# Structural and Biophysical Investigations into Ubiquitin Binding Proteins

Thomas Peter Garner, Msci. (Hons)

Thesis submitted to the University of Nottingham for the degree of
Doctor of Philosophy
July 2011

# Abstract

The complicated task of interpreting the many ubiquitin signals is mediated by specific ubiquitin binding domains. The investigations discussed in this thesis focus on two very different ubiquitin binding domains.

Chapters 3 to 5 detail the structural characterisation of the ubiquitin binding protein ZNF216. The structure of the ubiquitin binding Znf_A20 domain has been determined using multidimensional NMR techniques. A thermodynamic and structural characterisation of the interaction between the Znf_A20 and ubiquitin has been performed utilising chemical shift mapping, PRE based approaches, ESI-MS and ITC. The Znf_A20 domain forms a high affinity complex with Ub utilising a non-canonical binding site on ubiquitin centred at Asp58. The investigation was extended to the function of the Znf_A20 domain in the context of the full length protein. ZNF216, like many other ubiquitin receptors, has a 'hook and line' domain architecture with two independent domain separated by a long disordered linker.

Chapters 6, 7 and 8 focus on the UBA domain of p62. The p62-UBA domain has been identified as a 'hot spot' for mutations linked to Paget's disease of bone, a bone disorder which affects >3% of the over 55s. The dimerisation of the p62-UBA domain has been shown here to be a novel regulatory mechanism for the ubiquitin binding properties of p62. Modulation of both dimerisation and ubiquitin recognition are potential mechanisms by which mutations may disrupt p62 function and this prospect has been investigated here.

The final chapter of this thesis examines the possibility of cooperation between different ubiquitin binding domains by simultaneous interaction with Ub to form ternary complexes. Using the Znf_A20 and the p62-UBA as an example the formation of a ternary complex has been demonstrated. By examining the available ubiquitin complexes it has been suggested that the formation of Ub mediated ternary complexes is limited to only a few UBD pairings. The formation of Ub mediated ternary complexes may have interesting implications; for the formation of larger multi-protein complexes utilising ubiquitin as an interaction hub; on the recognition of poly-Ub with chain linkage specificity; and for Ub mediated signalling in general.

…I dedicate this thesis to my Grandfather Charles Richard Babington Elliot who was a great man…

# Acknowledgments

I wish to thank my supervisor Professor Mark Searle for providing me with the opportunity to carry out this work in his laboratory and for the constant help, support, and enthusiasm throughout my PhD. I would also like to thank Dr Robert Layfield for his countless pieces of advice, literary flare and biological tip offs.

I wish to give special thanks to Dr Jed Long who trained, suffered and fed me for 4 years, and with whom I have shared many ideas and enjoyed a fruitful collaboration.
I express my sincere gratitude to Dr Huw Williams for his friendship and assistance and for his technical wizardry which has got me out of many scrapes.

I must give a special mention to Drs Vicki Thurston and Anita Rea whose friendship made the lab feel like home and kept me going, and to Dr Graham Balkwill without whom I would never have started down this path.

I would like to take this opportunity to thank my undergraduate project student Liz Shedden for her hard work and dedication throughout.
I also would like to give a heart felt thank you to all members of the Searle group past and present; I would like to thank Drs Emma Simpson, Jill Watson, Tom Gallagher, Clare Evans and Ping Chen for their support in the early years, to my PhD Sister Laura Tye, to the project student and my good friend Anastasia Gadzhieva, to Kate Portman, Jenny Addlington, Liz Morris and John Edwards who have suffered me through my stressed and grumpy times with friendship patience and support, and finally to Alice Goode, Alex Cousins and John ‘the lion’ Phillips who have it all still to look forward to.
An extra thank you must go to all in the A20 Lab, especially to my good friends Shalu Sharma Kharkwal, Huey-Jen Fang, Jon Hopper, Chris Collier, Emily Golding and the ever present William Grainger.

I wish to thank all of my collaborators, Anthony Flint, Kamila Derecka, Malcolm Stevens, Dougal Ritson, Katarzyna Gluszyk, John Mosses, Jorg Linder, Christopher

Moody, Nico Dantuma, James Cavey, Dereen Najat, Neil Oldham, David Scott, Michael Williamson, Maya Pandya, Jeremy Craven and Joanna Strachan.
Finally I would like to thank the EPSRC and the University of Nottingham Chemistry department for funding.

## ..Acknowledgements 2...

I would like to thank my parents for letting me be me, for supporting me whatever choices I make and for always being proud of me whatever path I take.

To my Mother who has been through so much but always remains strong, I would not be the same person without you and I could not be more proud of you.

Thank you to my brother Nick for understanding and for keeping me grounded these thesis months and for teaching me to make things ‘fair’.

To my loving little sister Lucy who has always kept me going from a distance and kept me sane.

To Lucy R who helped me find me again

To the sacrifices I made unwillingly, you are my only regret…

# Table of Contents

# Abbreviations

| | |
|---|---|
| $^{15}$N{1H} NOE | $^{15}$N heteronuclear NOE |
| APS | Ammonium persulfate |
| AUC | Analytical Ultracentrifugation |
| BMRB | Biological magnetic resonance data bank |
| BSA | Bovine serum albumin |
| CD | Circular Dichroism |
| CID | Collision induced dissociation |
| CSD | Chemical shift deviation |
| CSI | Chemical shift index |
| DNA | Deoxyribonucleic acid |
| DSS | 4,4-dimethyl-4-silapentane-1-sulfonic acid |
| DTT | Dithiothreitol |
| DUB | Deubiquitinating |
| *E. Coli* | *Escherichia coli* |
| EDTA | Ethylene diamine tetraacetic acid |
| ESCRT | Endosomal sorting complex required for transport |
| ESI-MS | Electrospray Ionisation Mass Spectrometry |
| GST | Glutathione S-transferase |
| HDMS | High definition mass spectrometry |
| HSQC | Heteronuclear single quantum coherence |
| HyTEMPO | 4-hydroxy-2,2,6,6-tetramethylpiperidinyl-1-oxy |
| IκB | Inhibitor of κB |
| IPTG | Isopropyl-β-D-Thiogalactopyranoside |
| ISAC | Implicit saupe tensor alignment constraint |
| ITC | Isothermal titration calorimetry |
| MALDI | Matrix assisted laser desorption ionization |
| MTSL | S-(2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl) -methyl methanesulfonothioate |
| MWt | Molecular weight |
| NF- κB | Nuclear Factor κB |

| | |
|---|---|
| NMR | Nuclear Magnetic Resonance |
| NOE | Nuclear Overhauser effect |
| NOESY | NOE spectroscopy |
| OD | Optical density |
| PAG | Polyacrylamide |
| PB1 | Phox and Bem1p |
| PCR | Polymerase Chain Reaction |
| PDB | Paget's Disease of Bone |
| RPDB | RSC-Protein Data Bank |
| ppb | Parts per billion |
| ppm | Parts per million |
| PRE | Paramagnetic relaxation enhancement |
| PSVS | Protein structure validation software suite |
| PTM | Post-translational modification |
| qTOF | quadrupole-time of flight |
| RC | Random coil |
| RCI | Random coil index |
| RDC | Residual dipolar couplings |
| RSGI | RIKEN structural genomics initiative |
| S | Svedbergs |
| SAP | Stress associated protein |
| SDS-PAGE | Sodium dodecyl sulphate polyacrylamide gel electrophoresis |
| SEC | Size exclusion chromatography |
| Tc | Correlation time |
| TEMED | N,N,N,N-Tetramethylethylethylendiamine |
| TNFaIP3 | Tumor necrosis factor alpha(TNFα)-induced protein 3 |
| TOCSY | Total correlation spectra |
| Tris.HCl | tris(hydroxymethyl)aminomethane HCl |
| Ub | Ubiquitin |
| UBA | Ubiquitin associated domain |
| UBD | Ubiquitin binding domain |
| UPS | Ubiquitin Proteasome System |
| UV | Ultraviolet |
| wt | Wild type |

Amino Acid Codes

| Amino Acid | 3 letter Code | 1 letter Code | Amino Acid | 3 letter Code | 1 letter code |
|---|---|---|---|---|---|
| Alanine | Ala | A | Leucine | Leu | L |
| Arganine | Arg | R | Lysine | Lys | K |
| Asparagine | Asn | N | Methionine | Met | M |
| Asparatate | Asp | D | Phenylalanine | Phe | F |
| Cystein | Cys | C | Proline | Pro | P |
| Glutamaine | Glu | E | Serine | Ser | S |
| Glutamate | Gln | Q | Threonine | Thr | T |
| Glycine | Gly | G | Tryptophan | Trp | W |
| Histidine | His | H | Tyrosine | Tyr | Y |
| Isoleucine | Ile | I | Valine | Val | V |

# Chapter 1

## 1 Introduction

Proteins underpin the development, function and morphology of any given cell type. The many proteins which determine the behaviour of that cell are produced by selective gene expression and by alternative RNA splicing[1]. However a cell is not a static entity, it continuously responds to external and internal cues such as growth factors, DNA damage, quorum sensing molecules, cell-cell contact, mechanical loading and starvation. The adaption of a cell to these environmental changes may be mediated by a range of responses often resulting in changes in gene expression or splicing. However the cellular response to changing conditions is not limited to the level of protein production and is frequently mediated by the covalent modification of existing biomolecules[1,2]. In the case of proteins, almost every protein molecule will either be modified or be affected by a covalent post-translational modification (PTM) within its lifetime, the majority of which are reversible allowing for transient cellular responses and tight regulation[2]. The covalent modification of biomolecules in response to environmental changes is not limited to proteins and the reversible covalent modification of DNA, membrane lipids, and sugars are all common events covering every aspect of cellular biology [1,3].

Proteins are subject to a wide variety of PTMs including the isomerisation or hydroxylation of prolines, the methylation or acetylation of lysine side chains, phosphorylation of the alcohol groups of Ser, Thr and Tyr residues, and the addition of small proteins of the Ubiquitin (Ub) family to lysine residues[1]. The consequences of PTMs on protein structure and function are as varied as the modifications themselves. PTMs in many cases affect the chemical properties of the polypeptide chain resulting in conformational changes and/or changes in surface properties. Such changes may have a number of outcomes including allosteric regulation of binding interactions or enzyme activity, changes in oligomerisation state, or altered stability [1,4]. In a large number of cases PTMs result in recognition by specific PTM binding

domains which either recognise the modified target surface or directly recognise the PTM to facilitate downstream processes[1,2].

## 1.1 The Ubiquitination System

Ubiquitination and phosphorylation have emerged as the two key players in cell signalling events controlled by PTMs. Phosphorylation has been widely regarded as one of the most important cellular process since its discovery in the mid 1950s, resulting in the award of a Nobel Prize in Physiology and Medicine in 1992 to Edmond Fischer and Edwin Krebs for its discovery[5]. Protein phosphorylation affects the majority of proteins and mediates a response both through allosteric regulation of the modified protein and by recognition by a host of receptors containing phosphoprotein binding domains such as the SH2 domain[1]. Ubiquitination was discovered some 30 years later as a PTM involving the covalent attachment of the 76 amino acid protein Ub to lysine residues in target proteins. This was found to regulate protein stability resulting in protein degradation. Like phosphorylation the importance of the ubiquitination system was recognised in 2004 with the award of the Nobel Prize in Chemistry to Aaron Ciechanover, Avram Hershko and Irwin Rose for the discovery of the Ub system.

Ubiquitination is achieved via a highly specific three step enzyme system (Figure 1-1); the first step involves a Ub-activating enzyme termed the E1 which adenylates the C-terminus of Ub using ATP followed by the formation of a thioester bond between C-terminus of Ub and a catalytic cysteine in the E1. The second step is the 'charging' of an ubiquitin-conjugating enzyme or E2 by transfer of the thioester-linked Ub from the E1 to form a thioester with an active site cysteine in the E2 enzyme. The third and final step involves a complex between the 'charged' E2 a ubiquitin ligase or E3 ligase and the target substrate whereby the E3 facilitates transfer of the Ub unit from a thioester with the E2 to form an isopeptide bond between the C-terminal carboxylate of Ub and the ε-amino group of a target lysine side chain (Figure 1-1)[6]. Ubiquitin contains seven lysine residues and can be ubiquitinated to form poly-Ub chains of different linkage types. Ubiquitination of target proteins and poly-Ub chains can be disassembled by deubiquitinating enzymes (DUBs) in order to reverse the ubiquitination process[6,7]. To date there are only two known E1 enzymes in the human proteome, while around 40 E2s, approximately 600 E3s and 90 DUBs are known[7]. The large number of E2 and E3 enzymes has led to

the suggestion that target specificity (including poly-Ub linkage specificity) is determined by the many potential combinations of E2 and E3 enzyme and this is an area of intense research in the field of ubiquitination[2,7].

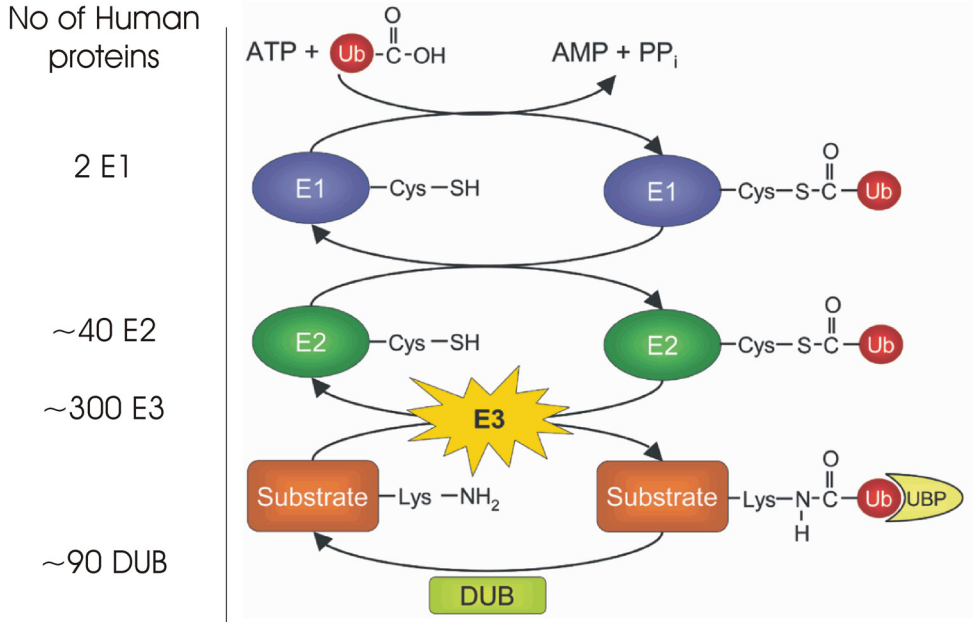



**Figure 1-1: The three step ubiquitination system taken from Passmore *et al* [6]. The approximate numbers of different enzyme type are included.**

### 1.1.1 Types of Ubiquitination

Ubiquitination can occur in a number of forms associated with distinct functions. The simplest level of variation in ubiquitination is in the number of Ub moieties; proteins can be mono-ubiquitinated where only a single Ub molecule is attached at a single position, multi-ubiquitination where mono-Ub is attached at multiple sites along the polypeptide chain or poly-ubiquitination where a chain of Ub (ubiquitinated Ub) is attached to a target protein at a single or multiple sites. The seven lysines and the N-terminus of Ub can all function as ubiquitination sites for the assembly of poly-Ub chains, the formation of poly-Ub chains of different linkage types and lengths creates a huge number of possibilities for Ub mediated signalling making the ubiquitination system an incredibly complex system[2,7].

Like phosphorylation ubiquitination is a reversible, highly regulated, PTM which mediates protein-protein interactions and alters the properties of target proteins. However by comparison phosphorylation is a much simpler modification, the larger size of Ub presents a larger surface for interaction with PTM binding

domains compared to phosphorylation and the ability of Ub to form polymers makes the ubiquitination system a much more complicated one[2].

The first identified poly-Ub linkage, which has since been demonstrated to be the most common Ub polymer *in vivo*, is the Lys48-linked poly-Ub chain[8]. In the early years of Ub research it was noted that the addition of Lys48 linked poly-Ub resulted in rapid degradation of the target protein by the 26 proteasome, a large protein complex which possess chymotrypsin-like, trypsin-like and peptidyl-glutamyl peptide-hydrolyzing-like proteolytic activity [9-11]. The research which followed identified the Ub-proteasome system (UPS) as the principal mechanism for selective degradation of cellular proteins. The turn over of key proteins including repressor proteins, transcription factors, and enzymes by the UPS is vital for cellular homeostasis and controlled cellular responses[2].

In a recent quantitative study the entire ubiquitin content of yeast was analysed, and relative abundances of ubiquitin linkages was determined[8]. Although Lys48-linked poly-Ub was the most commonly detected linkage type it only represented 29 % of the measured poly-Ub content, with the Lys11-linkage the second most popular with 28 % followed by Lys63 with 17 % and Lys6, Lys27, Lys29 and Lys33 with 11, 9, 3 and 3 % respectively[4]. Previous to this, the confirmation of all of the available chain linkages *in vivo* had not been established.

In recent years the functions of the different ubiquitination forms other than Lys48-linked are beginning to emerge; mono-ubiquitination has been linked to the regulation of histones and DNA damage, while multi-mono-Ub in conjunction with Lys63 linked polyubiquitination of cell-surface receptors is associated with internalisation and lysosomal degradation of those receptors[2]. Like Lys48-linked chains, poly-Ub linked via Lys11 and Lys29 have been linked to protein degradation via the UPS. Poly-Ub linked via Lys63 and Lys6 have links to DNA damage. Lys63-linked chains also play a key role in the regulation of cell signalling in particularly in pathways linked to NF-κB regulation. Recent reports have also highlighted the *in vivo* presence of linear poly-Ub chains with implications in cell signalling and kinase activation in the NF-κB pathway[12]. Another exciting field in which Ub mediated signalling appears to play a central role is that of selective autophagy and in particular the removal of misfolded protein aggregates, organelle and bacteria via the action of the Ub binding autophagy receptors p62 and NBR1, the linkage specificity

and processes which distinguish between protein degradation by the UPS or by autophagy is one where many questions remain.

## 1.2 Ubiquitin Binding Domains and Interpreting the Ubiquitination Code

While ubiquitination has the potential to alter the chemical properties of the modified protein[13], both its proteolytic and non-proteolytic functions are more commonly mediated via interaction with receptors containing a host of different Ub binding domains (UBDs). The many potential forms of ubiquitination are reflected in the large number of known UBD-containing proteins, to date there are over 20 known UBD families and over 200 proteins in the human genome which are predicted to contain at least one UBD[14-17]. UBDs are diverse in their secondary structure content, in their size, and in their affinity for Ub. However the molecular detail of how these different UBDs successfully discriminate between different ubiquitination types remains unclear in all but a few cases. Despite the versatility of chain linkage types and lengths poly-Ub chains are essentially repeating units of the same protein, it is therefore of particular interest how UBDs successfully interpret ubiquitination to mediate the correct process based on subtle differences.

### 1.2.1 Canonical Ubiquitin Binding Domains

The first described Ub binding domain was the Ub associated (UBA) domain[18] which, like the overwhelming majority of UBDs which have been identified since the UBA domain, binds to the hydrophobic β-sheet surface of Ub centred on the pairing of Ile44 and Val70 (Figure 1-2). Of the known UBDs over 17 families bind to the hydrophobic β-sheet with subtle variations in the contacts made with surrounding regions (Figure 1-1). As a result of the large number of binding partners the hydrophobic binding surface with Ile44 at its centre is considered the canonical Ub binding surface.

Early investigations had proposed links between UBD families and function, with the suggestion that UBA domains are almost exclusively associated with the UPS, however this has since proven not to be the case with no clear correlation between a given UBD type and a given function or cellular process[16].

