two-step affinity purification of multiubiquitylated proteins from Saccharomyces cerevisiae. *Methods Enzymol* 2005, *399*, 385-392.
[107] Strachan, J., Roach, L., Sokratous, K., Tooth, D., *et al.*, Insights into the molecular composition of endogenous unanchored polyubiquitin chains. *J Proteome Res* 2012, *11*, 1969-1980.
[108] Kaiser, S. E., Riley, B. E., Shaler, T. A., Trevino, R. S., *et al.*, Protein standard absolute quantification (PSAQ) method for the measurement of cellular ubiquitin pools. *Nat Methods* 2011, *8*, 691-696.
[109] Nakayasu, E. S., Ansong, C., Brown, J. N., Yang, F., *et al.*, Evaluation of selected binding domains for the analysis of ubiquitinated proteomes. *J Am Soc Mass Spectrom* 2013, *24*, 1214-1223.
[110] Hjerpe, R., Aillet, F., Lopitz-Otsoa, F., Lang, V., *et al.*, Efficient protection and isolation of ubiquitylated proteins using tandem ubiquitin-binding entities. *EMBO Rep* 2009, *10*, 1250-1258.
[111] Mark, K. G., Simonetta, M., Maiolica, A., Seller, C. A., Toczyski, D. P., Ubiquitin ligase trapping identifies an SCF(Saf1) pathway targeting unprocessed vacuolar/lysosomal proteins. *Mol Cell* 2014, *53*, 148-161.
[112] O'Connor, H. F., Lyon, N., Leung, J. W., Agarwal, P., *et al.*, Ubiquitin-Activated Interaction Traps (UBAITs) identify E3 ligase binding partners. *EMBO Rep* 2015, *16*, 1699-1712.
[113] Sims, J. J., Scavone, F., Cooper, E. M., Kane, L. A., *et al.*, Polyubiquitin-sensor proteins reveal localization and linkage-type dependence of cellular ubiquitin signaling. *Nat Methods* 2012, *9*, 303-309.
[114] van Wijk, S. J., Fiskin, E., Putyrski, M., Pampaloni, F., *et al.*, Fluorescence-based sensors to monitor localization and functions of linear and K63-linked ubiquitin chains in cells. *Mol Cell* 2012, *47*, 797-809.
[115] Chen, Z., Zhong, Y., Wang, Y., Xu, S., *et al.*, Ubiquitination-induced fluorescence complementation (UiFC) for detection of K48 ubiquitin chains in vitro and in live cells. *PLoS One* 2013, *8*, e73482.
[116] Mevissen, T. E., Hospenthal, M. K., Geurink, P. P., Elliott, P. R., *et al.*, OTU deubiquitinases reveal mechanisms of linkage specificity and enable ubiquitin chain restriction analysis. *Cell* 2013, *154*, 169-184.
[117] Thomson, J. J., Rays of positive electricity. *Philos. Mag.* 1910, *20*, 752-767.
[118] Aston, F. W., A positive ray spectrograph. *Philos. Mag.* 1919, *38*, 707–715.
[119] Dempster, A. J., A new method of positive ray analysis. *Phys. rev.* 1918, *11*, 316–324.

[120] Karas, M., Bahr, U., Ingendoh, A., Hillenkamp, F., *Angewandte Chemie International Edition in English* 1989, pp. 760-761.
[121] Tanaka, K., Waki, H., Ido, Y., Akita, S.*, et al.*, *Rapid Communications in Mass Spectrometry* 1988, pp. 151-153.
[122] Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., Whitehouse, C. M., Electrospray ionization for mass spectrometry of large biomolecules. *Science* 1989, *246*, 64-71.
[123] Loo, J. A., Studying noncovalent protein complexes by electrospray ionization mass spectrometry. *Mass Spectrom Rev* 1997, *16*, 1-23.
[124] Taylor, G., Disintegration of Water Drops in an Electric Field. *Proceedings of the Royal Society A* 1964, *280*, 383.
[125] Gomez, A. a. T., K. Q., Charge and fission of droplets in electrostatic sprays. *Physics of Fluids* 1994, *6*, 404-414.
[126] Kebarle, P., Verkerk, U. H., Electrospray: from ions in solution to ions in the gas phase, what we know now. *Mass Spectrom Rev* 2009, *28*, 898-917.
[127] Konermann, L., Ahadi, E., Rodriguez, A. D., Vahidi, S., Unraveling the mechanism of electrospray ionization. *Anal Chem* 2013, *85*, 2-9.
[128] Kebarle, P., A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry. *J Mass Spectrom* 2000, *35*, 804-817.
[129] de Hoffmann, E., Stroobant, V., *Mass Spectrometry: Principles and Applications*, Wiley 2007.
[130] Roepstorff, P., Fohlman, J., Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed Mass Spectrom* 1984, *11*, 601.
[131] Steen, H., Mann, M., The ABC's (and XYZ's) of peptide sequencing. *Nat Rev Mol Cell Biol* 2004, *5*, 699-711.
[132] Motoyama, A., Yates, J. R., 3rd, Multidimensional LC separations in shotgun proteomics. *Anal Chem* 2008, *80*, 7187-7193.
[133] Peng, J., Gygi, S. P., Proteomics: the move to mixtures. *J Mass Spectrom* 2001, *36*, 1083-1091.
[134] Larance, M., Lamond, A. I., Multidimensional proteomics for cell biology. *Nat Rev Mol Cell Biol* 2015, *16*, 269-280.
[135] Nilsson, T., Mann, M., Aebersold, R., Yates, J. R., 3rd*, et al.*, Mass spectrometry in high-throughput proteomics: ready for the big time. *Nat Methods* 2010, *7*, 681-685.
[136] Heck, A. J., Native mass spectrometry: a bridge between interactomics and structural biology. *Nat Methods* 2008, *5*, 927-933.
[137] Konermann, L., Douglas, D. J., Acid-induced unfolding of cytochrome c at different methanol concentrations: electrospray ionization mass spectrometry specifically monitors changes in the tertiary structure. *Biochemistry* 1997, *36*, 12296-12302.
[138] Breuker, K., McLafferty, F. W., Stepwise evolution of protein native structure with electrospray into the gas phase, 10(-12) to 10(2) s. *Proc Natl Acad Sci U S A* 2008, *105*, 18145-18152.
[139] Hogan, C. J., Jr., Ruotolo, B. T., Robinson, C. V., Fernandez de la Mora, J., Tandem differential mobility analysis-mass spectrometry reveals partial gas-phase collapse of the GroEL complex. *J Phys Chem B* 2011, *115*, 3614-3621.
[140] Afonso, J. P., Chintakayala, K., Suwannachart, C., Sedelnikova, S.*, et al.*, Insights into the structure and assembly of the Bacillus subtilis clamp-