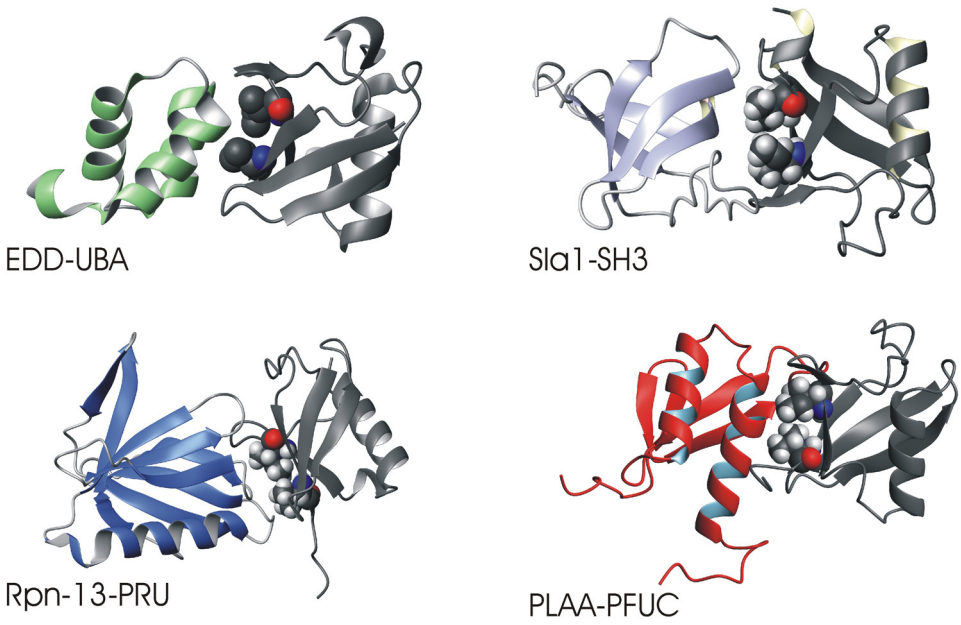



**Figure 1-2: Example of various Ile44 binding UBDs in complex with Ub including the UBA of EDD, the SH3 domain of Sla1 the PRU domain of Rpn13 and the PFUC domain of PLAA (PDB ID: 2QHO, 1Q0W, 2JT4, 1Q5W, 2Z59, and 2K8B) in all cases Ub is coloured Grey and Ile44 and Val70 are drawn as 'space filling'.**

### 1.2.2 Non Canonical Ubiquitin Binding Domains

Although for many years only UBDs which bound to the Ile44 binding site were identified, Ub presents around 450 $Å^2$ of solvent exposed surface area for interaction with binding partners. As such it is unsurprising that in recent years a number of non-canonical UBDs have been identified which utilise distinct binding patches including a generally polar surface centred on Asp58 utilised by the A20 zinc finger domain, the C-terminal flexible tail of Ub bound by the UBP domain, and Leu8 bound by the UBM domain (Figure 1-3)[15,16,19].

UBDs which utilise non-Ile44 based binding surfaces have the advantage of not competing with the many other Ile44 binding UBD families; however the functional significance of this has yet to be explored. New classes of UBD are continuously being reported giving new insights into new areas of Ub mediated signalling and expanding the current understanding of the complicated ubiquitination system.

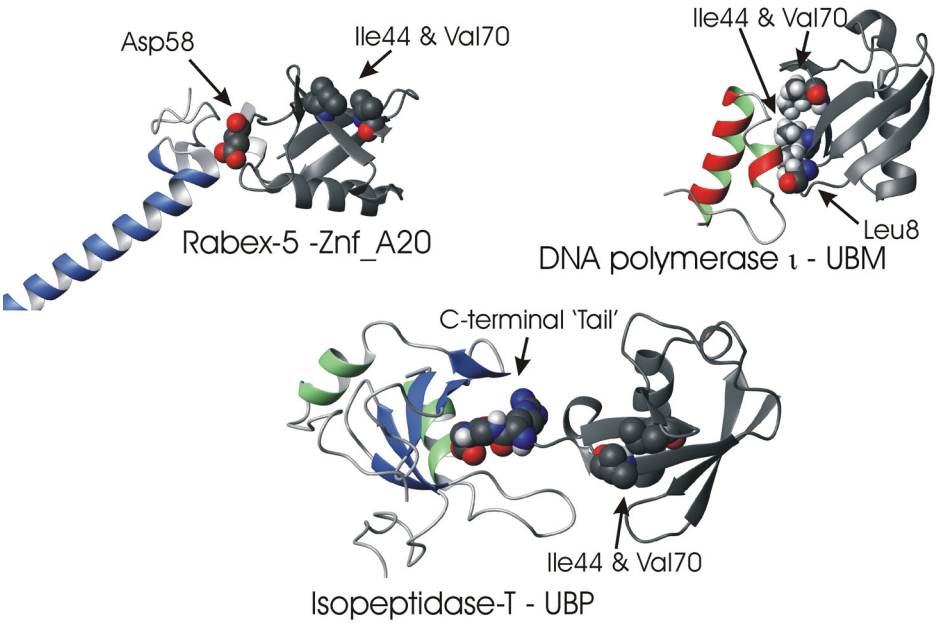



**Figure 1-3: Non-canonical Ub binding UBDs in complex with Ub including the Znf_A20 of Rabex-5, The UBP of isopeptidase T and the UBM of DNA polymerase ι (RPDB ID: 2C7N, 2G45 and 2KHW).**

The emergence of non-Ile44 binding UBDs opens up the interesting prospect of cooperation between UBDs. The binding of the majority of UBDs to the same binding patch indicates an evolutionary disadvantage to simultaneous interaction of two UBDs with Ub. Where UBDs are involved in very different downstream processes the limiting of Ub to binary interactions with UBDs is logical to prevent ambiguity in the down-stream response. However with more non-Ile44 binding UBDs being reported the prospect of cooperation or cross talk between different UBDs through the formation of Ub-mediated multi-UBD complexes may become a factor. A recent investigation into the interaction of Rabex-5 with Ub demonstrated that such a potential ternary complex was possible. Two independent groups observed two molecules of Rabex-5 contacting a single molecule of Ub using two different UBDs in the crystal lattice demonstrating the potential for such ternary complexes[20,21].

### 1.2.3 Poly-Ubiquitin Linkage Specificity

Precisely how the different properties of UBDs are used to respond to the various different forms of Ub involved in different cellular-processes remains unclear. The versatility of Ub to generate different signals from a simple building

block originates from its ability to form oligomers of different lengths and linkage types. It may therefore be expected that the different properties of UBD-families facilitate the interpretation of Ub-mediated signalling through poly-Ub recognition which is specific to a given chain length or linkage type. Despite this the majority of structural investigations of UBD interactions with Ub have focused only on mono-Ub, with only a handful of exceptions.

A number of UBA domains have demonstrated specificity for Lys48, Lys63 and linear-Ub. In the case of Lys48 specificity the UBA can insert between the two adjacent Ub moieties in a poly-Ub chain and make contact simultaneously with both Ub units. The contacts made to both Ub are determined largely by the conformational restrictions imposed by the linkage type and therefore effectively confer a linkage specificity (Figure 1-4)[22,23]. A similar mechanism describes the specificity of the NZF domain of TAB2 for Lys63 linked poly-Ub, which again sandwiches between two Ub units with contacts which are defined by the linkage type (Figure 1-4)[24]. In a number of other cases it is the spatial arrangement of multiple Ub interacting sites which determines the specificity. In the case of RAP80 which has two Ub interacting motifs (UIM) separated by a linker of conserved length, the extended nature of the UIM domain and of the spacing between them confers specificity for the highly extended Lys63 type poly-Ub chains (Figure 1-4)[2,15,16]. In a similar case, the positioning of two Ub interaction sites within the UBAN domain of NEMO allows the perfect spacing for recognition of the similarly extended conformation of linear Ub. Interestingly despite being very similar in appearance the UBAN is highly selective for linear Ub over Lys63 Ub demonstrating the exquisite specificity conferred by many UBDs[2,25,26].

Determining the specificity of Ub recognition by UBDs, and the molecular basis for that specificity (or lack of), is central to the understanding of Ub mediated signalling and the molecular detail of the associated cellular processes.

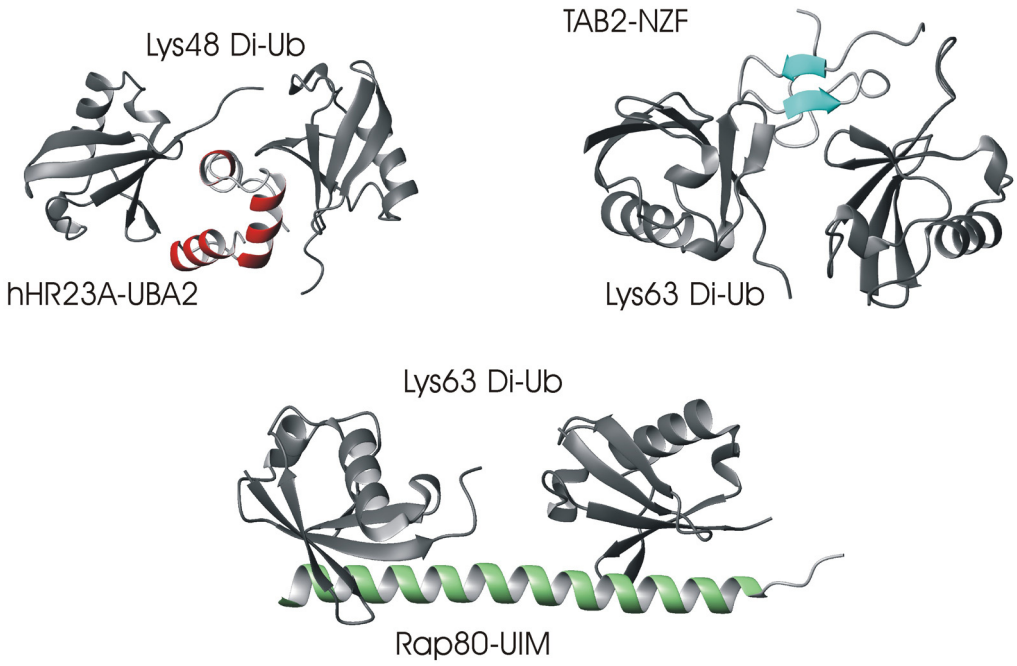



**Figure 1-4: UBD interactions with poly-Ub chains including the UBA2 of hHR23A with Lys48 linked Ub, the NZF of TAB2 in complex with Lys63 linked di-Ub and the DUIM of RAP80 in complex with Lys63 di-Ub (RPDB ID: 1ZO6, 3A9J and 3A1Q).**

## 1.3 NF-κB and Ubiquitin Mediated Signalling

The transcription factor NF-κB controls many processes including immunity, inflammation and apoptosis. As a result its activation is highly regulated with the majority of upstream processes being subject to regulation[27]. The regulation of NF-κB utilises a combination of PTMs including phosphorylation and a number of different Ub-modifications utilising at least three different poly-Ub chain linkage-types including Lys48, Lys63 and linear poly-Ub[27,28].

Under normal conditions NF-κB is retained in the cytoplasm by the inhibitor of NF-κB (IκB). IκB is under the control of the IκB Kinase (IKK) complex which phosphorylates IκB leading to its Lys48 linked polyubiquitination and degradation by the UPS. This frees NF-κB to translocate to the nucleus, where it drives gene expression. The role of Ub in NF-κB activation expands beyond UPS mediated degradation of the repressor element IκB, it is also intimately involved in activation of the IKK complex. The Lys63 E3 ligase TRAF6 is intimately involved in NF-κB activation. One of the proposed mechanisms is based on the activation of the TAK1

kinase complex which phosphorylates and activates the IKK complex. TRAF6 in conjunction with the E2 complex of UBC13 and UEV1A generates Lys63 poly-Ub chains on key substrates including RIP1. These are recognised by the binding partner of TAK1, TAB2, utilising a Lys63-specific NZF domain. Binding of TAB2 to Lys62-linked poly-Ub results in dimerisation and autophosphorylation of TAK1 to generate active TAK1 which then proceeds to phosphorylate and activate the IKK complex[29].

Ub also plays a role in direct activation of the IKK complex; the IKK complex is comprised of three components, the two kinases IKKα and IKKβ, and the regulatory subunit IKKγ which is more commonly known as NEMO. NEMO harbours a UBD termed a UBAN which shows high selectivity for linear Ub. The binding of linear Ub causes a significant conformational change which has been suggested to facilitate direct activation of the IKK complex by its regulatory subunit NEMO. In addition to the known roles of Ub in NF-κB a whole host of NF-κB regulating poly-Ub binding proteins of unknown role have recently been identified using micro arrays. There has also been a strong link between the activity of a number of DUBs and negative regulation of NF-κB including the TN-Fα inducible protein 3 (TNFaIP3), Cezanne and CYLD[30-32].

The NF-κB pathway offers an ideal model for investigating the interplay between various aspects of the ubiquitination system as well as being one of the most important cellular factors with links to a number of human diseases[33].

## 1.4 Summary, Aims and Objectives

The ubiquitination system represents a challenging cell signalling and regulatory system with vital roles in many cellular processes. The complexity of the Ub system arises from the size of the PTM which presents a number of binding surfaces for recognition by different binding partners, and more significantly by the formation of poly-Ub chains of distinct linkage types. The mechanism by which UBDs interpret the ubiquitination 'code' has been determined in a limited number of cases and there is a considerable amount which remains unknown.

The investigations presented in this thesis focus on two UBDs with links to the UPS and the regulation of NF-κB and which have links to human disease. The p62 protein is intimately involved in the regulation of NF-κB and autophagy. p62 harbours a C-terminal UBA domain which interacts non-covalently with Ub and has

been identified as a mutation 'hot spot' in Paget's disease of bone (PDB). PDB associated mutations in p62 appear to influence the recognition of Ub by the p62-UBA domain resulting in dysregulation of NF-κB activity. The mechanism by which the p62-UBA regulates its own Ub recognition is unique to UBA domains and offers interesting prospects for the mechanism by which its Ub related functions are disrupted by disease.

The other UBD-containing protein investigated in this thesis is the ZNF216 protein a novel-binder of Ub with links to the UPS, in particular the shuttling of muscle proteins to the proteasome during muscle atrophy. The ZNF216 protein binds to Ub using a non-canonical binding patch centred on Asp58 via a Znf_A20 domain. To date only a single example of a Asp58-binding UBD has been characterised leaving a number of questions over the key features of the interaction which define the binding patch. For example it is not known if the binding site and binding orientation is conserved across the Znf_A20 family or shows variation similar to that seen for the UBA family.

In order to gain insight into the function of these two proteins in disease and to contribute to the ever expanding body of understanding of the ubiquitination system, both of these two domains and their interaction with Ub were characterised structurally and thermodynamically. In Chapter 9 of this thesis the prospect of cooperation between these two domains, which share common functional pathways and several binding partners, has been investigated. This identified an exciting ternary complex between the two UBDs and Ub *in vitro*, this observation has been extended to examine all UBD families with interesting results.

# Chapter 2

## 2 Materials and Methods

### 2.1 Molecular Biology Techniques

Both the human p62/SQSTM1 and ZNF216 genes and various gene segments had been previously cloned into the pGEX-4T-1 plasmid by the group of Dr Robert Layfield and as such were available for transformation into appropriate *Escherichia coli* (*E. coli*) cell lines. The gene sequence for yeast ubiquitin was cloned previously into a pkk223-3 plasmid by Dr Geoff Platt[34]. The restriction sites and plasmids used for the various constructs used during the studies outlined here are detailed in Table 2-1. The wild type (wt) gene sequences encoding the proteins used in this study can be found in Appendix A, the encoded protein sequences are as follows:

p62 (*Homo sapiens*)

MASLTVKAYLLGKEDAAREIRRFSFCCSPEPEAEAEAAAGPGPCERLL
SRVAALFPALRPGGFQAHYRDEDGDLVAFSSDEELTMAMSYVKDDIFR
IYIKEKKECRRDHRPPCAQEAPRNMVHPNVICDGCNGPVVGTRYKCSV
CPDYDLCSVCEGKGLHRGHTKLAFPSPFGHLSEGFSHSRWLRKVKHGH
FGWPGWEMGPPGNWSPRPPRAGEARPGPTAESASGPSEDPSVNFLKNV
GESVAAALSPLGIEVDIDVEHGGKRSRLTPVSPESSSTEEKSSSQPSS
CCSDPSKPGGNVEGATQSLAEQMRKIALESEGRPEEQMESDNCSGGDD
DWTHLSSKEVDPSTGELQSLQMPESEGPSSLDPSQEGPTGLKEAALYP
HLPPEADPRLIESLSQMLSMGFSDEGGWLTRLLQTKNYDIGAALDTIQ
YSKHPPPL*

ZNF216 (*Rattus norvegicus*)

MAQETNQTPGPMLCSTGCGFYGNPRTNGMCSVCYKEHLQRQQNSGRMS
PMGTASGSNSPTSDSASVQRADATLNNCEGAAGSTSEKSRNVPVAALP
VTQQMTEMSISREDKITSPKTEVSEPVVTQPSPSVSQPSSSQSEEKAP
ELPKPKKNRCFMCRKKVGLTGFDCRCGNLFCGLHRYSDKHNCPYDYKA
EAAAKIRKENPVVVAEKIQRI*

Ubiquitin (*Saccharomyces cerevisiae*)

MQIFVKTLTGKTITLEVESSDTIDNVKSKIQDKEGIPPDQQRLIFAGK
QLEDGRTLSDYNIQKESTLHLVLRLRGG*

In the above sequences regions colored blue represent the main isolated domain constructs listed below.

In order to improve solubility during expression and aid purification all p62 and ZNF216 constructs were expressed as N-terminal fusions with Glutathione S-transferase (GST) with a thrombin cleavage site located between GST and the desired protein, leaving either a 'GS' or 'GSPEF' peptide at the N-terminus of the protein as an artifact from the cleavage site.

The ZNF216 1-60 construct was generated by the incorporation of a premature stop codon in ZNF216 at residue 61.

**Table 2-1: Table of Plasmids information for the proteins used in this study.**

| Clone | Plasmid | Cloning site | Affinity tag | Tag (Final) |
|---|---|---|---|---|
| Yeast ubiquitin | pKK233-3 | EcoR1-HindIII | NA | NA |
| p62 387-436 (UBA)* | pGEX-4T-1 | BamH1-Xho1 | GST | GS |
| p62 341-440* | pGEX-4T-1 | EcoR1-Xho1 | GST | GSPEF |
| p62 261 – 440 | pGEX-4T-1 | EcoR1-Xho1 | GST | GSPEF |
| p62 221-440 | pGEX-4T1 | EcoR1-XhoI | GST | GSPEF |
| ZNF216 1-220 | pGEX-4T-1 | BamH1-Xho1 | GST | GS |
| ZNF216 1-60 (Znf_A20)* | pGEX-4T-1 | BamH1-Xho1 | GST | GS |
| ZNF216 150-213 (Znf_AN1) | pGEX-4T-1 | BamH1-Xho1 | GST | GS |

*Mutants of these constructs are within the same construct as listed here.

From here onwards yeast ubiquitin, p62 387-436, ZNF216 1-60 and ZNF216 150-213 will be referred to as ubiquitin, p62-UBA, Znf_A20 and Znf_AN1 respectively.

### 2.1.1 Sterilization

All pipette tips, glassware, growth media, and buffers were sterilized by heating at 121 °C for 15 minutes in an autoclave (Philip Harris, Status Autoclave). Heat sensitive reagents were sterilized with a sterile 0.22μm syringe filter (Sartorius stedim). Disposables including 30 ml universals, Petri dishes, syringes, and needles were supplied sterile from the supplier.

All bench spaces were cleaned at regular intervals with TriGene (TriGene) and wiped down with 100 % ethanol before cell work or growths were carried out.

All cell work and growths were carried out in close proximity to a gas flame provided by a Bunsen burner.

All glassware, media and materials which were exposed to *E. coli* were sterilised by autoclavation or exposure to TriGene for 15 minutes. Disposables were autoclaved followed by incineration.

### *2.1.2* Overnight *E. coli* Cultures

Many of the molecular biology techniques require the use of overnight cultures. 15-20 ml Luria-Bertani (LB) broth containing, where appropriate, 270 µM ampicillin, were inoculated with *E. coli* picked as single colonies from LB agar plates or from *E. coli* glycerol stock and grown at 37 °C with agitation at 180 rpm for 16 hours before being used as appropriate.

XL1-Blue, BL21 (DE3) and C41 (DE3) *E. coli* strains were used in this study.

*Solutions:*

LB broth (Sigma): 10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl, autoclaved.

LB Agar (Sigma): 10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl, 15 g/L agar, autoclaved.

### 2.1.3 *E. coli* Glycerol Stocks

All *E. coli* strains used in this study were stored at -80 °C utilising glycerol as a cryoprotectant. 500 µl of overnight culture were mixed with 500 µl of 30 % glycerol in sterile eppendorfs and stored at -80 °C.

### 2.1.4 Agarose Gels

To monitor a number of molecular biology techniques it was necessary to visualise DNA samples and to estimate their content by size using agarose gels.

5µl of DNA was added to 1µl of 6 x DNA loading dye (Promega). A 2log DNA ladder (NEB) was run to check the size of DNA fragments.

Gels were set up in a Mini-sub® cell GT apparatus (BioRad) according to the manufacturer's recommendations. 1% agarose gels were prepared by dissolving 0.3g of agarose in 30ml of TAE and heated in a microwave until clear. Gel was cooled to 50 ºC prior to the addition of 3µl of ethidium bromide, poured and allowed to set for 30 minutes.

Gels were run with constant voltage of 100 V until sufficient separation had occurred. DNA bands were visualised using a G:BOX UV transilluminator (Syngene).

*Solutions:*

TAE: 40 mM tris(hydroxymethyl)aminomethane HCl (Tris.HCl) pH 8, 1 mM ethylene diamine tetraacetic acid (EDTA), 20 mM Acetic acid.