loader complex and its interaction with the replicative helicase. *Nucleic Acids Res* 2013, *41*, 5115-5126.
[141] Jurneczko, E., Cruickshank, F., Porrini, M., Clarke, D. J., *et al.*, Probing the conformational diversity of cancer-associated mutations in p53 with ion-mobility mass spectrometry. *Angew Chem Int Ed Engl* 2013, *52*, 4370-4374.
[142] Knapman, T. W., Valette, N. M., Warriner, S. L., Ashcroft, A. E., Ion Mobility Spectrometry-Mass Spectrometry of Intrinsically Unfolded Proteins: Trying to Put Order into Disorder. *Curr Anal Chem* 2013, *9*, 181-191.
[143] Hochberg, G. K., Ecroyd, H., Liu, C., Cox, D., *et al.*, The structured core domain of alphaB-crystallin can prevent amyloid fibrillation and associated toxicity. *Proc Natl Acad Sci U S A* 2014, *111*, E1562-1570.
[144] Uetrecht, C., Versluis, C., Watts, N. R., Wingfield, P. T., *et al.*, Stability and shape of hepatitis B virus capsids in vacuo. *Angew Chem Int Ed Engl* 2008, *47*, 6247-6251.
[145] Uetrecht, C., Rose, R. J., van Duijn, E., Lorenzen, K., Heck, A. J., Ion mobility mass spectrometry of proteins and protein assemblies. *Chem Soc Rev* 2010, *39*, 1633-1655.
[146] Loo, J. A., Berhane, B., Kaddis, C. S., Wooding, K. M., *et al.*, Electrospray ionization mass spectrometry and ion mobility analysis of the 20S proteasome complex. *J Am Soc Mass Spectrom* 2005, *16*, 998-1008.
[147] Wyttenbach, T., Grabenauer, M., Thalassinos, K., Scrivens, J. H., Bowers, M. T., The effect of calcium ions and peptide ligands on the relative stabilities of the calmodulin dumbbell and compact structures. *J Phys Chem B* 2010, *114*, 437-447.
[148] Sharon, M., Robinson, C. V., The role of mass spectrometry in structure elucidation of dynamic protein complexes. *Annu Rev Biochem* 2007, *76*, 167-193.
[149] van Duijn, E., Simmons, D. A., van den Heuvel, R. H., Bakkes, P. J., *et al.*, Tandem mass spectrometry of intact GroEL-substrate complexes reveals substrate-specific conformational changes in the trans ring. *J Am Chem Soc* 2006, *128*, 4694-4702.
[150] van Duijn, E., Bakkes, P. J., Heeren, R. M., van den Heuvel, R. H., *et al.*, Monitoring macromolecular complexes involved in the chaperonin-assisted protein folding cycle by mass spectrometry. *Nat Methods* 2005, *2*, 371-376.
[151] Vaughan, C. K., Gohlke, U., Sobott, F., Good, V. M., *et al.*, Structure of an Hsp90-Cdc37-Cdk4 complex. *Mol Cell* 2006, *23*, 697-707.
[152] Fuerstenau, S. D., Benner, W. H., Thomas, J. J., Brugidou, C., *et al.*, Mass Spectrometry of an Intact Virus *Angew Chem Int Ed Engl* 2001, *40*, 9822.
[153] Uetrecht, C., Versluis, C., Watts, N. R., Roos, W. H., *et al.*, High-resolution mass spectrometry of viral assemblies: molecular composition and stability of dimorphic hepatitis B virus capsids. *Proc Natl Acad Sci U S A* 2008, *105*, 9216-9220.
[154] Baldwin, A. J., Lioe, H., Robinson, C. V., Kay, L. E., Benesch, J. L., alphaB-crystallin polydispersity is a consequence of unbiased quaternary dynamics. *J Mol Biol* 2011, *413*, 297-309.
[155] Sharon, M., Taverner, T., Ambroggio, X. I., Deshaies, R. J., Robinson, C. V., Structural organization of the 19S proteasome lid: insights from MS of intact complexes. *PLoS Biol* 2006, *4*, e267.