### 2.1.5 Polymerase Chain Reaction (PCR)

Forward and backward mutagenic primers were designed in accordance with the QuickChange® site-directed mutagenesis kit manual (Stratagene) and purchased from the School of Biomedical Sciences Biopolymer synthesis and analysis laboratory (University of Nottingham). PCR site directed mutagenesis were performed using the QuickChange® site-directed mutagenesis kit. A thin walled PCR tube was prepared containing a 50 µl PCR mixture containing template DNA (10 ng), excess forward and reverse primers (125 ng), 1 x reaction buffer, and 1 mM dNTP mix before addition of *Pfu Turbo*® DNA polymerase (2.5 U). The PCR mixture was cycled using a Techgene PCR thermal cycler (Techne) according to the following program.

| | | | |
|---|---|---|---|
| 1 cycle | 95 °C | for | 30 seconds |
| n cycles | 95 °C | for | 30 seconds |
| | 55 °C | for | 60 seconds |
| | 68 °C | for | 1 minute/kb plasmid template DNA length |

(Point mutation n = 12, single amino acid changes n = 16, multiple amino acid changes n = 18).

Following PCR reactions N-6 methylated template DNA was removed by digestion with the *Dpn* I restriction endonuclease (10 U) at 37 °C for 1 hour before cooling to 4 °C prior to transforming into *E. coli* XL1-Blue.

A large set of the mutations used in this study had been made previously within the Searle group[34-36]. The following new mutations were made for the purpose of this study.

| | |
|---|---|
| Ubiquitin | K48C |
| ZNF216 | S61X (Stop) |
| Znf_A20 | F22Y (Made by Miss Joanna Strachan) |
| | V34K (Made by Miss Joanna Strachan) |

The sequences of the primers used for the generation of the above mutants can be found in Appendix B.

*Solutions:*

10 X Reaction Buffer: 10 mM KCl, 10 mM $(NH_4)SO_4$, 20 mM Tris.HCl pH 8.8, 2 mM $MgSO_4$, 0.1% Triton® X-100, 0.1 mg/ml Bovine serum albumin (BSA).

## 2.1.6 Transformations

### 2.1.6.1 Preparation of super competent cells

10 ml of LB broth were inoculated 1:50 with *E. coli* overnight culture and grown at 37 °C with agitation at 180 rpm until the optical density (OD) at 595 nm reached 0.3. At this time the cell cultures were harvested using a primo-R centrifuge (Heraeus) at 1000 g for 10 minutes. The supernatant was discarded and the cell pellet resuspended in 4 ml of 50 mM $CaCl_2$ at 4 °C for 3 hours to generate super competent cells. Super competent cells were centrifuged at 1000 g for 10 minutes, the supernatant discarded and the cell pellet resuspended in 800 µl of 50 mM $CaCl_2$, cells were separated into 100 µl aliquots for later use.

### 2.1.6.2 Heat Shock transformation

1-2 µl of DNA (100 ng/µl) was mixed with 100 µl of competent cells and incubated at 2 °C for 30 minutes. Heat shock was then performed by incubation at 42 °C for 45 seconds followed by cooling to 2 °C for three minutes prior to the addition of 400 µl of LB broth. Each transformation reaction was incubated at 37 °C with agitation at 180 rpm for 30 minutes before plating 250 µl onto an AGAR plate containing 270 µM ampicillin and incubated at 37 °C for 16 hours. For positive and negative controls plasmid DNA was replaced with pUC18 plasmid (Stratagene) or distilled water respectively.

## 2.1.7 DNA Isolation

Overnight culture containing *E. coli* XL1-Blue transformed with the desired plasmid was centrifuged at 1500 g for 10 min at 4ºC and the supernatant discarded. Plasmid DNA was isolated from the cell pellet using the QIAprep® spin Miniprep system (Qiagen) as per the manufacturer's instructions. This method relies on alkaline lysis of bacterial cells and subsequent absorption of the plasmid DNA upon

silica in the presence of high salt [37,38]. DNA is then eluted using 60 µl of elution buffer and purity confirmed by agarose gel (Section 2.1.4).

*Solutions:*

Elution Buffer: 10 mM Tris.HCl pH 8.5.

### 2.1.8 Sequencing

10µl DNA samples of (100 ng/µl) were submitted with 20 pmol/µl of plasmid primer to the School of Biomedical Sciences Biopolymer synthesis and analysis laboratory (University of Nottingham). Sequencing results were visualised using Chromas 2.21 (Technelysium).

## 2.2 Protein Overexpression and Purification

### 2.2.1 Small Scale Overexpression

In order to assess the protein expression quantity and solubility resulting from the plasmid transformed into either BL21(DE3) or C41(DE3) strains of *E. coli*, small scale over expressions were carried out under various conditions. 15 ml of LB broth containing ampicillin at 270 µM were inoculated 1 in 50 with an overnight culture of the desired transformant and grown at 37 °C with agitation at 180 rpm, until an OD at 595 nm of around 0.4 was reached. At this point the conditions were altered to match those at which expression was to be tested and incubated until the OD at 595 nm was between 0.6-0.7. At which point overexpression was induced by the addition of Isopropyl-β-D-Thiogalactopyranoside (IPTG) to a final concentration of 1 mM. Small scale over expressions were performed at different temperatures including 37, 30, 25 and 20 °C and incubated for up to 20 hours after expression was induced. After each small scale over expression a 1 ml sample was taken and centrifuged at 13,000 rpm, for 3 minutes in a Biofuge-pico bench top centrifuge (Heraeus). The supernatant was discarded and the pellet retained for analysis of the cells protein content by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE). Samples were similarly taken prior to induction as an indicator of back ground protein.

Each cell pellet was resuspended in lysis buffer A at a volume of 500 µl per unit of the final OD at 595 nm. Samples were sonicated at 8 microns for a total of 32 seconds in 8 second burst using a Soniprep 150 sonicator (MSE) and clarified by

centrifugation at 13,000 rpm for 10 minutes. 40 µl samples were taken prior to clarification representing the total cell lysate and after centrifugation representing the soluble fraction and analysed by SDS-PAGE.

*Solutions:*

Lysis buffer A: 10 mM Tris.HCl pH 7.4, 150 mM NaCl.

### 2.2.2 SDS-PAGE

SDS-PAGE was utilised to analyse protein containing samples by separation based on molecular weight allowing confirmation of desired protein product and an approximation of quantity.

The gel apparatus was set up according to the manufacturer's recommendations (BioRad). Polyacrylamide gels (PAG) were prepared incorporating a stacking gel above a resolving gel using the Mini-Protean®3 system (BioRad). Polymerisation reactions were instigated by the addition of N,N,N,N-Tetramethylethylethylendiamine (TEMED) to a final concentration of 0.1% (v/v) immediately prior to pouring.

Protein samples were boiled in 1 volume equivalent of 2 X gel loading buffer and loaded in 5-15 µl volumes alongside low molecular weight markers onto the stacking gel.

Gels were run on a BioRad SDS-PAGE kit in SDS running buffer at 140 V for around 1 hour. Gels were stained in protein stain for 1-3 hours before de-staining in destaining solution for 16 hours and visualising on a G:BOX light box (Syngene).

*Solutions:*

Stacking gel: 2 % Protogel® (15 % (w/v) acrylamide 0.4 % (w/v) bis-acrylamide (National Diagnostics)), 125 mM Tris.HCl pH 6.8, 0.1 % SDS, 1% (w/v) APS.

Resolving gel (20 %): 20 % Protogel®, 375 mM Tris.HCl pH 8.8, 0.1% (w/v) SDS, 1 % (w/v) APS.

2 X Loading Buffer: 100 mM Tris pH 6.8, 4% (w/v) SDS, 0.2 % (w/v) bromophenol blue, 20 % glycerol, 200 mM dithiothreitol (DTT).

Molecular Weight Markers: Albumin (66 kDa), Ovalbumin (45 kDa), glyceraldehyde-3-phosphate dehydrogenase (36 kDa), carbonic anhydrase (29 kDa), trypsinogen (24 kDa), trypsin inhibitor (20 kDa), α-lactalbumin (14.2 kDa) and aproprotinin (6.5 kDa).

SDS running buffer: 25 mM Tris.HCl, 250 mM Glycine, 0.1 % SDS pH 8.3.

Protein Stain: 0.25 % (w/v) Brilliant Blue R-250 (Fisher), 45 % (v/v) methanol, 10 % (v/v) glacial acetic acid.
Destaining Solution: 25 % (v/v) methanol, 10 % (v/v) glacial acetic acid.

### 2.2.3 Large Scale Overexpression

Large scale over expression of protein in *E. coli* was performed in 1 litre volumes of LB broth containing 270 μM ampicillin in a 2 litre baffled flask. Each growth was inoculated with 20 ml of overnight culture containing the desired transformant and grown at 37 °C with agitation until the OD at 595 nm reached 0.4. At this point the conditions were altered to the optimal conditions for expression as determined from the small scale over expression (Section 2.2.1) and grown until the OD at 595 nm reached 0.6. The cell culture was then induced by the addition of IPTG to a final concentration of 1 mM, at this point, and at appropriate intervals thereafter, ampicillin was re introduced to the culture to a total concentration of 270 μM. The induction conditions and times for each of the proteins used in this study are listed in Table 2-2. After induction cells were harvested by centrifugation at 4,500 g for 10 minutes in an Avanti J-25 I centrifuge with a JLA-10.500 rotor (Beckman). The supernatant was discarded and the pellet resuspended in 20 ml of lysis buffer A before centrifuging at 4,500 g for 10 minutes. The supernatant was discarded and the pellet stored at -80 °C for later use.

### 2.2.4 Minimal Media Growth

For growths of transformants containing ubiquitin LB broth was replaced with M9 minimal media. 1 litre of M9 minimal media was prepared in a 2 litre baffled flask by combining M9 salts with 20 ml of sterile filtered (0.22 μm sartorius stedim) M9 nutrient solution, ampicillin to a final concentration of 270 μM and biotin and thiamine to a final concentration of 10 mg/l.

45 ml of overnight culture was centrifuged at 900 g for 10 minutes and the supernatant discarded. The cell pellets were resuspended in 20 ml of M9 minimal media from the 1 litre volume and incubated at 37 °C with agitation at 180 rpm for 1 hour before inoculating into 1 litre of M9 minimal media. Growths from this point were carried out as with LB based over expression (Section 2.2.3).

*Solutions:*

M9 Salts: Calcium chloride dihydrate (15 mg/l), Di-Sodium orthophosphate (6 g/l), Magnesium sulphate (0.5 g/l), Potassium dihydrogen orthophosphate (3 g/l), NaCl (1 g/l).

M9 nutrient solution: 1 g Ammonium Chloride, 4 g glucose in 20 ml Milli-Q water.

### 2.2.5 Isotope Labelling

For investigation by NMR proteins were isotopically labelled with either $^{15}N$ or a combination of $^{15}N$ and $^{13}C$. M9 minimal media was prepared replacing the M9 nutrient solution with $^{15}N$ nutrient solution or $^{15}N/^{13}C$ nutrient solution and performed as above (Section 2.2.4).

In the case of growths including the ZNF216 constructs an additional 2 ml of trace metal solution was added to the M9 minimal media and $ZnCl_2$ added to the media at induction to a final concentration of 25 μM.

*Solutions:*

$^{15}N$ nutrient solution: 1 g $^{15}N$-Ammonium Chloride, 4 g glucose in 20 ml Milli-Q water.

$^{15}N/^{13}C$ nutrient solution: 1 g $^{15}N$-Ammonium Chloride, 4 g $^{13}C$-glucose in 20 ml Milli-Q water.

Trace metal solution: Iron Chloride (200 mg/l), Copper Chloride (10 mg/l), Manganese Chloride, disodium tetraborate (10 mg/l) Ammonium heptamolybdate tetrahydrate (10 mg/l).

### 2.2.6 GST Based Affinity Purification

The fusion of the GST protein to the N-terminus of recombinant proteins aids purification by the intrinsic ability of the GST protein to bind to reduced glutathione; this has been used extensively during this study as outlined below.

Cell pellets (From a 1 Litre growth) containing GST tagged proteins were defrosted at room temperature and resuspended in 7 ml of appropriate lysis buffer with DNAse 1 (5,000 U) (Sigma) and complete™ EDTA free protease inhibitor (Roche) (Table 2-2), before freezing at -80 °C for 20 minutes and thawed at room temperature. The cells were then lysed by sonication at 12 microns in 30 second burst for a total of 6 minutes. Cell lysates were clarified of cell debris by centrifugation at 45,000 g for 30 minutes (x 3) at 4 °C in an Avanti J-25 I centrifuge with a JL-20 rotor (Beckman), followed by filtering through a 0.22 μm syringe filter.

Clarified cell lysate was mixed with 1 ml of glutathione sepharose 4B (GE Healthcare) pre equilibrated with lysis buffer in a disposable gravity flow column, and tumbled at 4 °C for 3 hours. Unbound proteins were eluted and the column washed with 3 column volumes of lysis buffer followed by 2 column volumes of cleavage buffer. In the case of the full length ZNF216 and Znf_AN1 an extra wash step of 1 column volume of 2M NaCl was added to remove DNA.

Recombinant protein was removed from the glutathione sepharose bound GST by an on column cleavage at 4 °C for 18 hours with thrombin (5-10 U) in 1ml of cleavage buffer. Cleaved protein was eluted and the column washed with 1.5 ml of cleavage buffer. The combined 2.5 ml volume was then subject to further purification. In the case of ZNF216 and Znf_AN1 the column was cleaned with an additional 10 ml of Milli-Q water.

**Table 2-2: Growth conditions and lysis buffer used for the proteins used in this study.**

| Protein | Induction temperature /°C | Induction length / Hours | Lysis Buffer |
|---|---|---|---|
| ubiquitin | 30 | 16 | 5 X AE |
| p62-UBA | 30 | 8 | B |
| p62 341-440 | 30 | 8 | B |
| p62 261 -440 | 30 | 8 | B |
| p62 221-440 | 25 | 16 | B |
| ZNF216 | 20 | 18 | C |
| Znf_A20 | 20 | 20 | D |
| Znf_AN1 | 20 | 18 | C |

*Solutions:*

DNAse 1 Stock: DNAse 1 (Sigma) 10 mg/ml (50, 000 U/ml) 1M $MgCl_2$, 0.1M $MnCl_2$.

Lysis buffer B: 10 mM Tris.HCl pH 7.4, 150 mM NaCl, 4 mM DTT, 15 mM EDTA, 0.1% (v/v) Triton X-100.

Lysis buffer C: 10 mM Tris.HCl, pH 7.4 500 mM NaCl, 2 mM DTT, 0.1% (v/v) Triton X-100, 50 μM $ZnCl_2$.

Lysis buffer D: 10 mM Tris.HCl, pH 7.4 150 mM NaCl, 2 mM DTT, 0.1% (v/v) Triton X-100, 50 µM ZnCl2.

Cleavage buffer: 20 mM Tris.HCl, pH 8.4, 250 mM NaCl, 2.5 mM $CaCl_2$, 10 µM $ZnCl_2$ *

*$ZnCl_2$ was only included for ZNF216 preparations.

### 2.2.7 Fast Protein Liquid Chromatography (FPLC)

Fast protein liquid chromatography (FPLC) was used for large scale protein purification using an AKTA prime FPLC system (GE Healthcare) in conjunction with a variety of chromatographic media including ion exchange, gel filtration, and desalting columns. All buffers were autoclaved, filtered using a 0.2 µM cellulose-nitrate filter (Whatman) and degassed under vacuum. Protein fractions were monitored by changes in the absorbance at 280 nm. Fractions shown to contain protein were analysed by SDS-PAGE (Section 2.2.2).

### 2.2.8 Ion Exchange

Cell pellets containing ubiquitin (From 4 x 1 litre growths) were subjected to an initial purification step of cell lysis at pH 4 whereby many proteins precipitate and can be removed by centrifugation. The remaining proteins are removed by cation exchange.

Cell pellets containing ubiquitin were resuspended in 50 ml of 5 X AE buffer and sonicated at 12 microns in 30 second burst for a total of 7 minutes. Cell debris and precipitated proteins were removed by centrifugation at 45, 000 g for 45 minutes at 4 °C, this process was repeated with small additions of 5 ml 5 X AE followed by centrifugation until no further precipitation is witnessed. The supernatant was diluted up to 200 ml with Milli-Q water and loaded using a 50 ml Superloop™ (GE Healthcare), onto a 5 ml SPHP cation exchange column (GE Healthcare) which had been pre equilibrated with cation bind buffer. Bound protein was eluted using a linear gradient of 0-70% cation elution buffer 1 over 100 ml followed by a second gradient of 70 -100% over 20 ml. Both loading and gradient steps were performed at a flow rate of 2.5 ml/min with 10 ml fractions collected throughout.

ZNF216 and Znf_AN1 were subject to a cation exchange step after GST purification (Section 2.2.6). 10 ml of cation bind buffer B was added to the eluted protein from GST purification and loaded onto a SPHP cation exchange column

followed in tandem by an QHP anion exchange column pre equilibrated with cation bind buffer B. The anion exchange column was removed and protein eluted from the cation exchange column with a linear gradient of 0-30 % cation elution buffer B over 70 ml followed by a second gradient of 30-100 % over 30 ml. After completion the cation exchange column was replaced with the anion exchange column and a gradient of 0-100 cation elution buffer over 50 ml was performed. All loading and gradient steps were performed at a flow rate of 2 ml/min with fractions of 10 ml collected throughout.

*Solutions:*

5 X AE: 250 mM Acetic acid, 10 mM EDTA pH 4.0.

Cation bind buffer A: 50 mM Acetic acid, 2 mM EDTA, pH 4.0.

Cation elution buffer A: 50 mM Acetic acid, 2 mM EDTA, 2 M NaCl, pH 4.0.

Cation bind buffer B: 10 mM Tris pH 7.4, 50 μM $ZnCl_2$, 2 mM DTT.

Cation elution buffer B: 10 mM Tris pH 7.4, 50 μM $ZnCl_2$, 2 mM DTT, 2 M NaCl.

### 2.2.9 Gel filtration, Size Exclusion Chromatography

Size exclusion chromatography (SEC) of ubiquitin and p62 constructs was achieved using a HiLoad™ Superdex™ 200 column gel filtration column (GE Healthcare) pre equilibrated in gel filtration buffer A. Freeze dried ubiquitin cation exchange fractions were resuspended in 2.5 ml of Milli-Q water centrifuged at 4,500 g for 10 minutes and loaded onto the gel filtration column. In the case of proteins from GST purifications (Section 2.2.6) EDTA was added to a final concentration of 10 mM and the sample loaded directly onto the gel filtration column. Separation was achieved using a flow rate of 3 ml/min and fractions of 10 ml were collected up to 360 ml after injection.

SEC of Znf_A20 and the p62-UBA mutants E409K and G410K was achieved using a HiLoad™ Sephacryl™ S-100 HR gel filtration column pre equilibrated in gel filtration buffer B and gel filtration buffer A respectively. Separation was achieved using a flow rate of 2 ml/min and fractions of 10 ml were collected.

Size calibration of the HiLoad™ Sephacryl™ S-100 HR gel filtration column was achieved using the low molecular weight (MWt) gel filtration calibration kit (GE Healthcare) recorded in gel filtration buffer B at a flow rate of 2 ml/min.

*Solutions:*

Gel filtration buffer A: 30 mM potassium phosphate pH 7.0 (mix of mono- and di-basic powders), 100 mM NaCl.

Gel filtration buffer B: 10 mM Tris.HCl pH 7.4, 120 mM NaCl, 1 mM DTT, 25 µM $ZnCl_2$.

Protein Calibration Kit: Blue Dextran (2,000 kDa), Aprotinin (6.5 kDa), Ribonuclease A (13.7 kDa) Carbonic anhydrase (29 kDa), Ovalbumin (44 kDa) and Conalbumin (75 kDa).

*Data analysis:*

Determination of protein size was determined by generation of a standard curve using proteins from the calibration kit with the addition of ubiquitin: calibration of the column void volume was calculated using blue dextran (GE Healthcare) and the geometric column volume supplied by the manufacturer (GE HealthCare). The gel phase distribution coefficient of a set of calibration proteins and the investigated protein was calculated using Equation 2-1.

**Equation 2-1:** $$K_{av} = (Ve - V0)/(Vc - V0)$$

Where $K_{av}$ is the gel phase distribution coefficient, Vc is the geometric column volume V0 is the column void volume and Ve is the measure elution volume.

Molecular weight standard curves were constructed as a plot of Log MWt against $K_{av}$ to estimate the apparent MWt of the investigated protein.

Stoke's radius was calculated using two calibration curves; the first is a plot of $\sqrt{-(\log K_{av})}$ against Stoke's radius, following the manufacturers protocol provided with the Low MWt gel filtration calibration kit (GE HealthCare). The second was following the procedure of Siegel *et al*[39] by plotting Stoke's radius against $\sqrt[3]{K_{av}}$. Both approaches give linear plots which can be used to calculate the Stoke's radius of the investigated protein with good agreement between the two approaches (<5%). Aprotinin has been shown to deviate from linearity when used with Sephacryl columns due to potential interaction with the matrix; removal of Aprotinin from the data set improved the quality of the calibration and as therefore not used. The literature values of Stoke's radius for the calibration proteins and the determined calibration curves with and without Aprotinin can be found in Appendix C.

### 2.2.10 Concentration

After each FPLC step and after purification, each sample was concentrated. In most cases proteins were lyophilised by snap freezing in liquid nitrogen and freeze drying under high vacuum. The only exception to this was the ZNF216 full length protein which could not be resuspended after freeze drying and was instead concentrated by spin concentrating using a vivaspin20 spin concentrator (Vivascience) with a 3000 molecular weight cut off PES membrane as per the supplier's instruction.

### 2.2.11 Desalting

Pure protein samples were desalted before being used in biophysical studies. Samples in a volume of 2-2.5 ml were loaded onto a 25 ml HiTrap™ desalting column (GE Healthcare) pre equilibrated with Milli-Q water and run at a rate of 0.5 ml/min, fractions changed manually at the points between initial protein elution and the appearance of trace salt a indicated by changes in conductivity measured by the AKTA prime system.

The full length ZNF216 and Znf_AN1 precipitated when desalted into Milli-Q water and were therefore desalted into appropriate buffer for study and concentrated as required.

### 2.2.12 Purification Overviews

The order of purification steps employed for each protein is out lined below for each protein used in this study.

Ubiquitin: 1. Cation exchange 2.Gel filtration 3. Desalting.

p62 constructs (All) and Znf_A20*: 1. GST purification 2. Gel filtration 3. Desalting.

ZNF216 and Znf_AN1[#]: 1. GST purification 2. Ion exchange.

*Initial purification of Znf_A20 utilised ion exchange followed by desalting but later switched to the above with improved yields.

[#]Initial purification of ZNF216 and Znf_AN1 utilised gel filtration but resulted in very large apparent masses due to an interaction with an unknown impurity which was separated by the inclusion of the high salt wash and ion exchange. The impurity isolated by ion exchange was analysed by SDS-PAGE and ultraviolet (UV) spectroscopy using a NanoDrop ND-100 spectrophotometer (NanoDrop).

p62 341-440 P392L was supplied by Miss Dereen Najat and Dr Robert Layfield.

p62 341-440 P387L was supplied by Dr Jed Long.

### 2.2.13 MTSL Coupling

In order to record paramagnetic relaxation enhancement (PRE) experiments by NMR, the paramagnetic spin label *S*-(2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonothioate (MTSL) (Toronto Research Chemicals Inc) was covalently coupled to the cysteine side chain engineered into the ubiquitin K48C mutant. MTSL was coupled to the single incorporated cysteine by first reducing 30 mg of K48C ubiquitin in 2 ml of 50 mM DTT for 45 minutes prior to desalting (Section 2.2.11). 3 molar equivalents of MTSL were added directly to the desalt fraction and allowed to react in the dark at room temperate for 2 hours before another desalting step to remove un-reacted MTSL. The desalted protein was then freeze dried. A coupling efficiency of over 95% was confirmed by ESI-MS (Section 2.3.3)(Appendix D).

MTSL coupled ubiquitin K48C was reduced where appropriate with a 5 molar excess of buffered sodium ascorbate and reacted for 12 hours.

*Solutions:*

Buffered Sodium Ascorbate: 0.5 M Sodium Ascorbate, 5 mM Tris.HCl pH 7, and 50 mM NaCl.

### 2.2.14 Quantification

Desalted lyophilised protein was quantified by mass weight using an ALJ 120-4 balance (Kern), typical yields from a purification of a GST tagged protein was between 3 – 5 mg of protein per litre of growth medium. Ubiquitin was produced at quantities between 30-50 mg/litre. Where protein could not be desalted or lyophilised concentration was calculated using UV spectroscopy using the OD at 280 determined using a NanoDrop ND-100 spectrophotometer (NanoDrop) using Beer-Lambert's law (Equation 2-2).

**Equation 2-2:** $$OD = \varepsilon cl$$

Where OD is the recorded optical density at a given wavelength, ε is the molar extinction coefficient for that protein at the given wavelength, l is the path length in cm and c is the protein concentration. ε at 280 nm is calculated for proteins

from the additive contributions from cysteines (125 $M^{-1}$ $cm^{-1}$), tyrosines (1490 $M^{-1}$ $cm^{-1}$), and tryptophans (5500 $M^{-1}$ $cm^{-1}$)[40].

## 2.3 Biophysical Techniques

### 2.3.1 UV-CD Spectroscopy

UV - Circular Dichroism (CD) and UV spectra were recorded on an Pi-Star-180 Spectrophotometer (Applied Photophysics) interfaced with an Acorn Archimedes computer with inbuilt software (Applied Photophysics). The optical system was configured with a 75 W Xe lamp, circular light polarizer, and end-mounted photomultiplier. The machine was calibrated with D-camphor sulfonic acid (Sigma) prior to experimental work.

The temperature was regulated using an RTE-300 circulating programmable water bath (Neslab Inc) and a thermoelectric temperature controller (Melcor).

Protein CD spectra were recorded using one of two quartz cuvettes; for protein concentrations in the range of 1.5 – 0.1 mg/ml a 300 µl cuvette with a 1mm path length was used, in the concentration range of 0.15 – 0.01 mg/ml a 3 ml cuvette with a 10 mm path length was used. CD spectra of ubiquitin, p62 constructs and Znf_A20 were recorded in CD buffer A while the ZNF216 and Znf_AN1 were recorded in CD buffer B.

All CD data was converted using the APL data converter version 3 (Applied Photophysics) and analysed and sorted using Excel 2002 (Microsoft) before further analysis by other programs where appropriate.

*Solutions:*

CD Buffer A: 10 mM potassium phosphate pH 7 in Milli-Q water.

CD Buffer B: 100 mM potassium phosphate pH 7 in Milli-Q water.

#### 2.3.1.1 CD Scans

All samples were equilibrated at the appropriate temperature for around 20 minutes prior to experimental measurement. CD spectra were typically recorded in the wavelength range between 190 to 260 nm, data points were recorded at every 1 nm step. The number of counts was set to 10,000, adaptive sampling was enabled and set to 500,000, and entrance/exit slits of 3 nm were employed. CD spectra are

reported as the average of 2-5 scans subtracted from appropriate buffer blanks recorded in the same fashion.

*Data analysis:*

Initial recording of ellipticity were converted to molar ellipticity using Equation 2-3 which corrects for wavelength and concentration and is related to molar CD by Equation 2-4. CD data may also be presented as mean residue ellipticity which is simply equal to the molar ellipticity divided by the number of residues in the protein of study[41].

**Equation 2-3:** $$[\theta] = (100\theta) / (cl)$$

Where [θ] equals the molar ellipticity, θ is the recorded ellipticity, c is the protein concentration and l equals the path length used in cm.

**Equation 2-4:** $$[\theta] = 3298.2\ \Delta\varepsilon$$

Where Δε equals molar CD[42].

#### 2.3.1.2 Equilibrium Unfolding Profiles

Equilibrium thermal unfolding profiles were typically recorded in the range of 278 K up to 368 K with 0.5 K increments. A 30 s settling time at each temperature prior to recording giving a heating rate of around 0.3 K/min. CD spectra were recorded at 278 K and 368 K and the 2 wavelengths with the largest change in ellipticity selected for monitoring the unfolding process during thermal unfolding profiles. In all cases used in this study thermal unfolding profiles were recorded at 222 and 224 nm.

*Data analysis:*

Thermal unfolding profiles can be used to calculate the thermodynamic parameters for the folding transition including the free energy change, enthalpy change, heat capacity, and melting temperature, using the Gibbs-Helmholtz equation detailed in Equation 2-5[42].

Unprocessed thermal unfolding data was fitted to the following equations using the program IGOR Pro Version 5.0.5.7(WaveMetrics).

**Equation 2-5: Thermal Denaturation fitting equations**

**Equation 2-5.1** $$\Delta G = \Delta H\ (1-T/T_M)-\Delta Cp((Tm-T)+T\ln(T/T_M))$$

**Equation 2-5.2** $K = \exp(-\Delta G/(RT)$

**Equation 2-5.3** $\alpha = K/(1+K)$

**Equation 2-5.4** $[\theta]_t = \alpha([\theta]_F-[\theta]_U)+[\theta]_U$

Where ΔG represents the free energy change, ΔH equals the enthalpy change, T equals temperature, $T_M$ represents the melting temperature, ΔCp is the heat capacity change, K is the folding constant, R is the gas constant which is equal to 8.314472 $JK^{-1}mol^{-1}$, $[\theta]_t$ is the ellipticity at a given temperature, $[\theta]_F$ is the ellipticity of the folded state and $[\theta]_U$ is the ellipticity of the unfolded state.

$[\theta]_F$ and $[\theta]_U$ are corrected for temperature dependence by Equation 2-6.

**Equation 2-6** $[\theta]_F = F_c + F_mT$ **and** $[\theta]_U = U_c + U_mT$

Where $F_c$ and $U_c$ are the 100% folded and unfolded ellipticity at 0 K and $F_m$ and $U_m$ are the linear temperature corrections for the folded and unfolded states.

For initial fitting ΔCp is fixed to 0 to improve ease of fitting as advised by Greenfield *et al.* [42]. For fitting dimers or higher order structures Equation 2-5 was modified as described by Greenfield *et al.* [42].

#### 2.3.1.3 Dilution Studies

Dilution experiments were recorded on the p62-UBA and its various mutants. CD spectra and thermal unfolding profiles were recorded in the concentration range of 1.5 mg/ml to 0.1 mg/ml. An initial set of concentrations ranging from around 230 μM to 1 μM were recorded by systematic 1:1 dilutions with CD Buffer A. Any other desirable concentrations were recorded after initial analysis. After each dilution, samples were equilibrated for 1 hour at room temperature before experimentation. All concentrations were confirmed using a NanoDrop ND-100 spectrophotometer (NanoDrop) (Section 2.2.14).

*Data analysis:*

Apparent mid point was calculated at each concentration by Equation 2-5 and variations thereof or by using the first derivative of the folding curve as suggested by Greenfield *et al.* [42]. Apparent midpoint was plotted as a function of protein concentration to estimate the $T_m$ for the dimer and monomer.

#### 2.3.1.4 EDTA Induced Unfolding

In order to examine the structural consequences of removing zinc from the ZNF216 protein constructs, CD spectra were recorded continuously over a period of 2 to 3 hours after the addition of 5 mM EDTA to 30 μM protein. Single scans were recorded approximately one every 11 minutes, and the ellipticity as a function of time measured.

*Data analysis:*

Ellipticity as a function of time was used to measure approximate rate constant by fitting the observed EDTA induced unfolding process to a pseudo-first-order rate constant using Equation 2-7[43,44].

**Equation 2-7:** $$\theta_t = \theta_0 + (\Delta\theta * (1 - \exp(-k * t)))$$

Where $\theta_t$ equals the ellipticity at a given time, $\theta_0$ equals the ellipticity at t = 0, $\Delta\theta$ equals the maximum ellipticity change, k equals the rate constant, and t equals time in seconds.

In the case of full length ZNF216, two approximate rate constants originating from the two domains assuming independent unfolding transitions, fit to a double exponential described by Equation 2-8.

**Equation 2-8:** $$\theta_t = \theta_0 + (\Delta\theta_1 * (1 - \exp(-k_1 * t))) + (\Delta\theta_2 * (1 - \exp(-k_2 * t)))$$

Where $\Delta\theta_1$ and $\Delta\theta_2$ represent the two changes in ellipticity originating from the two domains and $k_1$ and $k_2$ are their respective rate constants.

### 2.3.2 Isothermal Titration Calorimetry (ITC)

Isothermal titration calorimetry (ITC) was performed to characterise protein-protein interactions. In particular to determine the affinity, stoichiometry, and enthalpy change of a given interaction. ITC was used to study the dimerisation of the p62 UBA domain and also the interaction between the Znf_A20 domain and ubiquitin.

ITC was carried out on a VP-ITC instrument (MicroCal) controlled by the VPViewer2000 version 1.4.27 (MicroCal).

All buffers and protein samples were degassed under vacuum for a minimum of 10 minutes before each experiment as to avoid air bubbles and ensure efficient injection and mixing. All ITC experiments were performed at 298 K.

#### 2.3.2.1 Dilution Experiments

Dilution experiments were performed on p62-UBA and its mutants including S399P, E409K, G410K, G411S, and G425R and on p62 341-440 and its mutants P387L and P392L. Typically protein at a concentration between 200 -350 μM in ITC Buffer A was diluted rapidly into the ITC cell containing ITC Buffer A. 55 sequential injections of 5 μl into a 1.424 ml cell with 300 second equilibration time after each injection was employed. A reference power of 5 μCal/sec was used.

The only exceptions to this were the p62-UBA mutants E409K and G410K where an initial concentration of 2 mM was used due to the weak nature of the dimerisation with those mutants.

The endothermic heat pulse resulting from dimer dissociation was measured and subtracted from appropriate blanks using buffer only. A control dilution experiment was also performed under identical conditions on a monomeric protein, namely the 2$^{nd}$ UBA domain from the human homologue of RAD23 A (hHR23A) (Residues 319-363) provided by Dr Jed Long and Miss Jennifer Adlington.

Similarly dilution experiments were recorded on ZNF216 and Znf_A20 recorded as above with initial protein concentrations in the range of 200-500 μM in ITC Buffer B.

*Solutions:*

ITC Buffer A: 50 mM potassium phosphate pH 7, 50 mM NaCl.

ITC Buffer B: 10 mM Tris.HCl pH 7.0, 150 mM NaCl 50 μM $ZnCl_2$.

*Data analysis:*

Data were analysed using the Origin software (MicroCal) to determine $K_{dim}$ and $\Delta H_{dim}$ using a dimer dissociation model with a standard correction factor for mixing artefacts.

#### 2.3.2.2 Binding Studies

Binding studies were carried out on the interaction between ubiquitin and ZNF216 or Znf_A20, all binding experiments were performed in ITC buffer B.

200-400 μM ubiquitin was added in 55 sequential injections of 5 μl into the ITC cell containing 15-35 μM ZNF216 or Znf_A20 with a 350 second equilibration time between each injection. A reference power of 25 μCal/s was used in all cases.

In order to investigate the effect of binding of ZNF216 to ubiquitin bound to the p62-UBA 300 μM Znf_A20 was injected in 55 sequential injections of 5 μl into

the ITC cell containing 20 μM ubiquitin or 20 μM ubiquitin in the presence of 400 μM p62-UBA with a 350 second equilibration time between each injection. A reference power of 25 μCal/s was used.

The exothermic heat pulse resulting from the interaction was measured and subtracted from appropriate blanks using buffer and only one binding partner.

*Data analysis:*

Data were analysed using the Origin software (MicroCal) to determine the association constant ($K_a$), the enthalpy and entropy change of the interaction and the interaction stoichiometry using a single site model. The free energy of binding (ΔG) was calculated using Equation 2-9.

**Equation 2-9:** $$\Delta G = -RT \ln K = \Delta H - T\Delta S$$

Where R equals the gas constant and T is the absolute temperature.

### 2.3.3 Electrospray Ionisation Mass Spectrometry (ESI-MS)

Electrospray ionisation mass spectrometry (ESI-MS) was used for determining the mass, oligomerisation state, metal ligation, and purity of purified proteins as well as for the study of protein-protein interactions to confirm stoichiometry and measure binding affinities.

ESI-MS was performed on a SYNAPT™ electrospray ionisation, high definition mass spectrometry (HDMS™) system with a Triwave™ ion mobility separation cell and a quadrupole time-of-flight (qTOF) mass analyser (Waters). Samples were injected using a mechanically driven injector and 100 μl syringe (Hamilton) at 5 μl/min. Instrument control and initial data analysis was performed using the Masslynx™ software (Waters).

Total protein concentration was kept in the range of 0.5 to 9 μM in MS Buffer for all experiments. All buffers and samples used for ESI-MS were stored in plastic disposables to prevent sodium contamination from glassware.

After initial mass calculations in MassLynx all results were sorted and analysed using Excel 2002 (Microsoft).

*Solutions:*

MS Buffer: 25 mM Ammonium Acetate pH 6.9 in Milli-Q water.

MS Stock: 250 mM Ammonium Acetate pH 6.9 in Milli-Q water.

#### 2.3.3.1 Native Mass Measurements

In order to measure the mass and purity of a purified protein, 90 µl of desalt fraction was made up to 25 mM Ammonium acetate using MS Stock, and diluted to an appropriate protein concentration with MS buffer.

ESI-MS was recorded on the diluted samples in positive ion mode with a 2.5 kV capillary voltage, a sample cone voltage of 30 V, a 100 L/hour desolvation gas at 323 K, trap and transfer collision energy voltages of 7.0 V and 6.0 V respectively and an intermediate vacuum pressure of 2 mbar.

*Data analysis:*

The apparent molecular mass was calculated from the mass to charge ratios recorded in positive ion mode using the mass lynx software, each mass to charge ratio can be used to calculate the molecular mass using Equation 2-10.

**Equation 2-10** $$MWt = (MZ) - Z$$

Where MWt is the molecules molecular weight, M is the measured mass to charge ratio of the ion and Z equals the charge-state of that ion.

#### 2.3.3.2 Protein Complex Measurements

ESI-MS titrations were performed to confirm binding stoichiometry and measure binding affinities for the interaction of the p62-UBA or p62 341-440 with ubiquitin and also the Znf_A20 with mono ubiquitin or K48 or K63-linked di-ubiquitin.

Titrations with p62-UBA and ubiquitin were recorded with the following concentrations of p62-UBA:ubiquitin (µM); 1:0.5, 1:0.7, 1:1, 1:1.2, 1:1.5, and 1:2. Samples were recorded in positive ion mode with a 2.5 kV capillary voltage, a sample cone voltage of 40 V, with 100 L/hour desolvation gas at 323 K, trap gas was set to 3 ml/min, trap and transfer collision energy voltages of 7.0 V and 4.0 V respectively, and an intermediate vacuum pressure of 4.0 mbar. Total experiment time for each point was 8 minutes. This was repeated on the p62-UBA mutants S399P, E409K G410K, G411S, and G425R.

Titrations with p62 341-440 and ubiquitin were recorded with the following concentrations of p62 341-440:ubiquitin (µM); 1:0.2, 1:0.3, 1.:0.4, 1:0.5, 1:0.6, 1:0.7, 1:0.8, 1:0.9, 1:1, 1:1.2, 1:1.5, and 1:2. Samples were recorded in positive ion mode with a 2.5 kV capillary voltage, a sample cone voltage of 60 V, with 100 L/hour

desolvation gas at 323 K, trap gas was set to 4.5 ml/min, trap and transfer collision energy voltages of 8.0 V for both and an intermediate vacuum pressure of 4.5 mbar. Total experiment time for each point was 8 minutes.

Titrations with Znf_A20 and ubiquitin were recorded with the following concentrations of Znf_A20:ubiquitin (μM); 1:0.5, 1:1, 2:1, 2:1.5, 2:2, 2.5:2, and 3:3. Samples were recorded in positive ion mode with a 2.5 kV capillary voltage, a sample cone voltage of 30 V, a 100 L/hour desolvation gas at 323 K, trap gas flow rate was set to 1.8 ml/min, trap and transfer collision energy voltages of 6 V for both, and an intermediate vacuum pressure of 3.00 mbar. Total experiment time for each point was 4 minutes.

To measure the stoichiometry and confirm the formation of a ternary complex between ubiquitin, Znf_A20, and p62-UBA, the three were mixed at the following concentrations of p62-UBA:ubiquitin:Znf_A20 (μM); of 3:1.5:1.5, 4:2:2, and 5:2:2. Titrations were carried out in positive ion mode with a 2.5 kV capillary voltage, a sample cone voltage of 40 V, a 100 L/hour desolvation gas at 323 K, a trap gas flow rate of 1.8 ml/min, trap and transfer collision energy voltages of 6.0 V for both, and an intermediate vacuum pressure of 2 mbar. Total experiment time for each point was 10 minutes.

Titrations between Znf_A20 and di-Ub linked by either K48 or K63 were also performed. K48 and K63-linked di-ubiquitin were purchased from Enzo life sciences and buffer exchanged into MS Buffer by continuous concentration and dilution using a vivaspin500, (Vivascience) with a 3000 molecular weight cut off PES membrane as per the supplier's instruction. The low concentrations of both di-Ub proteins made quantification using a NanoDrop (Section 2.2.14) impossible. The concentrations were therefore estimated indirectly by recording ESI-MS on mixtures of di-Ub and known concentrations of mono Ub or Znf_A20 and calculating the concentrations based on the intensities of the observed populations in the ESI-MS spectrum.

Proteins were mixed at Znf_A20:di-ubiquitin concentrations of (μM) 1:0.5, 1:1, 2:1, 2:1.5, 2:2 and 3:3. Samples were recorded in positive ion mode with a 2.5 kV capillary voltage, a sample cone voltage of 40 V, a 100 L/hour desolvation gas at 323 K, trap gas flow rate of 1.8 ml/min, trap and transfer collision energy voltages of 6 V for both, and an intermediate vacuum pressure of 2.07 mbar. Total experiment time for each point was 6 minutes.