[156] Sharon, M., Witt, S., Glasmacher, E., Baumeister, W., Robinson, C. V., Mass spectrometry reveals the missing links in the assembly pathway of the bacterial 20 S proteasome. *J Biol Chem* 2007, *282*, 18448-18457.
[157] Sharon, M., Witt, S., Felderer, K., Rockel, B.*, et al.*, 20S proteasomes have the potential to keep substrates in store for continual degradation. *J Biol Chem* 2006, *281*, 9569-9575.
[158] Taverner, T., Hernandez, H., Sharon, M., Ruotolo, B. T.*, et al.*, Subunit architecture of intact protein complexes from mass spectrometry and homology modeling. *Acc Chem Res* 2008, *41*, 617-627.
[159] Hernandez, H., Dziembowski, A., Taverner, T., Seraphin, B., Robinson, C. V., Subunit architecture of multimeric complexes isolated directly from cells. *EMBO Rep* 2006, *7*, 605-610.
[160] Barrera, N. P., Di Bartolo, N., Booth, P. J., Robinson, C. V., Micelles protect membrane complexes from solution to vacuum. *Science* 2008, *321*, 243-246.
[161] Liu, J., Konermann, L., Protein-protein binding affinities in solution determined by electrospray mass spectrometry. *J Am Soc Mass Spectrom* 2011, *22*, 408-417.
[162] Lanucara, F., Holman, S. W., Gray, C. J., Eyers, C. E., The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics. *Nat Chem* 2014, *6*, 281-294.
[163] Giles, K., Pringle, S. D., Worthington, K. R., Little, D.*, et al.*, Applications of a travelling wave-based radio-frequency-only stacked ring ion guide. *Rapid Commun Mass Spectrom* 2004, *18*, 2401-2414.
[164] Kanu, A. B., Dwivedi, P., Tam, M., Matz, L., Hill, H. H., Jr., Ion mobility-mass spectrometry. *J Mass Spectrom* 2008, *43*, 1-22.
[165] Pringle, S. D., Giles, K., Wildgoose, J.L., Williams, J.P., Slade, S.E., Thalassinos, K., Bateman, R.H., Bowers, M.T. and Scrivens, J.H. , An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oa-ToF instrument. *Int. J. Mass Spectrom.* 2007, *261*, 1-12.
[166] Ruotolo, B. T., Hyung, S. J., Robinson, P. M., Giles, K.*, et al.*, Ion mobility-mass spectrometry reveals long-lived, unfolded intermediates in the dissociation of protein complexes. *Angew Chem Int Ed Engl* 2007, *46*, 8001-8004.
[167] Ruotolo, B. T., Benesch, J. L., Sandercock, A. M., Hyung, S. J., Robinson, C. V., Ion mobility-mass spectrometry analysis of large protein complexes. *Nat Protoc* 2008, *3*, 1139-1152.
[168] Benesch, J. L., Aquilina, J. A., Ruotolo, B. T., Sobott, F., Robinson, C. V., Tandem mass spectrometry reveals the quaternary organization of macromolecular assemblies. *Chem Biol* 2006, *13*, 597-605.
[169] Benesch, J. L., Ruotolo, B. T., Mass spectrometry: come of age for structural and dynamical biology. *Curr Opin Struct Biol* 2011, *21*, 641-649.
[170] Konermann, L., Vahidi, S., Sowole, M. A., Mass spectrometry methods for studying structure and dynamics of biological macromolecules. *Anal Chem* 2014, *86*, 213-232.
[171] Rajabi, K., Ashcroft, A. E., Radford, S. E., Mass spectrometric methods to analyze the structural organization of macromolecular complexes. *Methods* 2015, *89*, 13-21.
[172] Sprangers, R., Velyvis, A., Kay, L. E., Solution NMR of supramolecular complexes: providing new insights into function. *Nat Methods* 2007, *4*, 697-703.

[173] Jeon, H. B., Choi, E. S., Yoon, J. H., Hwang, J. H., *et al.*, A proteomics approach to identify the ubiquitinated proteins in mouse heart. *Biochem Biophys Res Commun* 2007, *357*, 731-736.
[174] Meierhofer, D., Wang, X., Huang, L., Kaiser, P., Quantitative analysis of global ubiquitination in HeLa cells by mass spectrometry. *J Proteome Res* 2008, *7*, 4566-4576.
[175] Danielsen, J. M., Sylvestersen, K. B., Bekker-Jensen, S., Szklarczyk, D., *et al.*, Mass spectrometric analysis of lysine ubiquitylation reveals promiscuity at site level. *Mol Cell Proteomics* 2011, *10*, M110 003590.
[176] Xu, G., Paige, J. S., Jaffrey, S. R., Global analysis of lysine ubiquitination by ubiquitin remnant immunoaffinity profiling. *Nat Biotechnol* 2010, *28*, 868-873.
[177] Burande, C. F., Heuze, M. L., Lamsoul, I., Monsarrat, B., *et al.*, A label-free quantitative proteomics strategy to identify E3 ubiquitin ligase substrates targeted to proteasome degradation. *Mol Cell Proteomics* 2009, *8*, 1719-1727.
[178] Hor, S., Ziv, T., Admon, A., Lehner, P. J., Stable isotope labeling by amino acids in cell culture and differential plasma membrane proteome quantitation identify new substrates for the MARCH9 transmembrane E3 ligase. *Mol Cell Proteomics* 2009, *8*, 1959-1971.
[179] Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., *et al.*, Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* 2002, *1*, 376-386.
[180] Mayor, T., Graumann, J., Bryan, J., MacCoss, M. J., Deshaies, R. J., Quantitative profiling of ubiquitylated proteins reveals proteasome substrates and the substrate repertoire influenced by the Rpn10 receptor pathway. *Mol Cell Proteomics* 2007, *6*, 1885-1895.
[181] Wagner, S. A., Beli, P., Weinert, B. T., Nielsen, M. L., *et al.*, A proteome-wide, quantitative survey of in vivo ubiquitylation sites reveals widespread regulatory roles. *Mol Cell Proteomics* 2011, *10*, M111 013284.
[182] Udeshi, N. D., Mani, D. R., Eisenhaure, T., Mertins, P., *et al.*, Methods for quantification of in vivo changes in protein ubiquitination following proteasome and deubiquitinase inhibition. *Mol Cell Proteomics* 2012, *11*, 148-159.
[183] Udeshi, N. D., Svinkina, T., Mertins, P., Kuhn, E., *et al.*, Refined preparation and use of anti-diglycine remnant (K-epsilon-GG) antibody enables routine quantification of 10,000s of ubiquitination sites in single proteomics experiments. *Mol Cell Proteomics* 2013, *12*, 825-831.
[184] Rose, C. M., Isasa, M., Ordureau, A., Prado, M. A., *et al.*, Highly Multiplexed Quantitative Mass Spectrometry Analysis of Ubiquitylomes. *Cell Syst* 2016.
[185] Kirkpatrick, D. S., Hathaway, N. A., Hanna, J., Elsasser, S., *et al.*, Quantitative analysis of in vitro ubiquitinated cyclin B1 reveals complex chain topology. *Nat Cell Biol* 2006, *8*, 700-710.
[186] Dammer, E. B., Na, C. H., Xu, P., Seyfried, N. T., *et al.*, Polyubiquitin linkage profiles in three models of proteolytic stress suggest the etiology of Alzheimer disease. *J Biol Chem* 2011, *286*, 10457-10465.
[187] Blankenship, J. W., Varfolomeev, E., Goncharov, T., Fedorova, A. V., *et al.*, Ubiquitin binding modulates IAP antagonist-stimulated proteasomal degradation of c-IAP1 and c-IAP2(1). *Biochem J* 2009, *417*, 149-160.