*Data analysis:*

The apparent concentrations of free protein ([P]), protein-ligand complex ([PL]) and free ligand (Protein 2) ([L]) were calculated from the concentrations of total protein and ligand used, and the signal intensities of each species as done by Veros and Oldham[45]. It was assumed that P, L, and PL gave the same linear detector response. Plots of [PL]/[P] vs. [L] were used to determine $K_d$ according to Equation 2-11 [45]. Solvent exposed surface areas were estimated from the average charge-state of proteins in the native ESI-MS spectrum using the approach of Kaltashov and Mohimen, calculated using Equation 2-12 [46].

**Equation 2-11** $$K_d = ([P][L])/[PL]$$

Where [P] equals the concentration of protein 1, [L] equals the concentration of ligand (or protein 2) and [PL] equals the concentration of their complex.

**Equation 2-12** $$Z_{av} = (\sum ZI)/(\sum I)$$

Where $Z_{av}$ is the average charge-state for the species, Z is the charge-state of a given ion of that species and I is the intensity of the given ion.

Titrations involving di-ubiquitin were treated as two independent identical sites (Sites A and B) with binding constants described by Equation 2-13.

**Equation 2-13**

**Equation 2-13.1** $$K_1 = [PL]/[P][L]$$

**Equation 2-13.2** $$K_2 = [PL_2]/[PL][L]$$

Where, in this case, [P] and [L] equal the concentration of free di-ubiquitin and Znf_A20 respectively, [PL] represents the sum of the concentration of di-ubiquitin with only site A or B occupied (each with their own binding constants described by Equation 2-14), and [$PL_2$] represents a single di-ubiquitin molecule bound to two molecules of Znf_A20 (which is formed by two potential routes with binding constants described by Equation 2-14)[47].

**Equation 2-14**

**Equation 2-14.1** $$K_I = [PL_a]/[P][L]$$

**Equation 2-14.2** $$K_{II} = [PL_b]/[P][L]$$

**Equation 2-14.3** $$K_{III} = [PL_2]/[PL_a][L]$$

**Equation 2-14.4** $K_{IV} = [PL_2]/[PL_b][L]$

Where $[PL_a]$ equals the concentration of single bound complex where site A is occupied and $[PL_b]$ equals the concentration of single bound complex where site B is occupied. $K_1$ and $K_2$ above can therefore be described in terms of the individual equilibrium components from which it is made by Equation 2-15.

**Equation 2-15:**

**Equation 2-15.1:** $K_1 = K_I + K_{II}$

**Equation 2-15.2:** $1/K_2 = (1/K_{III}) + (1/K_{IV})$

From Equation 2-14 and Equation 2-15 it is possible to see that if all sites are of equal affinity and non cooperative, i.e. $K_I = K_{II} = K_{III} = K_{IV}$ then $K_1 = 4 \times K_2$. Therefore a four fold ratio between the two determined constants strongly suggests two equal and independent sites. Where the two sites show negative cooperativity or are not equivalent then $K_1 > 4 \times K_2$, while if the system is positively cooperative then $K_1 < 4 \times K_2$ [47].

### 2.3.3.3 Dilution Experiments

Dilution experiments were performed to assess the suitability of ESI-MS to study dimerisation of the p62-UBA domain and confirm its stoichiometry. Dilution experiments were performed on p62-UBA and its mutants including S399P, E409K G410K, G411S, and G425R. ESI-MS were recorded at concentrations of 0.5, 1.0, 2.0, and 4.0 μM. ESI-MS was recorded in positive ion mode with a 2.5 kV capillary voltage, a sample cone voltage of 30 V, a 100 L/hour desolvation gas at 323 K, trap and transfer collision energy voltages of 7.0 V and 6.0 V respectively, and an intermediate vacuum pressure of 2 mbar. Total experiment time for each point was 6 minutes.

The ability to form heterodimers was assessed by mixing wild type (wt) p62-UBA with the p62-UBA mutants S399P, G411S, and G425R at concentrations of 2:2 μM and ESI-MS recorded as in the dilution experiments.

*Data analysis:*

The apparent concentrations of free monomer and dimer could not be distinguished due to having identical m/z in many cases. Instead the apparent average charge-state of the monomer, calculated by Equation 2-12, was plotted as a function of concentration; a consistent shift in average charge-state is an indicator of

concentration dependent changes in conformation or oligomerisation state populations.

#### 2.3.3.4 Collision Induced Dissociation (CID) - MS/MS

Collision induced dissociation (CID) is the gas phase dissociation of protein-protein complexes using elevated acceleration voltages; this gives rise to high energy collisions with argon gas in the trap or transfer chamber of a Triwave separation cell. This allows confirmation of the nature of a proposed protein complex ion in terms of its composition, and in favourable cases order of contacts in multi-protein complexes[48].

CID was performed on the proposed ternary complex of p62-UBA, ubiquitin and Znf_A20 at concentrations of p62-UBA:ubiquitin:Znf_A20 of 5:2:2 μM. MS/MS was recorded in positive ion mode with a 2.5 kV capillary voltage, a sample cone voltage of 40 V, a 100 L/hour desolvation gas at 323 K, trap collision energy voltage of 6.0 V, a trap gas at 1.8 ml/min, and an intermediate vacuum pressure of 2 mbar. The ternary complex 9+ ion (m/z = 2322) was isolated using the quadrupole and accelerated through the transfer chamber with collision energies ranging from 6 to 22 V. Total experiment time for each point was 10 minutes.

This was repeated on complexes of p62-UBA:ubiquitin at 3:6 μM isolating the complex 7+ ion (m/z = 2043) and on the Znf_A20-Ubiquitin complex at 3:3 μM isolating the complex 8+ ion (m/z = 1894.8).

*Data analysis:*

The relative signal intensities were used to calculate the percentage breakdown of the parent ion into daughter ions of smaller molecular weight. Daughter ion masses calculated from Equation 2-10. This allowed determination of the nature of daughter ions and their populations relative to the parent ion, with increased collision energy.

### 2.3.4 Analytical Ultracentrifugation (AUC)

Analytical Ultracentrifugation was performed and interpreted by Dr David Scott from the school of Biosciences at the University of Nottingham. Data was obtained on an XL-I analytical ultracentrifuge (Beckman) using absorbance optics at 230 nm or using interference optics, utilising a protocol previously defined[49]. Znf_A20 or full length ZNF216 in the concentration range of 0.1- 1 mg/ml were

recorded in 10 mM Tris.HCl pH 7.0, 100 mM NaCl, 50 μM $ZnCl_2$. Sedimentation velocity experiments were carried out at 40 000 rpm using an AnTi60 rotor, at 20 °C, and at 3 loading concentrations to check for self-association. Cells were scanned every 10 minutes.

*Data analysis:*

Raw data was analysed by Dr David Scott. All data were analysed using ULTRASCAN. Molecular weight was determined using the AUC determined sedimentation coefficient (s) and diffusion coefficient (D) using Equation 2-16.

**Equation 2-16:** $$s/D = (M (1 – \upsilon\rho) / (RT)$$

Where M is the molecular weight, υ is the partial specific volume, ρ is the solution density, R equals the gas constant, and T is the absolute temperature.

Where applicable the calculated sedimentation coefficient can be related to the Stoke's radius calculated from SEC (Section 2.2.9) using Equation 2-17[50].

**Equation 2-17:** $$6\pi\eta Rs = (M (1 – \upsilon\rho) / (Ns)$$

Where η is the solvent viscosity and Rs is the Stoke's radius. This can be used to compare the measured Stoke's radii calculated from AUC and SEC or rearranged (in the case of the Znf_A20) to calculate M[50]. The molecular weight determination from AUC for the Znf_A20 resulted in high error due to inaccuracies in measuring s and D for a small protein. The use of the SEC determined Rs in calculating the molecular weight from the AUC measured value of sedimentation coefficient was therefore invaluable[50].

## 2.4 Nuclear Magnetic Resonance Spectroscopy (NMR)

Nuclear Magnetic Resonance (NMR) experiments were recorded on an Avance™ 600 MHz (14.1 T field strength) NMR spectrometer (Bruker) fitted with a TXI triple resonance probe with z-axis gradients (Bruker). Instrument control and initial data analysis was performed using Topspin 2.1 (Bruker).

All spectra were referenced internally in the proton dimension to the methyl peak of 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS)(Fluka). $^{15}N$ and $^{13}C$ spectra were referenced indirectly using DSS and the IUPAC standard chemical shift ratios of 0.101329118 and 0.251449530 for $^{15}N$:$^{1}H$ and $^{13}C$:$^{1}H$ respectively[51].

*Data analysis:*

All 1D spectra were apodized with an exponential window function with line broadening of 2 Hz.

All 2D and 3D data were zero-filled to 2048 in the direct dimension and double their acquired points (up to 1024 points) all other dimensions prior to Fourier transform, optimised with a shifted sine squared apodization in both dimensions, and treated to an automatic polynomial baseline correction.

All NMR spectra were processed in Topspin 2.1 (Bruker) and analysed using CCPNMR (Version 1.1.15) [52].

### 2.4.1 Sample Preparation

Lyophilised constructs of p62, ubiquitin, or Znf_A20* were dissolved in 600 µl of NMR Buffer A, centrifuged at 13, 000 rpm for 1 minute in a bench top centrifuge, and placed in a standard 5 mm 528-PP-7 NMR tube (Wilmad)(*Labelled Znf_A20 was dissolved instead in NMR Buffer B).

ZNF216 and Znf_AN1 were prepared instead by buffer exchanging by desalting and/or spin concentrating into 600 µl (Sections 2.2.10 and 2.2.11) of NMR buffer B, centrifuged at 13, 000 rpm in a bench top centrifuge, placed in a standard 5 mm 528-PP-7 NMR tube (Wilmad), purged with nitrogen gas, and sealed.

All NMR experiments were recorded with the observed protein concentration in the range of 0.65 to 1.2 mM, at 298 K unless otherwise stated. Higher concentrations were used during titrations however these proteins were not being observed at the time.

For $D_2O$ exchange experiments, NMR samples prepared as above were freeze dried and resuspended in 600 µl of 99.9% $D_2O$, centrifuged at 13, 000 rpm for 20 seconds in a bench top centrifuge, and placed in a standard 5 mm 528-PP-7 NMR tube (Wilmad). The time taken between initial exposure to $D_2O$ and data acquisition was timed on a stop watch and was typically between 20-30 minutes.

*Solutions:*

NMR Buffer A: 50 mM potassium phosphate pH7, 50 mM NaCl, 0.04% $NaN_3$ (w/v) 10 % $D_2O$.

NMR Buffer B: 5 mM Tris.HCl 50 mM NaCl, 50 µM ZnCl2, 0.04% $NaN_3$ (w/v) 10 % $D_2O$.

### 2.4.2 RDC Sample preparation

Samples for residual dipolar coupling (RDC) measurements were prepared by soaking samples (prepared as above) into 5% or 7% PAG gels in an RDC gel press (Worldwide glass resources) over night. Protein soaked gels were then loaded into a standard 5 mm 528-PP-7 NMR tube (Wilmad) (with the bottom removed to allow loading) using an RDC gel press (Worldwide glass resources) as per the manufacturer's instructions.

5% and 7% PAG gels were prepared from 30% Protogel® diluted with Milli-Q water and 1% APS (w/v). Polymerisation was initiated by addition of TEMED to a final concentration of 0.1% (v/v) and immediately added to 6 mm plastic tubing. The gel was allowed to set before washing with Milli-Q water to remove un-reacted Protogel®. The gel was then swelled in Milli-Q water for 24 hours before cutting to 21 mm (600 µl /(2.5 $mm^2$ x π)) and air drying for 2 weeks at room temperature.

### 2.4.3 1D and Homonuclear 2D experiments

All one dimensional (1D) experiments were recorded using standard Bruker pulse sequences over 32768 data points with spectral width of 20.027 ppm (12019.23 Hz). Water suppression was achieved using either a WATERGATE pulse sequence[53] a presaturation pulse or using excitation sculpting[54]. 1D experiment performed on isotopically labelled protein were performed with appropriate decoupling using standard Bruker pulse sequences.

All 1D and homonuclear 2D experiments had transmitter frequencies set to water at ~2823 Hz (4.7 ppm).

#### 2.4.3.1 Homonuclear 2D Experiments

$^1H$-$^1H$ nuclear Overhauser effect (NOE) spectroscopy (NOESY), and total correlation spectroscopy (TOCSY) using an MLEV17 sequence[55] were recorded on protein concentrations in the range of 0.9 and 1.2 mM using standard Bruker pulse sequences.

NOESY were recorded on p62 341-440 and p62 221-440 in 99.9% $D_2O$ recorded with 2048 in f2 and 560 points in f1, a spectral width of 10.0138 (6009.615), a NOESY mixing time of 150 ms, and 64 scans. A TOCSY spectrum was recorded with 2048 points in f2 and 512 points in f1, a spectral width of 10.0138

(6009.615 Hz), 48 scans, a TOCSY mixing time of 70 ms, and a presaturation pulse to suppress any residual water peak.

NOESY were recorded on Znf_A20 and its mutants F22Y and V34K, with 2048 points in f2 and 1024 points in f1, a spectral width of 11.479 ppm (6887.052 Hz), 48 scans and WATERGATE water suppression. NOESY were recorded at mixing times of 50, 150 and 250 ms. A TOCSY spectrum was recorded over 2048 points in f2 and 512 points in f1, a spectral width of 11.4759 (6887.052 Hz), 32 scans, a TOCSY mixing time of 70 ms, and WATERGATE water suppression.

NOESY spectra with a mixing time of 150 ms and TOCSY were repeated at 283 K, and at 298 K in 99.9% $D_2O$ recorded as above. However in $D_2O$ excitation sculpting was used to suppress residual water in place of WATERGATE.

*Data analysis:*

Cross-peaks in NOESY spectra were assigned, sorted and classified in CCPNMR (Version 1.1.15) and used to generate distance restraints, according to the $r^{-6}$ relationship between NOE intensity and the average internuclear separation. Distance boundaries were estimated in CCPNMR (Version 1.1.15) using Equation 2-18 based on a default reference distance and intensity within CCPNMR. Observed NOEs were classified as 'strong', 'strong-medium', 'medium-weak', 'weak', or 'very weak' and given maximum internuclear separation boundaries of 2.5, 2.8, 4, 5, and 6 Å respectively. All classes were given a minimum separation of 1.8 Å. Duplicate NOE from different spectra were merged and averaged. Distance restraints were exported as an 'ARIA' distance constraints list using CCPNMR (Version 1.1.15) and used for structure generation (Section 2.5.3).

**Equation 2-18:** $$D_{ij} = \sqrt[6]{\{D_{ref}^6/(I_{ij}/I_{ref})\}}$$

Where $D_{ij}$ represents the average internuclear-distance of protons 'i' and 'j', $D_{ref}$ is the reference internuclear distance, $I_{ij}$ equals the NOE intensity between protons 'I' and 'j' and $I_{ref}$ equals the reference intensity for $D_{ref}$.

### 2.4.3.2 Half filtered NOESY

$^{13}C$ and $^{15}N$-half-filtered NOESY experiments were recorded in order to measure intermolecular NOE between $^{13}C$-$^{15}N$-protein and unlabelled protein. 2D $^{13}C$ and $^{15}N$-half-filtered NOESY experiments were recorded on 1 mM $^{13}C$-$^{15}N$-ubiquitin in the presence of 1.5 mM unlabelled Znf_A20 and on $^{13}C$-$^{15}N$-Znf_A20 in the

presence of 2 mM ubiquitin following the method of Otting *et al* [56]. $^{13}C$ and $^{15}N$-half-filtered NOESY were recorded with 2048 points in f2 and 512 points in f1, spectral widths of 11.97 ppm (7183.9 Hz) in both dimensions, a NOESY mixing time of 250 ms, and 96 scans [56].

### 2.4.4 HSQC/ TROSY 2D Experiments

$^{15}N$ and $^{13}C$- heteronuclear single quantum coherence (HSQC) were recorded on U-$^{15}N$-labelled and U-$^{15}N/^{13}C$ labelled proteins using the 'pep'-HSQC pulse sequence[57]. For the ZNF216 both $^{15}N$-HSQC and $^{15}N$-TROSY were recorded[58].

The sensitivity of the backbone amide proton chemical shift and that of its bound nitrogen, were utilised extensively during this study to investigate structural perturbations (or local environmental changes) caused by mutation, construct extension, protein-protein interactions, and temperature. The $^{15}N$-HSQC was used as the primary tool for their measurement and also used to measure relaxation properties including $^{15}N$ heteronuclear NOE and relaxation enhancement caused by paramagnetic agents.

$^{13}C$-HSQC[57] were recorded on Znf_A20 in the presence and absence of 2 mM ubiquitin with 2048 points in f2 and 256 points in f1 over spectral widths of 11.97 ppm (7183.908 Hz) in the proton dimension and 90 ppm (13581.894 Hz) in the nitrogen dimension, transmitter frequencies of 2823.43 Hz (4.705 ppm) in the proton dimension and 7114.9 Hz (47.149 ppm) in the carbon dimension, and 32 scans.

#### 2.4.4.1 HSQC Titrations

The primary use of the $^{15}N$-HSQC was to monitor protein-protein interactions. Unlabelled protein at concentrations ranging from 0.1 up to 6 mM was titrated into $^{15}N$-labelled protein at 1 mM. The only exception to this was $^{15}N$ p62 341-440 which was performed at a concentration of 0.6 mM.

Ubiquitin was titrated into 1 mM $^{15}N$-p62-UBA and its various mutants including S399P, E409K G410K, G411S and G425R by addition of a concentrated 60 mM stock of ubiquitin to final concentrations of between 0.1-6 mM of ubiquitin. $^{15}N$-HSQC were recorded with 2048 points in f2 and 512 points in f1, spectral widths of 6.4887 ppm (3894.081 Hz) and 31.00 ppm (1885.352 Hz) for proton and nitrogen dimensions respectively, a proton transmitter frequency of 4350.9 Hz (7.25 ppm), a nitrogen transmitter frequency of 7175.66 Hz (118 ppm) and 16 scans.

Ubiquitin was also titrated into 1 mM $^{15}$N-Znf_A20 by addition of a concentrated 60 mM stock of ubiquitin to final concentrations of between 0.1 – 4 mM ubiquitin. $^{15}$N-HSQC were recorded with 2048 points in f2 and 488 points in f1, a spectral width of 5.01 ppm (3004.808 Hz) and 30.00 ppm (1824.531 Hz) for proton and nitrogen dimensions respectively, a proton transmitter frequency of 4981.08 Hz (8.30 ppm), and a nitrogen transmitter frequency of 7054.03 Hz (116 ppm) , and 16 scans. After saturation of Znf_A20, lyophilised-p62-UBA was dissolved into the sample to a concentration of 0.5 mM and a $^{15}$N-HSQC was recorded as above; this was repeated to a final concentration of 4 mM p62-UBA.

Unlabeled p62-UBA and its mutants including S399P, E409K, G411S, and G425R, p62-341-440 and its mutant P392L, and Znf_A20 were titrated (Separately) into 1 mM $^{15}$N-ubiquitin. 1200 µl of 1 mM $^{15}$N-ubiquitin was prepared and split into two 600 µl samples. One was used as the sample and the other used to dissolve lyophilised unlabeled-p62 to a final concentration of 6 mM to generate a concentrated stock. The two samples were mixed in various quantities according to Equation 2-19 to generate a range of concentrations between 0.1 and 6 mM of unlabelled protein. In the case of Znf_A20 the concentrated stock was only required to be at 4 mM.

**Equation 2-19:** $V2 = ((M1-M3)*V1)/(M1-M2)$

Where V2 equals the volume to be removed from one sample to be replaced by the other, M1 is the concentration of the current sample, M3 is the desired concentration, V1 is the volume of the sample, and M2 is the concentration of the concentrated stock.

$^{15}$N-HSQC of ubiquitin were recorded with 2048 points in f2 and 716 points in f1, over a spectral width of 5.98 ppm (3591.954 Hz) and 40.00 ppm (2432.713 Hz) for proton and nitrogen dimensions respectively, proton transmitter frequency of 4801.04 Hz (8 ppm), a nitrogen transmitter frequency of 7175.66 Hz (118 ppm), and 10 scans.

A pre-prepared sample of 1mM $^{15}$N-ubiqutin with 2 mM Znf_A20 was used to dissolve lyophilised p62-UBA to a concentration of 0.5 mM and a $^{15}$N-HSQC was recorded as above; this was repeated to a final concentration of 4 mM p62-UBA.

$^{15}$N-ubiquitin was also mixed with 2 mM of the Znf_A20 mutants F20Y and V32K and compared to wild type (wt) Znf_A20 to examine localised changes at the interface in the ubiquitin $^{15}$N-HSQC.

*Data analysis:*

Changes in the $^{15}$N-HSQC were used to examine changes in protein structure caused by mutants, or to examine binding patches through titrations, by examining the residue specific changes in chemical shift measured as chemical shift perturbations (CSP), calculated by Equation 2-20.

**Equation 2-20:** $$\mathbf{CSP = \sqrt{((\Delta^{1}H)^{2} + (\Delta^{15}N/5)^{2})}}$$

Where $\Delta^{1}$H and $\Delta^{15}$N are the observed changes in proton and nitrogen chemical shift changes in response to a mutation or a binding event.