[188] Newton, K., Matsumoto, M. L., Wertz, I. E., Kirkpatrick, D. S., *et al.*, Ubiquitin chain editing revealed by polyubiquitin linkage-specific antibodies. *Cell* 2008, *134*, 668-678.
[189] Huang, F., Kirkpatrick, D., Jiang, X., Gygi, S., Sorkin, A., Differential regulation of EGF receptor internalization and degradation by multiubiquitination within the kinase domain. *Mol Cell* 2006, *21*, 737-748.
[190] Sliter, D. A., Kubota, K., Kirkpatrick, D. S., Alzayady, K. J., *et al.*, Mass spectrometric analysis of type 1 inositol 1,4,5-trisphosphate receptor ubiquitination. *J Biol Chem* 2008, *283*, 35319-35328.
[191] Matiuhin, Y., Kirkpatrick, D. S., Ziv, I., Kim, W., *et al.*, Extraproteasomal Rpn10 restricts access of the polyubiquitin-binding protein Dsk2 to proteasome. *Mol Cell* 2008, *32*, 415-425.
[192] Wu, S. L., Kim, J., Bandle, R. W., Liotta, L., *et al.*, Dynamic profiling of the post-translational modifications and interaction partners of epidermal growth factor receptor signaling after stimulation by epidermal growth factor using Extended Range Proteomic Analysis (ERPA). *Mol Cell Proteomics* 2006, *5*, 1610-1627.
[193] Xu, P., Peng, J., Characterization of polyubiquitin chain structure by middle-down mass spectrometry. *Anal Chem* 2008, *80*, 3438-3444.
[194] Lee, A. E., Castaneda, C. A., Wang, Y., Fushman, D., Fenselau, C., Preparing to read the ubiquitin code: a middle-out strategy for characterization of all lysine-linked diubiquitins. *J Mass Spectrom* 2014, *49*, 1272-1278.
[195] Cannon, J. R., Martinez-Fonts, K., Robotham, S. A., Matouschek, A., Brodbelt, J. S., Top-down 193-nm ultraviolet photodissociation mass spectrometry for simultaneous determination of polyubiquitin chain length and topology. *Anal Chem* 2015, *87*, 1812-1820.
[196] Valkevich, E. M., Sanchez, N. A., Ge, Y., Strieter, E. R., Middle-down mass spectrometry enables characterization of branched ubiquitin chains. *Biochemistry* 2014, *53*, 4979-4989.
[197] Sokratous, K., Strachan, J., Roach, L. V., Layfield, R., Oldham, N. J., Cyclisation of Lys48-linked diubiquitin in vitro and in vivo. *FEBS Lett* 2012, *586*, 4144-4147.
[198] Sokratous, K., Roach, L. V., Channing, D., Strachan, J., *et al.*, Probing affinity and ubiquitin linkage selectivity of ubiquitin-binding domains using mass spectrometry. *J Am Chem Soc* 2012, *134*, 6416-6424.
[199] Kimura, Y., Yashiroda, H., Kudo, T., Koitabashi, S., *et al.*, An inhibitor of a deubiquitinating enzyme regulates ubiquitin homeostasis. *Cell* 2009, *137*, 549-559.
[200] Braten, O., Shabek, N., Kravtsova-Ivantsiv, Y., Ciechanover, A., Generation of free ubiquitin chains is up-regulated in stress and facilitated by the HECT domain ubiquitin ligases UFD4 and HUL5. *Biochem J* 2012, *444*, 611-617.
[201] Li, W., Tu, D., Brunger, A. T., Ye, Y., A ubiquitin ligase transfers preformed polyubiquitin chains from a conjugating enzyme to a substrate. *Nature* 2007, *446*, 333-337.
[202] Piotrowski, J., Beal, R., Hoffman, L., Wilkinson, K. D., *et al.*, Inhibition of the 26 S proteasome by polyubiquitin chains synthesized to have defined lengths. *Journal of Biological Chemistry* 1997, *272*, 23712-23721.
[203] Amerik, A. Y., Swaminathan, S., Krantz, B. A., Wilkinson, K. D., Hochstrasser, M., In vivo disassembly of free polyubiquitin chains by yeast