Titrations of unlabelled proteins into $^{15}$N-ubiquitin were used to calculate an apparent $K_d$ for the interaction using Equation 2-21 in the program IGOR Pro Version 5.0.5.7(WaveMetrics). Residues with large final CSP were selected and the $K_d$ calculated from a global fit to them all simultaneously.

**Equation 2-21:** $$\mathbf{\delta = \Delta\delta \{[L] /(K_d + [L])\}}$$

Where δ is the observed CSP at a given ligand concentration, Δδ is the final CSP at saturation, and [L] equals the ligand (unlabeled protein) concentration.

Titration data for constructs of the p62-UBA were also analysed using a competitive equilibrium model defined by the two equations of Equation 2-22 using the program DynaFit (V 3.28.070, Biokin)[59,60]. Initial fits were poorly defined, however fixing the value of $K_I$ to those determined by ITC allowed determination of $K_{II}$ (Equation 2-22). An example of the model script used for running DynaFit can be found in (Appendix E).

**Equation 2-22:** $$\mathbf{L + L \leftrightarrows LL : K_I} \qquad \mathbf{L + P \leftrightarrows PL : K_{II}}$$

Large cross interface CSP seen in the $^{15}$N-HSQC of ubiquitin caused by the Znf_A20 mutants F20Y and V32K (compared to wt Znf_A20) were used to imply relatively close-contact between the residue showing large CSP in ubiquitin and the mutated residue in Znf_A20. These were used as an inter-residue distance restraint for complex-modelling with a maximum distance boundary of 5.5 Å (Section 2.5.5).

As for the ITC data, the ΔG of binding was determined from the equilibrium constants using Equation 2-9.

#### 2.4.4.2 Dilution experiments

Dilution experiments were recorded on the $^{15}$N-p62-UBA mutants E409K and G410K to measure $K_{dim}$. $^{15}$N-HSQC were recorded in the concentration range of 1 mM to 0.1 mM with varying parameters optimised for time efficiency and signal strength. The number of scans was varied from 16 to 768, with 2048 complex points in f2, the number of points in f1 was varied between 512 and 256 to reduce time in high scan experiments. A spectral width of 5.01 ppm (3004.808 Hz) and 35.00 ppm (2128.628 Hz) for proton and nitrogen dimensions respectively was used, with proton transmitter frequency of 4921.07 Hz (8.2 ppm), and a nitrogen transmitter frequency of 7297.28 Hz (120 ppm).

The concentration at all points was confirmed using a NanoDrop ND-100 spectrophotometer (NanoDrop) (Section 2.2.14).

*Data analysis:*

Residues were selected which had clear well resolved peaks for all observed states at all concentrations and their relative intensities used to estimate the population of all observed species in Excel 2002 (Microsoft). $K_{dim}$ was calculated using the population of each species as a function of concentration using Equation 2-23 in Excel 2002 (Microsoft).

**Equation 2-23** $$\boldsymbol{K_{dim} = [P_M]^2/[P_{Dm}]}$$

Where $[P_M]$ and $[P_{Dm}]$ equal the calculated concentrations of free monomeric protein and dimeric protein respectively. $K_{dim}$ can therefore be determined from the gradient of a linear fit to plots of $[P_M]^2$ against $[P_{Dm}]$.

Dilution experiments were compared to those performed previously for the p62-UBA wt protein performed by our collaborators Dr M. J. Pandya, Dr C. J. Craven and Professor M. P. Williamson[61].

#### 2.4.4.3 Amide Temperature Dependence

The $^{15}$N-HSQC was used to monitor the temperature dependent changes in the chemical shift of the backbone amide-proton of the p62-UBA and its G425R mutant. Amide-proton temperature-coefficients were measured in order to study

hydrogen bonding and potential low energy excited states as done by Williamson *et al*[62].

The NMR temperature was calibrated using the standard protocol with a standard methanol test sample (Bruker) over a wide range of temperatures from 278 K to 310 K. An ethylene glycol sample was used to calibrate temperatures from 290 K to 323 K. Though both samples gave near identical correction factors there were some discrepancies. As a consequence, for the purpose of investigating potential low energy excited states only data using the methanol calibration were used.

$^{15}$N-HSQC were recorded on 1 mM protein with 200 µM DSS, in the temperature range 278 K to 323 K in steps of 2 or 3 K. $^{15}$N-HSQC were recorded with 16 scans with 2048 points in f2 and 576 points in f1, over spectral widths of 6.4487 ppm (3894.08 Hz) and 36 ppm (2189.443 Hz) in the proton and nitrogen dimensions respectively, transmitter frequencies of 4350.94 Hz (7.25 ppm) in the proton dimension and 7206.06 Hz (118.5 ppm) in the nitrogen, and a relaxation delay of 1 second. All spectra were referenced internally to DSS.

*Data analysis:*

Amide proton chemical shifts were plotted as a function of temperature and fitted to a linear correlation using IGOR Pro Version 5.0.5.7(WaveMetrics) to extract amide temperature coefficients in ppb/K. Deviation from linearity in the temperature correlation was determined by close examination of the residuals from the linear plot.

Temperature coefficients were plotted against chemical shift deviation (CSD) defined by Equation 2-24 to generate 'Andersen' plots[63].

**Equation 2-24** $$\mathbf{CSD = \delta H^N - \delta H^N_{rc}}$$

where $\delta H^N$ represents the observed amide proton chemical shift and $\delta H^N_{rc}$ denotes the random coil chemical shift for a given residue type. The random coil values from the Andersen *et al* work were used so that a proper comparison could be made with their data[63]. In accordance with Andersen *et al*, the $\delta H^N$ was used from the lowest temperature recorded, in this case 278 K[63].

Hydrogen bonded amides were identified from the p62-UBA domain dimer NMR structure (RPDB 2KNZ) using standard geometrical criteria available in MOLMOL[64], a maximum acceptor distance of 2.5Å from the amide proton and angle smaller than 40º between backbone nitrogen, amide proton and the acceptor atom were used[65].

#### 2.4.4.4 $^{15}$N Heteronuclear NOE

The $^{15}$N heteronuclear NOE ($^{15}$N($^{1}$H) NOE) was measured using standard $^{15}$N-HSQC based pulse sequences with and without proton saturation during the relaxation delay prior to the starting 90°-$^{15}$N-pulse in an interleaved manner[66]. $^{15}$N($^{1}$H) NOE experiments were recorded on a 1 mM sample of $^{15}$N Znf_A20 in the presence and absence of 4 mM ubiquitin or 4 mM ubiquitin with 4 mM p62-UBA. $^{15}$N($^{1}$H) NOE experiments were recorded with 2048 points in f2 and 488 points in f1, over spectral widths of 11.9705 ppm (7183.97 Hz) and 30 ppm (1824.531 Hz) in the proton and nitrogen dimensions respectively, transmitter frequencies of 2822.8 Hz (4.704 ppm) in the proton dimension and 7054.03 Hz (116 ppm) in the nitrogen, and a relaxation delay of 5 seconds.

*Data analysis:*

Interleaved $^{15}$N-HSQC were split into $^{15}$N-HSQC with and without proton saturation using Topspin 2.1 (Bruker). The $^{15}$N($^{1}$H) NOE was calculated using Excel 2002 (Microsoft) according to Equation 2-25.

**Equation 2-25:** $$^{15}\mathrm{N}(^{1}\mathrm{H})\ \mathrm{NOE} = (I_{sat}/I_{eqm})$$

Where $I_{sat}$ and $I_{eqm}$ represent the intensity of a given peak in a $^{15}$N($^{1}$H) NOE experiment with and without proton saturation respectively.

#### 2.4.4.5 Paramagnetic Relaxation Enhancement (PRE)

Paramagnetic relaxation enhancement (PRE) was performed using the soluble paramagnetic agent 4-hydroxy-2,2,6,6-tetramethylpiperidinyl-1-oxy (HyTEMPO) (Sigma), or MTSL covalently coupled to ubiquitin via the K48C mutant (Section 2.2.13).

PRE experiments using HyTEMPO were performed on 1 mM $^{15}$N-p62-UBA sample in the presence and absence of 4 mM ubiquitin: HyTEMPO was dissolved as a concentrated stock on NMR Buffer A to a final concentration of (500 mM) and added to the $^{15}$N-p62-UBA sample in a stepwise manner to final concentrations of 5, 10, 15 and 30 mM. At each concentration $^{15}$N-HSQC were recorded identical to those for previous titration studies (Section 2.4.4.1) with an extended relaxation delay of 5 seconds.

PRE experiments using MTSL-K48C-ubiquitin were performed on both $^{15}$N-p62-UBA and $^{15}$N-Znf_A20. $^{15}$N-p62-UBA and $^{15}$N-Znf_A20 were saturated with

MTSL-K48C-ubiquitin (4 and 2 mM respectively) and $^{15}$N-HSQC recorded identical to those for previous titration studies (Section 2.4.4.1) with an extended relaxation delay of 5 seconds: $^{15}$N-HSQC were recorded before and after reduction of the ubiquitin coupled paramagnetic MTSL group using ascorbic acid (Section 2.2.13).

*Data analysis:*

Residue specific PRE values were calculated in Excel 2002 (Microsoft) according to Equation 2-26, plotted as a function of residue number, and examined by plotting onto the surface of the protein.

**Equation 2-26:** $$\mathbf{PRE = (I_{para} / I_{dia})}$$

Where $I_{para}$ and $I_{dia}$ equal the intensity of the $^{15}$N-HSQC peak in the presence of paramagnetic agent and without paramagnetic agent (or with reduced/ diamagnetic agent) respectively.

PRE data were converted into approximate distance restraints for initial complex modelling similar to previous studies[67,68]: PRE values of < 0.1 were given a range of 2–15 Å, between 0.15 and 0.4 a range of 10 to 19.5 Å , 0.4 to 0.6 a range of 15 – 22 Å, and 0.6 -0.8 were given a range of 15 – 25 Å: values over 0.8 were not used.

The K48C ubiquitin mutation also resulted in cross interface CSPs seen in the $^{15}$N-HSQC of Znf_A20 when compared to wt-ubiquitin. CSPs were used to generate inter-residue distance restraint for complex modelling with a maximum distance boundary of 5.5 Å (Section 2.5.5).

#### 2.4.4.6 Residual Dipolar Coupling (RDC)

Backbone $^{1}D_{NH}$ residual dipolar couplings (RDC) were obtained from the difference in $^{1}J_{NH}$ scalar couplings between an isotropic (solution) environment and with partial alignment of the protein using 5 % or 7 % PAG (Section 2.4.1). RDCs were measured using standard interleaved $^{1}$H-$^{15}$N- IPAP-HSQC spectra [69]. RDC were measured for 1 mM $^{15}$N-Znf_A20 in the presence and absence of 2 mM ubiquitin and on 1 mM $^{15}$N-labelled ubiquitin in the presence of 2 mM Znf_A20 [69]. All $^{1}$H-$^{15}$N- IPAP-HSQC spectra were recorded with 2048 points in f2 and 976 points in f1, over a spectral width of 11.97 ppm (7183.908) Hz) in the proton dimension, a proton transmitter frequency of 2822.8 Hz (4.704 ppm), nitrogen

spectral widths and transmitter frequencies corresponding to those of the relevant $^{15}$N-HSQC (Section 2.4.4.1), and 16 scans.

*Data analysis:*

$^{1}$H-$^{15}$N-IPAP-HSQC spectra[69] were split into their component spectra using Topspin 2.1 (Bruker). RDC values were calculated from the difference between the calculated $^{1}J_{NH}$ coupling in the isotropic and aligned spectra using Excel 2002 (Microsoft). Values in ppm were converted into Hz by multiplication by the $^{15}$N-larmor-frequency at 14.1 T of 60.796 Hz.

To avoid complications during structure refinement caused by motional averaged RDC values, only RDC values from residues showing high values of $^{15}$N($^{1}$H) NOE (>0.6) were used in structure generation. RDC values were used as raw RDC values without prior knowledge of the orientational axis using the implicit saupe tensor alignment constraint (ISAC) method[70].

In the case of HADDOCK, docking RDC restraints were used as intervector projection angles[71] as per the instructions in HADDOCK[72,73].

## 2.4.5 Heteronuclear 3D Experiments for Assignment

All 3D experiments had their proton transmitter frequencies set to water at 2823 Hz (4.7 ppm).

### 2.4.5.1 3D NOESY Spectra

$^{15}$N-HSQC-NOESY were used to assign the HSQC of various mutants of p62-UBA in the free and bound state by comparison to previously recorded wild type data. To generate distance restraints for the free and bound state of Znf_A20, and to confirm the assignment ubiquitin bound to Znf_A20.

$^{15}$N-HSQC-NOESY were acquired using standard Bruker pulse sequences[74,75] on $^{15}$N-p62-UBA mutants including S399P, E409K, G410K, G411S, and G425R both free and in the presence of 4 mM ubiquitin, p62-341-440, on $^{15}$N-Znf_A20 in the presence and absence of 2 mM ubiquitin, on $^{15}$N-ubiquitin in the presence of 2 mM Znf_A20 and on $^{15}$N-Znf_AN1. All $^{15}$N samples were at a concentration of around 1 mM.

In the case of ZNF216 a $^{15}$N –TROSY-NOESY was recorded instead using a standard Bruker pulse sequence.

All $^{15}$N-HSQC NOESY and $^{15}$N –TROSY-NOESY experiments were recorded with 32 scans, 2048 points in f3, 32 points in f2 and 360 points in f1, a spectral width of 11.97 ppm (7183.908 Hz) in both proton dimensions, spectral width and transmitter frequency in the nitrogen dimension corresponding to those used in the relevant $^{15}$N-HSQC for a given protein (Section 2.4.4), and a NOESY mixing time of 150 ms.

In the case of Znf_A20, Znf_AN1, ZNF216 and p62-341-440, $^{15}$N-HSQC-NOESY spectra were recorded alongside a $^{15}$N-HSQC-TOCSY (mlev), recorded using standard bruker pulse sequences with identical parameters to the NOESY with a TOCSY mixing time of 70 ms.

A $^{13}$C-HSQC-NOESY[74,75] was recorded on $^{15}$N-$^{13}$C-Znf_A20 in order to generate distance restraints for structure generation purposes. $^{13}$C-HSQC-NOESY were recorded with 2048 points in f3, 32 points in f2, 128 points in f1, spectral widths of 14.9847 ppm (8992.806 Hz) in both proton dimensions and 80 ppm (12072.795 Hz) in the carbon dimension, and transmitter frequencies of 2823.43 Hz (4.705 ppm) in the proton dimensions and 7132.8 Hz (47.268 ppm) in the carbon dimension.

*Data analysis:*

3D NOESY data were combined with NOESY data from homonuclear 2D NOESY data and processed in an identical manner (Section2.4.3.1).

#### 2.4.5.2 Backbone assignment data

Standard backbone assignment experiments for the assignment of protein $N^H$, $H^N$, C', Cα and Cβ chemical shifts were performed on a 1 mM $^{15}$N-$^{13}$C-Znf_A20 sample using standard Bruker pulse sequences. These experiments included HNCO[76-78], HN(CA)CO[79], CACBNH[80], CACB(CO)NH[80] and a HCCH-TOCSY[81].

HNCO[76-78], HN(CA)CO[79], were recorded with 48 and 64 scans respectively, with 2048 points in f3, 32 points in f2, 64 points in f1, a spectral width of 11.9705 ppm (7183.908 Hz) in the proton dimension, 30 ppm (1824.53) in the nitrogen dimension and 25 ppm (3773.234 Hz) in the carbon dimension, with transmitter frequencies of 21821.63 Hz (4.7 ppm), 7054.03 Hz (116 ppm) and 26558.45 Hz (176 ppm) respectively.

CACBNH[80] and CACB(CO)NH[80] were recorded 80 and 48 scans respectively, with 2048 points in f3, 32 points in f2, 96 points in f1, a spectral widths

of 11.97 ppm (7183.9 Hz) in the proton dimension, 30 ppm (1824.531 Hz) in the nitrogen dimension and 74.96 ppm (11312.212 Hz) in the carbon dimension, transmitter frequencies of 2823.43 Hz (4.705 ppm), 7054.03 Hz (116 ppm) and 5884.76 Hz (38.99 ppm) respectively.

A HCCH-TOCSY[81] was recorded with 16 scans, 2048 points in f3, 64 points in f2 and 128 points in f1, over a spectral width off 11.97 ppm (7183.908 Hz) for the proton dimensions and 74.9 ppm (11312.218 Hz) for the carbon dimension, transmitter frequencies of 2823.43 Hz (4.705 ppm) and 5884.74 Hz (38.99 ppm) respectively, and a TOCSY mixing time of 24 ms.

The HNCO and HN(CO)CACB experiments were repeated on the Znf_A20 in the presence of 2 mM ubiquitin.

*Data analysis:*

The above experiments were used to assign the complete backbone resonances for the Znf_A20 protein using the standard NMR protein assignment methodologies[82].

$N^H$, $H^N$, C', Cα, and Cβ chemical shifts were exported from CCPNMR (Version 1.1.15) using the CCPNMR FormatConverter (Version 1.0.19) in the standard biological magnetic resonance data bank (BMRB) and 'TALOS' format. Chemical shifts were used to predict protein dihedral angles including φ, Ψ, ω and $\chi^1$ angles using the program PREDITOR[83]. PREDITOR also calculates the chemical shift predicted random coil index (RCI) as a measurement of protein disorder.

Φ and Ψ angles were also predicted from $N^H$, $H^N$, C', Cα, and Cβ chemical shifts using the program TALOS[84] and compared to those calculated from PREDITOR. Where large disagreement occurred those values were considered potentially unreliable and discarded.

## 2.5 Computational Bioinformatics and Structure Generation

### 2.5.1 BioInformatics

In order to identify highly conserved features of Znf_A20 domains sequence alignment of Znf_A20 domains from other proteins and species were examined. Potential Znf_A20 domains were selected from confirmed-proteins containing known Znf_A20 domains using a search for the term 'A20' in UniProt[85]. Sequence

alignment was performed on the full length proteins and on the Znf_A20 sequences alone in ClustalX 2.0.11[85,86]. Many hits also contained Znf_AN1 domains, the identified Znf_AN1 domains were also aligned in Clustal X[86].

Similarly Uniprot was used to identify suitable sequences for ubiquitin, and UBA domains for sequence alignment in Clustal X[86].

Sequence alignments were used to propose domain consensus sequences and displayed as sequence logos[87] using the WebLogo software (version 2.8.2, http:weblogo.Berkley.edu)[88]. The output colour scheme was modified to match that used by Clustal X[86].

The sequence for the p62 and ZNF216 protein was subjected to analysis using SMART[89], PROSCAN[90], and ELM[91] to search for potential domains or functional motifs, and using the DisProt server (www.disprot.org) to identify intrinsically disordered regions using the programs DisProt[92], IUPred[93], DISpro[94] and DISOPRED2[95] .

The program APSSP2 was used for secondary structure prediction using the APSSP2 web server (www.imtech.res.in/raghava/apssp2/).

### 2.5.2 Homology Modelling

Homology models of yeast ubiquitin were generated using SWISS-MODEL[96,97]. The free ubiquitin solution structure (RPDB ID 1D3Z) and ubiquitin bound to the Znf_A20 of Rabex-5 (RPDB ID 2FIF chain E) were used as templates using an alignment of human and yeast ubiquitin generated in Clustal X[86].

A second free homology model of yeast ubiquitin was generated with chemical shift refinement using yeast ubiquitin chemical shifts provided within the Searle group (Un published data), without a pre selected template using the program CS23D[98].

All homology models were subjected to molecular minimisation performed by Dr Huw Williams. Molecular minimisations were performed using the SANDER modules of AMBER. Starting structures were assembled in the xLEaP program including the introduction of all missing protons. Explicit solvent was added (TIP3P Water) using a truncated-octahedron geometry and padding to 9Å. Minimisation was achieved using 50 steps of steepest descent minimisation followed by conjugate gradient minimisation until the structure converged or 5000 steps were completed.

#### 2.5.2.1 Modelling Potential Ternary Complex

Structures of ubiquitin binding domains (UBD) in complex with ubiquitin were selected from the RPDB providing at least one representative from each of the 20 known UBD families[2,16,17]. All selected complex were aligned by performing an RMSD fit to the backbone atoms of ubiquitin residues Gln2- Arg72 using the program MOLMOL[64]. Complexes were analysed on the likelihood of the bound UBDs to be able to occupy ubiquitin simultaneously based on steric clashes between the two UBD in the aligned complex. Steric clashes were determined using the CalcVdw function in MOLMOL to calculate van der waals violations. The ability to form ternary complex was rated from -3 (highly unlikely) to 3 (Highly likely) on this basis.

The following UBD families (and RPDB IDs) were analysed; Znf_A20 (2C7N), MIU (2C7N), UIM (1Q0W), UBA (1WR1/2DEN), CUE (1OTR), NZF (1Q5W), UEV (1S1Q), PRU (2Z59), Znf_UBP/PAZ (2G45), UBCH8 (2KJH), GLUE (2DX5), PFUC (2K8B), GAT (1YD8), UBM (2KHW), SH3 (2JT4), DUIM (2D3G), UBCH5 (2FUH), VHS (3LDZ) and MPN (2ZNV). No RPDB was available for the UBZ domain but the alignment provided by Bomar *et al* [99] was used as a guide for analysis.

### 2.5.3 NMR Based Structure generation

Inter proton distance restraints generated from NMR data recorded on the Znf_A20 protein which had been previously categorised according to distance boundaries (Section 2.4.3.1) were further categorised according to the proton type for use in structure generation. These include, 'Average' for protons which represent the sum of multiple protons observed as a single resonance due to degeneracy including methyl groups and aromatic ring protons, 'HN to HN' including $H^N$ to $H^N$ and $H^N$ to Hα NOEs, 'intra residue' representing NOE between protons of the same residue number except $H^N$ to Hα, 'sequential' including NOE between protons in residue 'i' and either 'i'+1 or 'i'-1 that do not fall into the 'HN to HN' or 'average' categories, 'short' representing NOE between protons from residues which are closer than 4 residues apart in the primary sequence and not belong in any previous category, and 'long' representing any remaining NOEs.

Dihedral restraints calculated using PREDITOR[83] and not in large disagreement with those calculated in TALOS[84] (Section 2.4.5.2) were used

alongside NOE derived distance restraints and RDC restraints (Section 2.4.4.6). For early modelling $\chi^1$ angle restrains were omitted. After initial rounds of modelling $\chi^1$ angles were incorporated only for those residues which are likely to be buried or have restricted rotation around the Cα-Cβ bond in accordance with the findings of Wishart *et al* [83].

The structure of the Znf_A20 in its free and ubiquitin bound state were determined using a simulated annealing protocol implemented in XPLOR-NIH (version 2.1.6). This consisted of a three step protocol with the first step being a simulated annealing protocol derived from that proposed by Nilges *et al*[100,101] which used NOE and dihedral restraints, followed by a refinement step with the addition of RDC with NOE and dihedral restraints, followed by a final minimisation step[100,101].

Biophysical and bioinformatics data had shown the Znf_A20 domain to require zinc for correct folding. The Znf_A20 domain contains only 4 Cys residues and these were chosen as the 4 residues to coordinate the zinc atom. The 4 coordination zinc centre was included in all stages as covalent bonds between the Cys SG atoms of residues 14, 18, 30 and 33 and a $Zn^{2+}$ atom with a tetrahedral geometry around the $Zn^{2+}$, and S- $Zn^{2+}$ bond lengths of 2.3 Å. The Cys Hγ proton was removed.

A total of 722 NOE distance restraints, 177 dihedral restraints and 30 backbone $^1D_{NH}$ RDC parameters were used for the free Znf_A20: For the ubiquitin bound form of Znf_A20 a total of 345 NOE distance restraints, 161 dihedral restraints and 30 backbone $^1D_{NH}$ RDC parameters were used.

For structure calculation the PARALLHDG[101] force field was used with additional defined parameters for the $Zn^{2+}$ atom via the param19.ion parameter file and the toph19.ion topology file.

The simulated annealing protocol was carried out on an initial randomised, extended starting structure with no conformational restriction other than the zinc coordination sphere. The protocol consisted of an initial phase of high temperature Cartesian dynamics at 1000 K with a time step of 0.005 ps over 20,000 steps using the Verlet integrator.

The second refinement step included a slow cooling step from 1000 K to 100 K in steps of 50 K, with a time step of 0.005 ps over 40, 000 steps in which the relative weightings of all non-bonded energy terms were increased from 10 % of their default value up to their force field defaults. The NOE and dihedral restraints

were given a constant scale of 50 and 200 respectively. RDC restraints were incorporated without prior knowledge of the orientational axis using the using the ISAC[70] method by recalling the TENSO statement which calculates the rhombicity and amplitude from the structures during the run. RDC restraints were increased throughout the refinement stage from 1 % of its default up to its force field default. These refined structures were then refined with another 10,000 cooling steps.

An initial 200 structures were generated, structures with no NOE violations over 0.3 Å or dihedral angle violations of more than ± 5° were deemed acceptable. Of the accepted structures the 20 lowest energy structures were selected to represent the ensemble. The 20 accepted structures were analysed for fit to the RDC restraints using the program PALES[102] and found to be in excellent agreement.

In the case of the bound Znf_A20 structure the randomised initial structure was replaced with the lowest energy structures for the free Znf_A20, only 100 initial structures were generated and 10 final structures selected.

Final structures were visualised using the program MOLMOL[64]. MOLMOL was also used for initial inspection including Ramachandran plots, van der Waals clashes, and calculation of per residue RMSD. All structure figures were generated using MOLMOL[64].

### 2.5.4 Analysis and Validation

Analysis of the structural ensemble was performed using the RSC-protein data bank (RPDB), protein structure validation software suite (PSVS)[103]. The PSVS integrates analyses from several widely-used structure quality evaluation tools, including RPF[104], PROCHECK[105], MolProbity[106], Verify3D[107], Prosa II[108], the RPDB validation software, and various other structure-validation tools. PSVS provides standard constraint analyses, statistics on the RPDB validation goodness of fit between structures and experimental data, and knowledge-based structure quality scores in standardized format suitable for database integration. Global quality measures are reported as Z scores, based on calibration with a set of high-resolution X-ray crystal structures.

NMR chemical shifts were predicted for all protons, main chain carbon atoms and Cβ atoms from the solved structure of the Znf_A20 domain using the program SHIFTX[109] and compared to the recorded chemical shifts.

The 20 lowest energy structures and the NMR restraints used to derive them were deposited in the RPDB and BMRB databases using the ADIT-NMR tool (http://deposit.bmrb.wisc.edu/bmrb-adit/). Deposited data included the assigned chemical shifts, NOE distance constraints, dihedral restraints and RDC restraints. Coordinates and structure details for the Znf_A20 was deposited under the accession code 2KZY at the RPDB and at the BMRB under the accession number 17023.

### 2.5.5 NMR Based protein-protein complex formation

High ambiguity driven biomolecular docking (HADDOCK)[72] was performed to generate a model of the complex between Znf_A20 and ubiquitin using the program HADDOCK 2.0: A combination of CSP data from NMR titrations (Section 2.4.4.1), RDC data for the two proteins bound to its partner (Section 2.4.4.6), intermolecular-NOE recorded from half-filtered-NOE (Section2.4.3.2), PRE data using $^{15}$N-Znf_A20 with MTSL-K48C-ubiquitin (Section 2.4.4.5), and mutation based CSP data using the Znf_A20 mutants F22Y and V34K with $^{15}$N-ubiquitin (Section 2.4.4.1) were used to generate a family of structures.

CSPs were used to generate ambiguous interaction restraints (AIR) between solvent exposed residues with high CSPs in the two proteins to be docked as described by the HADDOCK 2.0 protocol[72].

To avoid the direct use of alignment tensors in docking, RDC were used as intervector projection angle (VEAN) restraints[71]; RDC restraints for both ubiquitin and Znf_A20 were analysed using the program 'prediction of alignment from structure' (PALES)[110] to extract the tensor parameters Axx, Ayy, and Azz which were used to generate VEAN restraints[71]. VEAN restraints were split into intramolecular and intermolecular VEAN restraints for use in docking[71]. The VEAN restraints generated from the tensor-parameters calculated in PALES based on ubiquitin or on Znf_A20 were significantly different; as a result the HADDOCK protocol was performed twice with each set of starting structures, once with each VEAN restraint set. The VEAN data set which resulted in the least violations was selected for the final runs.

Several starting structures were tried for docking including the free and bound solution structures of Znf_A20 described above (Section 2.5.3), the NMR structure of human ubiquitin (1D3Z), the crystal structure of human ubiquitin bound to the Znf_A20 domain of Rabex-5 (2FIF chain E)[20,21], and yeast ubiquitin homology

models made from both using SWISS-MODEL[96,97]. All gave similar results with the lowest number of violations arising from ubiquitin structures based on the Rabex-5-Znf_A20-bound ubiquitin structures.

HADDOCK docking was performed using the standard HADDOCK protocol described by Dominguez *et al*[72]. Step one involves initial rigid body docking where the two proteins are positioned 150 Å apart and orientated randomly relative to one another followed by energy minimisation. The second stage consists of three simulated annealing refinements. The first simulated annealing involves 1000 steps from 2000 to 50 K with 8 fs time steps, with the two proteins are considered as rigid bodies. The second simulated annealing involves 4000 steps from 2000 to 50 K with 4 fs time steps where the side chains at the interface are allowed to move. The third simulated annealing involves 1000 steps from 500 to 50 K with 2 fs time steps, with both side chains and backbone atoms at the interface allowed to move.[72]. Standard protocol was followed utilising the PARALLHDG 5.3[101] force field with electrostatics switched off.

A total of 13 AIR restraints, 4 NOE derived distance restraints, 29 PRE derived distance restraints, 4 mutation CSP derived distance restraints, and 284 intermolecular and 390 intramolecular VEAN restraints generated from a total of 92 RDC restraints. Distance restraints and intermolecular VEAN restraints were used throughout the protocol while the intramolecular VEAN restraints were excluded from the initial rigid body docking and used during the remainder of the protocol.

Initial docking and NMR CSP was consistent with the formation of a hydrogen bond between Znf_A20-Ser 33 and ubiquitin-Asp 58, as such the docking was repeated with this incorporated.

A total of 500 structures were generated from rigid body docking with 50 selected for semi- and fully-flexible docking. Selection of structures from the rigid body docking was based primarily on the distance violation energies.

The 10 lowest energy (HADDOCK energies) structures and the NMR restraints used to derive them were deposited in the RPDB and BMRB databases using the ADIT-NMR tool (http://deposit.bmrb.wisc.edu/bmrb-adit/). Coordinates and structure details for the Znf_A20 in complex with mono-Ub were deposited under the accession code 2L00 at the RPDB and at the BMRB under the accession number 17024.

# Chapter 3

# 3 Solution Structure of the Znf_A20 Domain of ZNF216

## 3.1 Introduction

### 3.1.1 Muscle Atrophy and the UPS

The pathways controlling protein synthesis and protein breakdown have an important role in the control of cell size[111]. This is particularly true in muscle which continually adapts to changing environment and energy requirements, and where control of cell size is predominantly controlled by rates of protein synthesis and degradation[111,112].

Hypertrophy or cell growth is largely controlled by activation of the Akt pathways induced by IGF-1 and insulin to stimulate protein synthesis and suppress protein degradation [111](Figure 3-1). A decrease in cell size and therefore muscle size is termed muscle atrophy and is mainly caused by the loss of organelles, cytoplasm and proteins[111]. The increased breakdown of proteins in skeletal muscle during muscle atrophy is characteristic of several clinical conditions including hyperthyroidism, sepsis, cancer cachexia, and AIDS[112,113] and may be triggered by TNF-α (Figure 3-1).

The Ub proteasome system (UPS) is the major mechanism for controlled protein degradation in eukaryotes and is key to the control of the majority of cellular processes. The UPS includes sequential multi-step reactions including Ub-conjugation to target proteins by E1, E2, and E3 enzymes, followed by recognition by shuttling proteins and transportation to, or direct recognition by, the 26S proteasome resulting in proteolysis of the target protein by the proteasome[9,11,16].

Genes that are up regulated during, and directly mediate, muscle atrophy are termed ‘atrogenes’ and include several components of the UPS. This had led many to conclude that the UPS is central to the degradation of myofibril proteins during muscle atrophy [112].

The two muscle-specific E3 Ub ligases MAFbx/Atrogin-1 and MuRF-1 are both up regulated in atrophying muscle and the genetic knock-out of these genes results in a resistance to muscle atrophy induced by denervation[112-114]. Transcription of MAFbx/Atrogin-1 and MuRF-1 are both under the control of members of the foxo transcription factor family which are also classed as atrogenes[112]. The phosphorylation of foxo family members by Akt results in their retention in the cytoplasm and this appears to be key to down regulation of muscle protein degradation during hypertrophy (Figure 3-1)[111].

An overview of the key signalling events which control hypertrophy and atrophy is shown in Figure 3-1 taken from the 2008 review by Sandri[111].

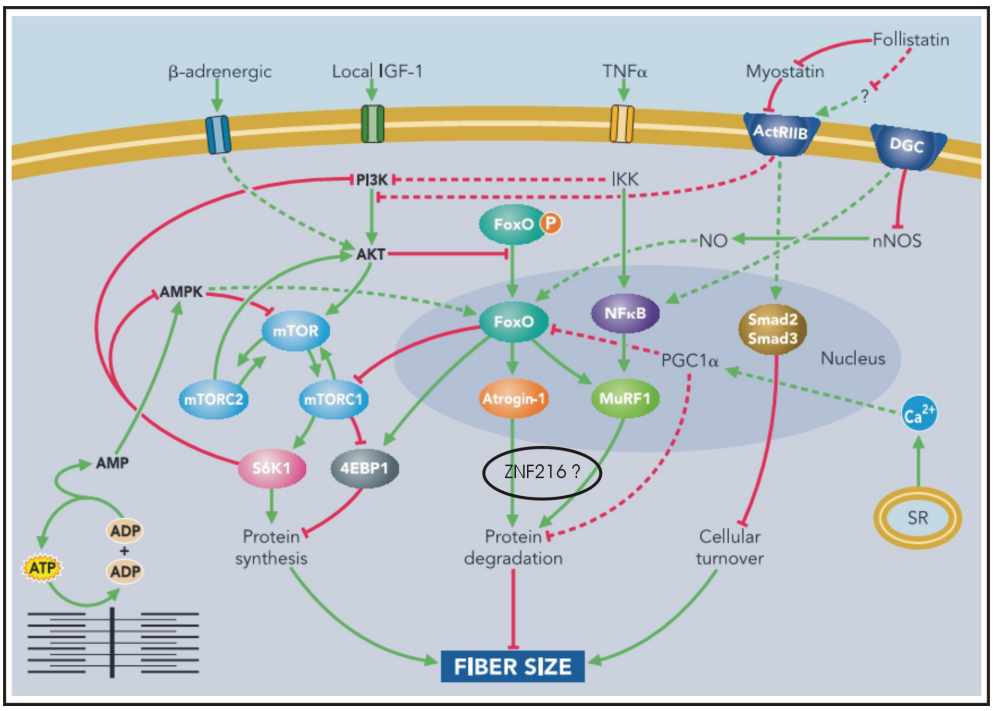



**Figure 3-1: Signalling pathway associated with control of muscle fibre size taken from Sandri 2008[111], the potential site for ZNF216 function has been added.**

### 3.1.2 ZNF216 and Muscle Atrophy

The ZNF216 protein is a 213 amino acid protein which was first identified in 1998 as being linked with autosomal recessive non-syndromic hearing loss (ARNSHL) due to its expression in the cochlear and its location on the DFNB7/11 locus linked to ARNSHL, although no disease links were subsequently found[115]. Later studies have reported high expression of ZNF216 in skeletal muscle (and to a lesser extent in bone) with dramatically weaker expression in other tested tissues[113].

The ZNF216 protein has been linked to a variety of functions including a potential modulator of RANKL-induced osteoclast differentiation and an inhibitor of NF-κB in skeletal muscle. Interactions between ZNF216 and proteins including IKKγ, RIP and TRAF6, have been reported as well as homo-oligomerisation, however its precise function is yet to be fully elucidated[116,117].

Recent studies have shown that, like MAFbx/Atrogin-1 and MuRF-1, the ZNF216 protein is up regulated in models of muscle atrophy including dexamethasone treatment of C2C12 myotubes, denervation and fasting. Like MAFbx/Atrogin-1 and MuRF-1, when ZNF216 is knocked out in mice, the muscle shows a resistance to denervation induced muscle atrophy. In addition ZNF216 also appears to be under the control of foxo and induced by cytokines such as TNF-α.

Although the precise cellular role of ZNF216 remains unclear, work by Hishiya *et al* has provided convincing evidence that the role of ZNF216 in muscle atrophy is as part of UPS mediated protein degradation[113]. Using yeast two hybrid, co-immunoprecipitation, and GST-pull down experiments Hishiya *et al* showed that the ZNF216 protein can bind non-covalently with Ub and ubiquitinated proteins using its N-terminal A20 type zinc finger and that the ZNF216 protein can also interact non-covalently with the 26S proteasome in a Znf_A20 and Ub independent manner[113]. Interestingly the attenuation of denervation induced muscle atrophy in ZNF216 null mice is coupled with the accumulation of Ubiquitinated proteins in the cytoplasm suggesting an impaired UPS. Furthermore cellular association of ZNF216 with the proteasome was significantly increased in models of muscle atrophy. The apparent reduction of UPS function in the absence of ZNF216, the increase in proteasomal association of ZNF216 induced by atrophy, and the interaction of ZNF216 with both Ub and the proteasome, provides compelling evidence for its role as a muscle specific proteasomal shuttling protein, acting downstream of MAFbx/Atrogin-1 and MuRF-1, facilitating the transport of ubiquitinated proteins to the proteasome for degradation in a similar fashion to proteins such as RAD23 and p62 in other tissues(Figure 3-1).

The prospect of ZNF216 functioning as a proteasomal shuttling protein provides further support for the notion that the UPS is central to the mechanism of muscle atrophy, with yet another atrogene being identified as a key component of the UPS.

### 3.1.3 The Znf_A20 Domain

Zinc fingers are among the most abundant proteins in eukaryotic genomes; they represent an interesting and diverse class of protein motifs exemplified by them being designated 'molecule of the month' by the RPDB in March 2007[118,119]. The coordination of a zinc atom by (predominantly) cysteine or histidine side chains gives rise to the formation of highly stable structures for relatively small domains. Many classes of zinc finger exist with the majority being domains of 30-40 amino acids in size[118,119]. The first zinc fingers were identified almost 25 years ago as DNA binding elements, it is now known that zinc fingers are hugely diverse in architecture and function, with many single, double, and multiple zinc binding motifs, performing functions including as transcription factors, ribosomal proteins, chaperones, lipid binders, and Ub binding domains[2,118,119].

As its name suggests the ZNF216 protein contains two zinc finger domains, an N-terminal A20 type zinc finger domain shown to mediate interactions with Ub, RIP, and IKKγ[113], and a C-terminal AN1 type zinc finger which binds two zinc atoms via two Cys3/His1 sites. Although the Znf_AN1 has been proposed to mediate an interaction with TRAF6 it has no confirmed function in any protein [113,117,120].

The Znf_A20 domain was originally identified in the TN-Fα inducible protein 3 (TNFaIP3) which harbours 7 tandem Znf_A20 domains and is a negative regulator of NF-κB signalling. TNFaIP3 demonstrates both deubiquitinating (DUB)-enzyme and Ub ligase activities, shown to catalyse first the removal of a lys63-linked poly-Ub chain and then the ligation of a lys48-linked Ub chain to the TNF receptor-1 binding protein RIP[32]. The Znf_A20 domain is a Cys2/Cys2 zinc finger which has since been identified in a variety of different proteins including a large family of stress associated proteins (SAP) in plants[120] and several Ub binding regulators of NF-κB signalling including Rabex-5, AWP1 and TEX27[20,21,121-124]. The precise function of the Znf_A20 domain is still undefined. In TNFaIP3 the A20 domains have been linked to self association and possesses Lys48-Ub ligase activity[32,125], while the Znf_A20 domains of ZNF216, AWP1, and TEX27 are linked to non-covalent interactions with Ub without evidence for Znf_A20 mediated ligase activity. Znf_A20 mediated self association has also been reported for SAP8 in plants[126] which also demonstrated an interaction between its Znf_A20 domain and its own

AN1 type zinc finger domain, an interaction which has also been suggested in ZNF216[117,126].

X-ray crystal structures of Rabex-5 and NMR derived solution structures for Znf_A20 domains 1, 4 and 7 from TNFaIP3 confirm the formation of the Cys2/Cys2 zinc finger which forms a so called 'treble-clef' type zinc finger. The Znf_A20 is comprised of a zinc knuckle of non-regular but stable secondary structure followed by a short helix[20,21,127].

Several functions have been ascribed to the Znf_A20 domain representing an interesting and diverse set of protein-protein interactions, although many of these are yet to be confirmed by appropriate biophysical approaches.

### 3.1.4 Summary and Aims

The ZNF216 protein has been shown to play a key role in muscle atrophy. In recent years the mechanisms controlling muscle atrophy have attracted much attention from the scientific community due to their relevance in various fields including ageing, quality of life, sports medicine and the prognosis of many diseases including cancer cachexia and AIDS[111,112]. The emerging role of components of the UPS including the ZNF216 protein in the control of muscle atrophy is becoming clear. In the case of the ZNF216 protein its Znf_A20 domain appears to be key to its function as a proteasomal shuttling protein. The Znf_A20 represents a domain for which a lot of questions still remain, with only three structures of the free domain available in the RPDB, which all originate from the same protein.

In order to better understand the molecular detail of the function of ZNF216, the presence of a Znf_A20 domain within the ZNF216 N-terminal region has been confirmed. Its oligomerisation state has been determined and its solution was structure elucidated using a combination of NMR spectroscopy and biophysical approaches.