Ubp14 modulates rates of protein degradation by the proteasome. *Embo Journal* 1997, *16*, 4826-4838.
[204] Dayal, S., Sparks, A., Jacob, J., Allende-Vega, N.*, et al.*, Suppression of the deubiquitinating enzyme USP5 causes the accumulation of unanchored polyubiquitin and the activation of p53. *J Biol Chem* 2009, *284*, 5030-5041.
[205] Hanna, J., Hathaway, N. A., Tone, Y., Crosas, B.*, et al.*, Deubiquitinating enzyme Ubp6 functions noncatalytically to delay proteasomal degradation. *Cell* 2006, *127*, 99-111.
[206] Besche, H. C., Haas, W., Gygi, S. P., Goldberg, A. L., Isolation of mammalian 26S proteasomes and p97/VCP complexes using the ubiquitin-like domain from HHR23B reveals novel proteasome-associated proteins. *Biochemistry* 2009, *48*, 2538-2549.
[207] Zhang, N. Y., Jacobson, A. D., Macfadden, A., Liu, C. W., Ubiquitin chain trimming recycles the substrate binding sites of the 26 S proteasome and promotes degradation of lysine 48-linked polyubiquitin conjugates. *J Biol Chem* 2011, *286*, 25540-25546.
[208] Lee, B. H., Lu, Y., Prado, M. A., Shi, Y.*, et al.*, USP14 deubiquitinates proteasome-bound substrates that are ubiquitinated at multiple sites. *Nature* 2016, *532*, 398-401.
[209] Lass, A., Cocklin, R., Scaglione, K. M., Skowyra, M.*, et al.*, The loop-less tmCdc34 E2 mutant defective in polyubiquitination in vitro and in vivo supports yeast growth in a manner dependent on Ubp14 and Cka2. *Cell Div* 2011, *6*, 7.
[210] Ouyang, H., Ali, Y. O., Ravichandran, M., Dong, A.*, et al.*, Protein aggregates are recruited to aggresome by histone deacetylase 6 via unanchored ubiquitin C termini. *J Biol Chem* 2012, *287*, 2317-2327.
[211] Hao, R., Nanduri, P., Rao, Y., Panichelli, R. S.*, et al.*, Proteasomes activate aggresome disassembly and clearance by producing unanchored ubiquitin chains. *Mol Cell* 2013, *51*, 819-828.
[212] Nanduri, P., Hao, R., Fitzpatrick, T., Yao, T. P., Chaperone-mediated 26S Proteasome Remodeling Facilitates Free K63 Ubiquitin Chain Production and Aggresome Clearance. *J Biol Chem* 2015, *290*, 9455-9464.
[213] Banerjee, I., Miyake, Y., Nobs, S. P., Schneider, C.*, et al.*, Influenza A virus uses the aggresome processing machinery for host cell entry. *Science* 2014, *346*, 473-477.
[214] Chen, J., Chen, Z. J., Regulation of NF-kappaB by ubiquitination. *Curr Opin Immunol* 2013, *25*, 4-12.
[215] Rajsbaum, R., Garcia-Sastre, A., Versteeg, G. A., TRIMmunity: the roles of the TRIM E3-ubiquitin ligase family in innate antiviral immunity. *J Mol Biol* 2014, *426*, 1265-1284.
[216] Pertel, T., Hausmann, S., Morger, D., Zuger, S.*, et al.*, TRIM5 is an innate immune sensor for the retrovirus capsid lattice. *Nature* 2011, *472*, 361-365.
[217] Fletcher, A. J., Christensen, D. E., Nelson, C., Tan, C. P.*, et al.*, TRIM5alpha requires Ube2W to anchor Lys63-linked ubiquitin chains and restrict reverse transcription. *EMBO J* 2015.
[218] Rajsbaum, R., Versteeg, G. A., Schmid, S., Maestre, A. M.*, et al.*, Unanchored K48-linked polyubiquitin synthesized by the E3-ubiquitin ligase TRIM6 stimulates the interferon-IKKepsilon kinase-mediated antiviral response. *Immunity* 2014, *40*, 880-895.

[219] Zeng, W., Sun, L., Jiang, X., Chen, X., *et al.*, Reconstitution of the RIG-I pathway reveals a signaling role of unanchored polyubiquitin chains in innate immunity. *Cell* 2010, *141*, 315-330.
[220] Miranzo-Navarro, D., Magor, K. E., Activation of duck RIG-I by TRIM25 is independent of anchored ubiquitin. *PLoS One* 2014, *9*, e86968.
[221] Jiang, X., Kinch, L. N., Brautigam, C. A., Chen, X., *et al.*, Ubiquitin-induced oligomerization of the RNA sensors RIG-I and MDA5 activates antiviral innate immune response. *Immunity* 2012, *36*, 959-973.
[222] Dynek, J. N., Goncharov, T., Dueber, E. C., Fedorova, A. V., *et al.*, c-IAP1 and UbcH5 promote K11-linked polyubiquitination of RIP1 in TNF signalling. *EMBO J* 2010, *29*, 4198-4209.
[223] Tokunaga, F., Nishimasu, H., Ishitani, R., Goto, E., *et al.*, Specific recognition of linear polyubiquitin by A20 zinc finger 7 is involved in NF-kappaB regulation. *EMBO J* 2012, *31*, 3856-3870.
[224] Wang, C., Long, W., Peng, C., Hu, L., *et al.*, HTLV-1 Tax Functions as a Ubiquitin E3 Ligase for Direct IKK Activation via Synthesis of Mixed-Linkage Polyubiquitin Chains. *PLoS Pathog* 2016, *12*, e1005584.
[225] Sambrook, J., Russell, D. W., The inoue method for preparation and transformation of competent e. Coli: "ultra-competent" cells. *CSH Protoc* 2006, *2006*.
[226] Liu, H., Naismith, J. H., An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol* 2008, *8*, 91.
[227] Wisniewski, J. R., Zougman, A., Nagaraj, N., Mann, M., Universal sample preparation method for proteome analysis. *Nat Methods* 2009, *6*, 359-362.
[228] Perkins, D. N., Pappin, D. J., Creasy, D. M., Cottrell, J. S., Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 1999, *20*, 3551-3567.
[229] The Universal Protein Resource (UniProt) in 2010. *Nucleic Acids Res* 2010, *38*, D142-148.
[230] Searle, B. C., Scaffold: a bioinformatic tool for validating MS/MS-based proteomic studies. *Proteomics* 2010, *10*, 1265-1269.
[231] Jumper, C. C., Schriemer, D. C., Mass spectrometry of laser-initiated carbene reactions for protein topographic analysis. *Anal Chem* 2011, *83*, 2913-2920.
[232] Bush, M. F., Hall, Z., Giles, K., Hoyes, J., *et al.*, Collision cross sections of proteins and their complexes: a calibration framework and database for gas-phase structural biology. *Anal Chem* 2010, *82*, 9557-9565.
[233] Roy, A., Kucukural, A., Zhang, Y., I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* 2010, *5*, 725-738.
[234] Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., *et al.*, MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 2010, *66*, 12-21.
[235] Case, D. A., Cheatham, T. E., 3rd, Darden, T., Gohlke, H., *et al.*, The Amber biomolecular simulation programs. *J Comput Chem* 2005, *26*, 1668-1688.
[236] Bertini, I., Case, D. A., Ferella, L., Giachetti, A., Rosato, A., A Grid-enabled web portal for NMR structure refinement with AMBER. *Bioinformatics* 2011, *27*, 2384-2390.

[237] Avvakumov, G. V., Walker, J. R., Xue, S., Allali-Hassani, A., *et al.*, Two ZnF-UBP domains in isopeptidase T (USP5). *Biochemistry* 2012, *51*, 1188-1198.
[238] Koradi, R., Billeter, M., Wuthrich, K., MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graph* 1996, *14*, 51-55, 29-32.
[239] Pai, M. T., Tzeng, S. R., Kovacs, J. J., Keaton, M. A., *et al.*, Solution structure of the Ubp-M BUZ domain, a highly specific protein module that recognizes the C-terminal tail of free ubiquitin. *J Mol Biol* 2007, *370*, 290-302.
[240] Labute, P., Protonate3D: assignment of ionization states and hydrogen coordinates to macromolecular structures. *Proteins* 2009, *75*, 187-205.
[241] **Amerik, A.**, Swaminathan, S., Krantz, B. A., Wilkinson, K. D., Hochstrasser, M., In vivo disassembly of free polyubiquitin chains by yeast Ubp14 modulates rates of protein degradation by the proteasome. *EMBO J* 1997, *16*, 4826-4838.
[242] Wang, C. H., Chen, G. C., Chien, C. T., The deubiquitinase Leon/USP5 regulates ubiquitin homeostasis during Drosophila development. *Biochem Biophys Res Commun* 2014, *452*, 369-375.
[243] Lindsey, D. F., Amerik, A., Deery, W. J., Bishop, J. D., *et al.*, A deubiquitinating enzyme that disassembles free polyubiquitin chains is required for development but not growth in Dictyostelium. *J Biol Chem* 1998, *273*, 29178-29187.
[244] Doelling, J. H., Yan, N., Kurepa, J., Walker, J., Vierstra, R. D., The ubiquitin-specific protease UBP14 is essential for early embryo development in Arabidopsis thaliana. *Plant J* 2001, *27*, 393-405.
[245] Komander, D., Lord, C. J., Scheel, H., Swift, S., *et al.*, The structure of the CYLD USP domain explains its specificity for Lys63-linked polyubiquitin and reveals a B box module. *Mol Cell* 2008, *29*, 451-464.
[246] Strachan, J., et al, Broad utility of an affinity-enrichment strategy for unanchored polyubiquitin. *Journal of Proteomics and Bioinformatics* 2013, *S7*.
[247] Ritorto, M. S., Ewan, R., Perez-Oliva, A. B., Knebel, A., *et al.*, Screening of DUB activity and specificity by MALDI-TOF mass spectrometry. *Nat Commun* 2014, *5*, 4763.
[248] Nielsen, M. L., Vermeulen, M., Bonaldi, T., Cox, J., *et al.*, Iodoacetamide-induced artifact mimics ubiquitination in mass spectrometry. *Nat Methods* 2008, *5*, 459-460.
[249] Xu, G., Jaffrey, S. R., Proteomic identification of protein ubiquitination events. *Biotechnol Genet Eng Rev* 2013, *29*, 73-109.
[250] Piotrowski, J., Beal, R., Hoffman, L., Wilkinson, K. D., *et al.*, Inhibition of the 26 S proteasome by polyubiquitin chains synthesized to have defined lengths. *J Biol Chem* 1997, *272*, 23712-23721.
[251] Setsuie, R., Sakurai, M., Sakaguchi, Y., Wada, K., Ubiquitin dimers control the hydrolase activity of UCH-L3. *Neurochem Int* 2009, *54*, 314-321.
[252] Lu, Y., Lee, B. H., King, R. W., Finley, D., Kirschner, M. W., Substrate degradation by the proteasome: a single-molecule kinetic analysis. *Science* 2015, *348*, 1250834.
[253] Bjorkhaug, L., Molnes, J., Sovik, O., Njolstad, P. R., Flatmark, T., Allosteric activation of human glucokinase by free polyubiquitin chains and

its ubiquitin-dependent cotranslational proteasomal degradation. *J Biol Chem* 2007, *282*, 22757-22764.
[254] Pertel, T., Hausmann, S., Morger, D., Zuger, S.*, et al.*, TRIM5 is an innate immune sensor for the retrovirus capsid lattice. *Nature* 2011, *472*, 361-365.
[255] Nakajima, S., Lan, L., Wei, L., Hsieh, C. L.*, et al.*, Ubiquitin-specific protease 5 is required for the efficient repair of DNA double-strand breaks. *PLoS One* 2014, *9*, e84899.
[256] Izaguirre, D. I., Zhu, W., Hai, T., Cheung, H. C.*, et al.*, PTBP1-dependent regulation of USP5 alternative RNA splicing plays a role in glioblastoma tumorigenesis. *Mol Carcinog* 2012, *51*, 895-906.
[257] DeForte, S., Uversky, V. N., Resolving the ambiguity: Making sense of intrinsic disorder when PDB structures disagree. *Protein Sci* 2016, *25*, 676-688.
[258] Hu, M., Li, P., Li, M., Li, W.*, et al.*, Crystal structure of a UBP-family deubiquitinating enzyme in isolation and in complex with ubiquitin aldehyde. *Cell* 2002, *111*, 1041-1054.
[259] Hu, M., Li, P., Song, L., Jeffrey, P. D.*, et al.*, Structure and mechanisms of the proteasome-associated deubiquitinating enzyme USP14. *EMBO J* 2005, *24*, 3747-3756.
[260] Dwivedi, P., Wu, C., Matz, L. M., Clowers, B. H.*, et al.*, Gas-phase chiral separations by ion mobility spectrometry. *Anal Chem* 2006, *78*, 8200-8206.
[261] Jurneczko, E., Barran, P. E., How useful is ion mobility mass spectrometry for structural biology? The relationship between protein crystal structures and their collision cross sections in the gas phase. *Analyst* 2011, *136*, 20-28.
[262] Jurneczko, E., Cruickshank, F., Porrini, M., Nikolova, P.*, et al.*, Intrinsic disorder in proteins: a challenge for (un)structural biology met by ion mobility-mass spectrometry. *Biochem Soc Trans* 2012, *40*, 1021-1026.
[263] Liu, L., Bagal, D., Kitova, E. N., Schnier, P. D., Klassen, J. S., Hydrophobic protein-ligand interactions preserved in the gas phase. *J Am Chem Soc* 2009, *131*, 15980-15981.
[264] Breuker, K., Bruschweiler, S., Tollinger, M., Electrostatic stabilization of a native protein structure in the gas phase. *Angew Chem Int Ed Engl* 2011, *50*, 873-877.
[265] Wyttenbach, T., Bowers, M. T., Structural stability from solution to the gas phase: native solution structure of ubiquitin survives analysis in a solvent-free ion mobility-mass spectrometry environment. *J Phys Chem B* 2011, *115*, 12266-12275.
[266] Chernushevich, I. V., Thomson, B. A., Collisional cooling of large ions in electrospray mass spectrometry. *Anal Chem* 2004, *76*, 1754-1760.
[267] Scott, D., Layfield, R., Oldham, N. J., Ion mobility-mass spectrometry reveals conformational flexibility in the deubiquitinating enzyme USP5. *Proteomics* 2015, *15*, 2835-2841.
[268] Kaltashov, I. A., Mohimen, A., Estimates of protein surface areas in solution by electrospray ionization mass spectrometry. *Anal Chem* 2005, *77*, 5370-5379.
[269] Kaltashov, I. A., Abzalimov, R. R., Do ionic charges in ESI MS provide useful information on macromolecular structure? *J Am Soc Mass Spectrom* 2008, *19*, 1239-1246.

[270] Scott, D., Layfield, R., Oldham, N. J., Structural insights into interactions between ubiquitin specific protease 5 and its polyubiquitin substrates by mass spectrometry and ion mobility spectrometry. *Protein Science* 2015, *24*, 1257-1263.
[271] Jumper, C. C., Bomgarden, R., Rogers, J., Etienne, C., Schriemer, D. C., High-resolution mapping of carbene-based protein footprints. *Anal Chem* 2012, *84*, 4411-4418.
[272] Sokratous, K., Roach, L. V., Channing, D., Strachan, J., *et al.*, Probing Affinity and Ubiquitin Linkage Selectivity of Ubiquitin-Binding Domains Using Mass Spectrometry. *Journal of the American Chemical Society* 2012, *134*, 6416-6424.
[273] Gao, Y., Li, Y., Zhang, C., Zhao, M., *et al.*, Enhanced Purification of Ubiquitinated Proteins by Engineered Tandem Hybrid Ubiquitin-binding Domains (ThUBDs). *Mol Cell Proteomics* 2016.
[274] Scott, D., Garner, T. P., Long, J., Strachan, J., *et al.*, Mass spectrometry insights into a tandem ubiquitin-binding domain hybrid engineered for the selective recognition of unanchored polyubiquitin. *Proteomics* 2016.
[275] Reyes-Turcu, F. E., Horton, J. R., Mullally, J. E., Heroux, A., *et al.*, The ubiquitin binding domain ZnFUBP recognizes the C-terminal diglycine motif of unanchored ubiquitin. *Cell* 2006, *124*, 1197-1208.
[276] Heideker, J., Wertz, I. E., DUBs, the regulation of cell identity and disease. *Biochem J* 2015, *467*, 191.

Appendix I

Buffer	Composition
5 X AE	250 mM Acetic acid, 10 mM EDTA (pH 4)
Acetate buffer	100mM NaOAc, 500mM NaCl (pH 4)
Ampicillin	100 mg/ml [286 mM] stock solution
Anionic elution buffer	5 mM potassium phosphate, 2 M NaCl (pH 7)
Anionic loading buffer	5 mM potassium phosphate (pH 7)
Blocking buffer	5% (w/v) Marvel, TBS-Tween (0.1%)
Buffer A	1.1M Tris, 0.1% (w/v) SDS, 30% (w/v) glycerol (pH 8.8)
Buffer B	1.1M Tris, 0.1% (w/v) SDS (pH8.8)
Cationic elution buffer	1 X AE buffer, 2 M NaCl
Cationic loading buffer	1 X AE buffer
Coomassie blue destain	10% (v/v) methanol, 10% (v/v) glacial acetic acid
Coomassie Blue stain	50% (v/v) methanol, 20% (v/v) glacial acetic acid, 0.12% (w/v) Coomassie Brilliant Blue R-250
Coupling buffer	100mM NaHCO3, 500mM NaCl (pH 8.3)
10 X DUB buffer	500 mM Tris, 1.5 M NaCl 100 mM DTT (pH 7.5)
DNAse I	10 mg/ml (50,000 U/ml) in 1 M $MgCl_2$, O.1 M MnCl
Electrode buffer	0.025M Tris, 0.19M glycine, 0.1% w/v SDS
Ethanolamine solution	1 M ethanolamine (pH 8.0)
SDS-PAGE gel application buffer (GAB)	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 (pH6.8)
Homogenising buffer	50 mM Tris, (pH 7.5), 150 mM NaCl, 0.5% (v/v) NP-40, N-ethylmaleimide (NEM; 5 mM), MG-132 (20 μM), mammalian protease inhibitor cocktail and mammalian phosphatase inhibitor cocktail , both 0.1% (v/v)
LB Agar	10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl, 15 g/L agar, 2.5 g/L select agar. *Autoclaved, then cooled to ~50 °C, at which point [286 μM] Ampicillin was added and the molten gel poured into petri dishes (~25 ml/dish)
LB broth	10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl
Ponceau S solution	0.1% (w/v) Ponceau S, 10% (v/v) glacial acetic acid
Protease inhibitor cocktail	AEBSF (2 mM), Aprotinin (0.3 μM), Bestatin (130 μM), EDTA (1 mM), E-64 (14 μM), Leupeptin (1 μM)
Sepharose wash buffer	50 mM Tris (pH 7.5)
Stacking buffer	0.14M Tris, 0.1% (w/v) SDS (pH6.8)
TAE buffer	40 mM Tris-acetate, 1 mM EDTA (pH 8)
TBS	20 mM Tris, 150 mM NaCl (pH 7.5)
TCB	20 mM Tris, 150 mM NaCl, 2.5 mM $CaCl_2$ (pH 8.4)
Tris buffer	100mM Tris, 500 mM NaCl (pH 8)
Ubiquitin GF buffer	5 mM Potassium phosphate (pH 7)
USP5 GF buffer	20 mM Tris, 0.5 M NaCl 2 mM DTT (pH 8)
Western transfer buffer	25 mM Tris, 192 mM glycine, 20% (v/v) methanol

Appendix II

Protein encoded	Primary sequence
USP5, Short Isoform	MAELSEEALLSVLPTIRVPKAGDRVHKDECAFSFDTPESEGGLYICMNTFLGFGKQYVERHFNKTGQRVY LHLRRTRRPKEEDPATGTGDPPRKKPTRLAIGVEGGFDLSEEKFELDEDVKIVILPDYLEIARDGLGGLPDI VRDRVTSAVEALLSADSASRKQEVQAWDGEVRQVSKHAFSLKQLDNPARIPPCGWKCSKCDMRENLW LNLTDGSILCGRRYFDGSGGNNHAVEHYRETGYPLAVKLGTITPDGADVYSYDEDDMVLDPSLAEHLSHF GIDMLKMQKTDKTMTELEIDMNQRIGEWELIQESGVPLKPLFGPGYTGIRNLGNSCYLNSVVQVLFSIPDF QRKYVDKLEKIFQNAPTDPTQDFSTQVAKLGHGLLSGEYSKPVPESGDGERVPEQKEVQDGIAPRMFKALIG KGHPEFSTNRQQDAQEFFLHLINMVERNCRSSENPNEVFRFLVEEKIKCLATEKVKYTQRVDYIMQLP VPMDAALNKEELLEYEEKKRQAEEEKMALPELVRAQVPFSSCLEAYGAPEQVDDFWSTALQAKSVAVKTT RFASFPDYLVIQIKKFTFGLDWVPKKLDVSIEMPEELDISQLRGTGLQPGEEELPDIAPPLVTPDEPKAP MLDESVIIQLVEMGFPMDACRKAVYYTGNSGAEAAMNWVMSHMDDPDFANPLILPGSSGPGSTSAAADPP PEDCVTTIVS MGFSRDQALKALRATNNSLE RAVDWIFSHI DDLDAEAAMD ISEGRSAADS ISESVPVGPK VRDGPGKYQLFAFISHMGTSTMCGHYVCHIKKEGRWVIYNDQKVCASEKPPKDLGYIYFYQRVAS	
ZnF-UBP (USP5)	KQEVQAWDGE VRQVSKHAFS LKQLDNPARI PPCGWKCSKC DMRENLWLNL TDGSILCGRRYFDGSGG NNHAVEHYRETGY PLAVKLGTIT PDGADVYSYD EDDMVLDPSL AEHLSHFGIDMLKMQKTDK	
UBA2 (hHR23a)	QVTPQEKEAI ERLKALGFPE SLVIQAYFAC EKNENLAANF LLSQNFDDE	
Monoubiquitin	MQIFVKTLTG KTITLEVEPS DTIENVKAKI QDKEGIPPDQ QRLIFAGKQL EDGRTLSDYN IQKESTLHLVL RLRGG	

*Residues mutated in the investigations presented in this thesis