## 3.2 ZNF216 Residues 11-35 Contain an A20 Type Zinc Finger

### 3.2.1 Sequence Alignment

The ZNF216 protein is 213 amino acids in length and has only two domains predicted by SMART[89]. Residues 11-35 are predicted to form an A20 type zinc

finger (Znf_A20) domain consistent with previous reports[113,117] while residues 154-191 are predicted to form an AN1 type zinc finger (Znf_AN1).

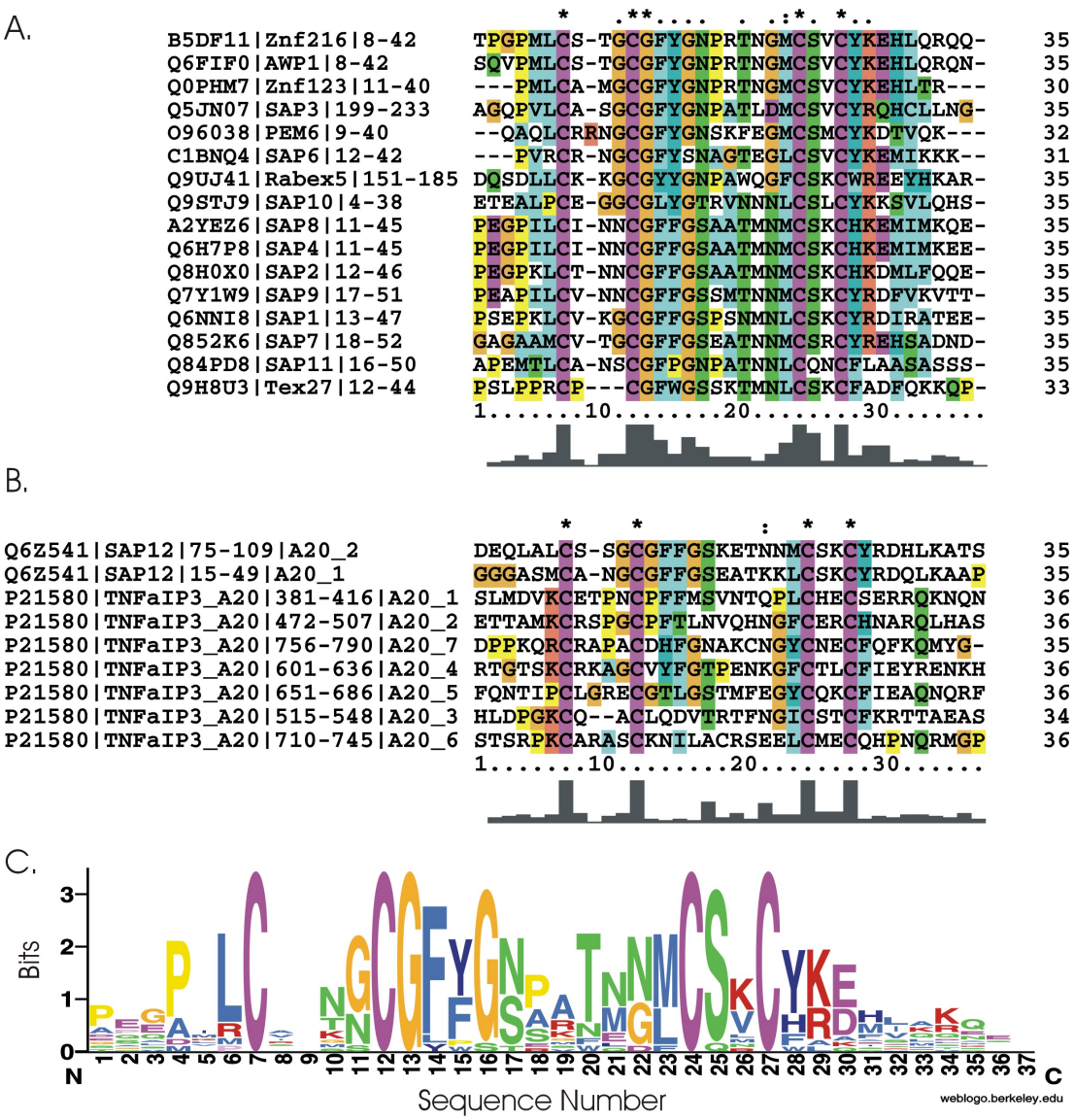



**Figure 3-2: Sequence alignment of Znf_A20 domains (A) From proteins containing a single Znf_A20. (B) Those with multiple Znf_A20. Each protein is labelled by its UniProt accession number followed by its common name, the domain residue range and Znf_A20 number where multiple Znf_A20 are found. (C) The sequence alignment from proteins containing a single Znf_A20 were used to generate a consensus sequence logo.**

The ZNF216 Znf_A20 domain was aligned with a set of confirmed Znf_A20 domains from the UniProt[85] databases aligned in Clustal X[86]. On first appearance the Znf_A20 seems to possess little sequence conservation with only the four cysteine

residues being conserved across all identified Zn_A20 domains; however two of the originally aligned proteins harbour multiple Znf_A20 domains, the stress associated protein (SAP) 12 which contains two and the tumor necrosis factor alpha (TNFα)-induced protein 3 (TNFaIP3) which contains seven. Of the seven Znf_A20 domains only the fourth Znf_A20 domain is proposed to be key to the proteins function[32]. The removal of these Znf_A20, which have the potential for functional redundancy due to duplication of the domain, results in some potentially common features within the Znf_A20 domain, however the sequence similarity remains low (Figure 3-2).

Other than the position of the four cysteine residues which are expected to complex the single zinc atom, several characteristics appear to be common to Znf_A20 domains. The degree of conservation is illustrated by the consensus sequence logo[87,88] shown in Figure 3-2C. Using the numbering used in Figure 3-2 the four cysteine residues occupy numbers 7, 12, 24, and 27 but the following features are also common; a proline residue at position 4 represents the first common residue, followed by a hydrophobic residue at position 6 and 23 prior to the first and third cysteines, glycines at positions 11, 13, and 16 with position 13 being completely conserved across the selected sequences, a di-aromatic motif at positions 14 and 15, polar residues at 17, 20, and 25 with serine preferred at position 25, and a motif corresponding to an aromatic followed by a basic residue and an acidic residue at positions 28 – 30 immediately following the fourth cysteine.

To date there are four Znf_A20 structures deposited in the RPDB (www.pdb.org), three solution NMR structures determined by the RIKEN structural genomics initiative (RSGI), representing the first, fourth and seventh Znf_A20 domains from TNFaIP3 (RPDB ID: 2EQG, 2EQE and 2EQF), and an x-ray crystal structure of the Znf_A20 domain of Rabex-5 in complex with mono Ub solved independently by two groups (RPDB ID: 2FIF and 2C7N)[20,21]. Although all four structures confirm the binding of zinc by the four cysteines, the three free solution structures from TNFaIP3 have many of the common characteristics listed above missing with only the fourth Znf_A20 domain resembling the proposed consensus sequence. In fact despite the TNFaIP3 being the namesake of the Znf_A20 domain, the seven which it harbours could in sequence be considered atypical, lacking most of the common features (Figure 3-2). The fourth available Znf_A20 structure from the Rabex-5 protein in a non-covalent complex with mono-Ub and again is atypical in the sense that, in this case, the Znf_A20 during evolution became a fusion with a

second Ub binding domain (UBD), with the helix of the Znf_A20 extending to form an MIU. However this structure does provides some potential insight into the conservation of some residues: The di-aromatic motif at positions 14 and 15 and the serine at position 25 (Figure 3-2) are described as being essential for the interaction of the Znf_A20 domain with Ub, with additional contributions from the polar residue at position 17[20,21]. The authors also suggest that the conserved pattern of glycine and prolines may be required for stable folding[20,21].

The Znf_A20 of ZNF216 has little sequence similarity with the available solution structures and while it shares common features with that of Rabex-5, the Znf_A20 of Rabex-5 is a fusion with an MIU domain and its structure is that of its complex with Ub, both of which are aspects which could influence the observed fold of the Znf_A20. The shortage of Znf_A20 domains in the RPDB, particularly those containing the consensus sequence proposed here further necessitates the need for structural and functional studies of this small zinc finger domain such as that possessed by ZNF216.

### 3.2.2 The Znf_A20 is Monomeric and Binds a Single Zinc Atom

The Znf_A20 domain was expressed in *E. coli* (C41-DE3) as a GST-fusion protein and purified after removal of the GST tag to high purity confirmed by ESI-MS (Section 2.3.3).

#### 3.2.2.1 Size Exclusion Chromatography

As part of its purification protocol size exclusion chromatography (SEC) was performed using a HiLoad™ Sephacryl™ S-100 HR gel filtration column. During SEC it was observed that the Znf_A20 domain eluted at a volume consistent with a much larger protein. SEC was repeated with a set of low MWt proteins (GE HealthCare) which were used to generate a calibration curve; the gel phase distribution coefficient for Znf_A20 gives an apparent native mass of the Znf_A20 domain of 12 kDa, almost double the expected 6.6 kDa (Figure 3-3).

Although the apparent increase in mass could be attributed to self association of the Znf_A20 as reported for other Znf_A20 domains[125,126], other attributes must be taken into account. The calculated gel phase distribution coefficient of a given protein is affected by shape as much as mass and as such is more accurately correlated with Stoke's (Hydrodynamic) radius[39]. The data that was used for the

calculation of apparent MWt can also be used to calculate the proteins Stoke's radius (Appendix C)[39]. From SEC of the Znf_A20 a Stoke's radius of 16.1 ± 1.0 Å was determined using the method of Siegel *et al*[39].

The Znf_A20 construct used in this study is a total of 62 residues in length, with the Znf_A20 predicted to occupy around half of that, the remainder is predicted to be disordered. Such a large portion of disordered protein would likely cause an increase in the Stoke's radius which would result in a larger than expected apparent mass. This would be consistent with the slightly smaller than doubled mass which could be the result of an averaging of the apparent mass due to the equilibrium between monomer and dimer. However if this were the case one might expect broadening of the elution peak in the size exclusion which is not the case (Figure 3-3). Neither explanation can be ruled out on the basis of size exclusion alone, it is also not assumed that a combination of both of the above explanations is not the cause of the increased apparent mass and further experiments are required.

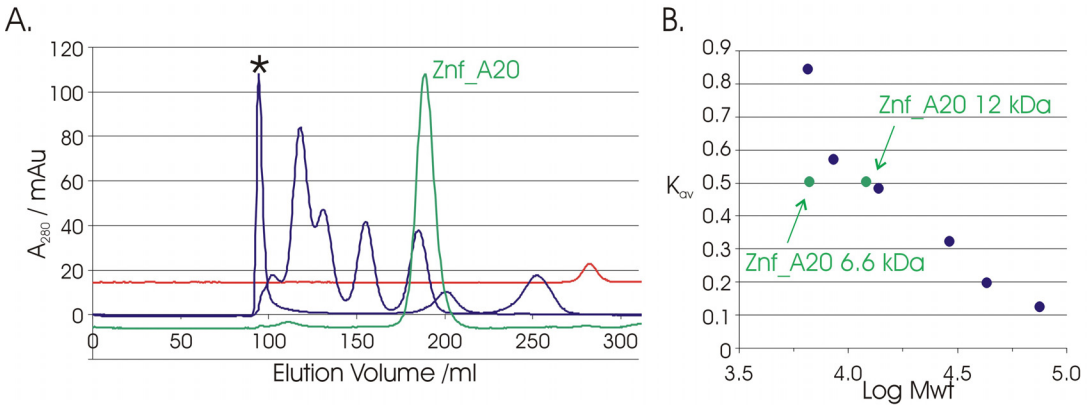



**Figure 3-3: (A) SEC trace for Znf_A20 (Green) Salt (Red) and Low MWt protein (Blue)(* denotes blue dextran). (B) The standard curve used to calculate the apparent native mass of Znf_A20. The expected mass of 6.5 kDa for monomeric Znf_A20 and the calculated mass are shown in green and labelled.**

#### 3.2.2.2 ITC, AUC and ESI-MS Confirm the Monomeric State

The dilution of oligomeric proteins into the calorimetric cell of an ITC instrument, results in an endothermic dissociation peak. The continuous injection of concentrated protein into the ITC cell until saturation is reached can be used to calculate dimer dissociation constants (Section 2.3.2.1). Injection of 200-500 µM Znf_A20 into buffer in the ITC cell gives no endothermic signal comparable to that seen for the RAD23-UBA which is known to be monomeric (Unpublished data from

Miss J. Adlington) (Figure 3-4). The lowest concentration experienced in the experiment is after the first injection, in this case representing a dilution from 200 µM to 0.7 µM, the absence of an endothermic signal would suggest a dimer dissociation constant of below 0.5 µM or a monomeric system.

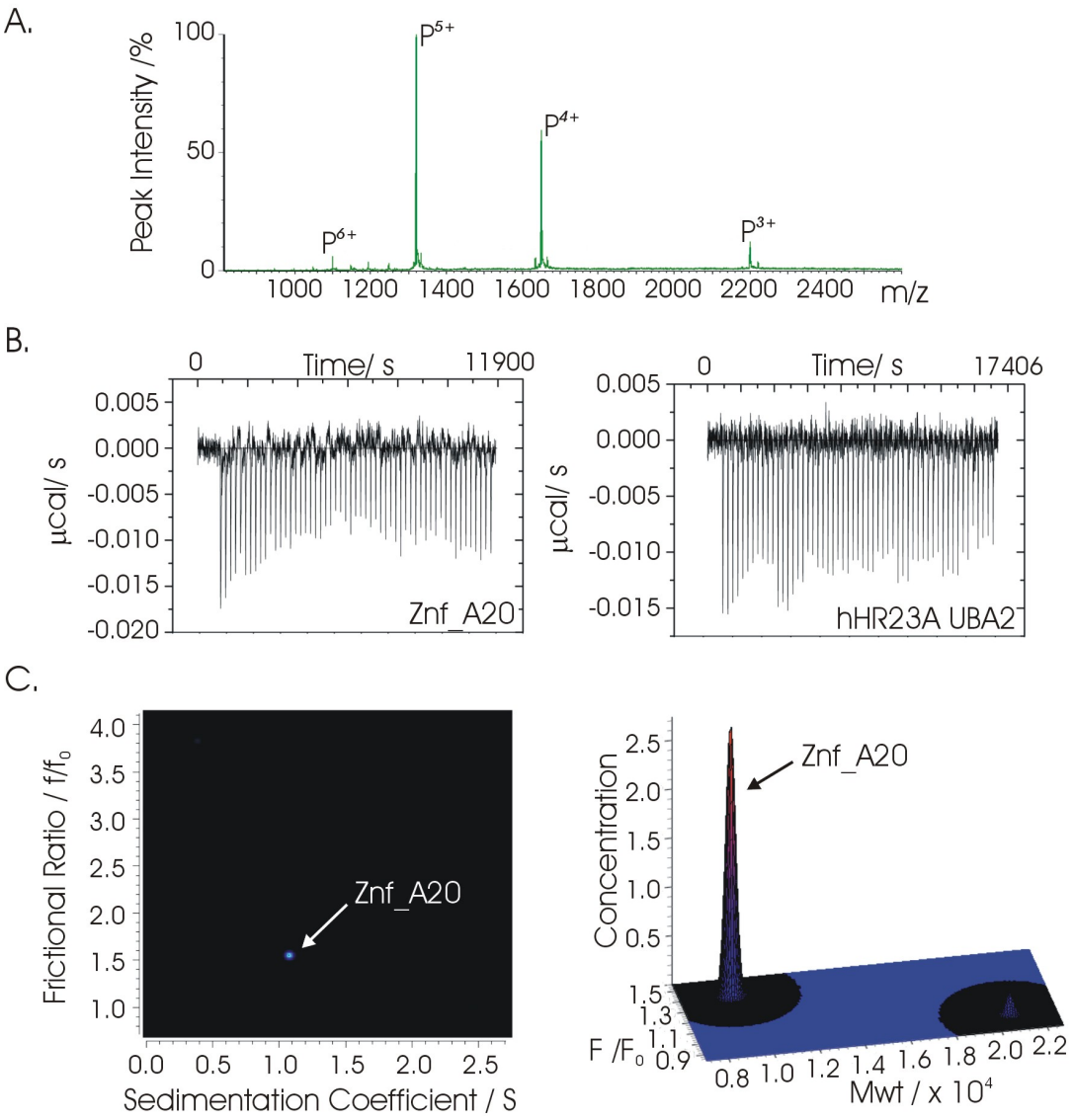



**Figure 3-4: (A) Native ESI-MS spectrum for the Znf_A20. (B) ITC dilution thermal output for the Znf_A20 and RAD23-UBA. (C) AUC data for the Znf_A20.**

In order to distinguish between a monomer or high affinity oligomer, mass measurements were made using native ESI-MS and AUC at medium to low concentrations. AUC was performed by Dr Dave Scott on Znf_A20 in the concentration range 0.1 mg/ml to 1 mg/ml (15-150 µM) using a Beckman XL-I

analytical ultracentrifuge, using absorbance optics measured at 230 nm. The AUC determined molecular weight showed a homogeneous distribution with a sedimentation coefficient of 1.05 ± 0.5 Svedbergs (S) and an apparent MWt of around 9 kDa. Given the high error for low molecular weight proteins using sedimentation techniques and the presence of a single well defined species, this is consistent with a stable monomer of 6.5 kDa (Figure 3-4). A sedimentation coefficient of 1.05 ± 0.5 Svedbergs with a monomer molecular weight of 6595.3 gives an estimated Stoke's radius of 16.9 ± 0.5 Å, while a dimer molecular weight of 13190.6 gives an estimated Stoke's radius of 34 Å. Combining the AUC determined sedimentation coefficient of 1.05 ± 0.5 S and the SEC determined Stoke's radius of 16.1 ± 1.0 Å gives a molecular weight of 6.2 ± 0.4 kDa. The AUC and SEC data therefore strongly suggests that the Znf_A20 is monomeric in solution. The sedimentation coefficient, Stoke's radius and frictional ratio for the Znf_A20 suggest a non spherical structure which is reasonably extended for its small size potentially explaining its large apparent mass by SEC.

In order to support this ESI-MS was recorded on Znf_A20 in the low concentration range, of 0.5-3 μM corresponding to the lowest concentration points measured during the ITC experiments. Unlike other ionisation methods such as matrix assisted laser desorption ionization (MALDI), ESI gives multiple charge-states; the number and average charge-state are proposed to be indicators of the protein fold or more accurately the solvent exposed surface area in solution or during early desolvation[46,128]. Largely unfolded states give broader charge-state distributions as a result of increased dynamics or structural heterogeneity and a higher average charge-state as a result of increased solvent exposed surface area[46,128]. The Znf_A20 domain gives a narrow charge-state distribution with an average charge-state of 5.05 consistent with a folded domain with a solvent exposed surface area of around 3765 $Å^2$ according to the data of Kaltashov *et al*[46]. In an ESI-MS the mass to charge ratio of a monomer and dimer can be equal where the dimer has double the mass and double the charge giving rise to identical values in an ESI-MS experiment. However the dimeric species would be expected to generate extra peaks between the 'monomeric' peaks due to charge-states with odd numbers. The complete absence of any such species in this case is consistent with the observed mass to charge species representing the monomeric species only (Figure 3-4).

The ESI-MS gives a monomeric weight of 6595.3 ± 0.4; the theoretical mass of the apoprotein is 6533.2, the mass of a single zinc atom is 65.38, giving a combined mass of 6598.5. The recorded mass of 6595.3 is consistent with the apoprotein bound to a single zinc atom coordinated by the four cysteine residues, the ~4 Da differences corresponds to the loss of the four cysteine thiol protons to form thiolates (Figure 3-4)[129].

#### 3.2.2.3 Circular Dichroism Confirms the Structural Requirement for Zinc

Circular Dichroism was performed to investigate the secondary structure content of the Znf_A20 domain and more specifically the effect of zinc binding to the structure of the Znf_A20 domain.

CD spectra were recorded on Znf_A20 at a concentration of 25- 30 μM. CD spectra were then recorded continuously over a period of two hours after an injection of EDTA to a final concentration of 5 mM. The addition of EDTA results in the removal of zinc ions from solution preventing it from binding to the Znf_A20 once released. If the zinc is required for correct folding of the Znf_A20 this should result in an inability of the protein to refold following zinc release coupled unfolding[43].

The CD spectrum of Znf_A20 contains minima at 222 and 205 nm consistent with a folded protein which is largely helical in nature[41]. The addition of EDTA to a final concentration of 5 mM induced a slow unfolding transition which reached its end point after around 2 hours. This represents a slow EDTA induced unfolding transition with an approximate rate of $3.1 \times 10^{-4} \pm 0.5 \times 10^{-4}$ $s^{-1}$ (Figure 3-5). The binding of zinc to EDTA is rapid and would not be rate limiting in this case, the slow observed rate can therefore be attributed to the slow unfolding /zinc release rate of the Znf_A20[43]. The presence of an apparent isodichroic point at around 200 nm suggests a 2 state folding transition with no detectable intermediates, however this would require a more detailed kinetic analysis (Figure 3-5)[130].

The CD spectrum of the EDTA treated Znf_A20 is consistent with a largely unfolded protein; the difference spectrum between zinc bound and EDTA treated Znf_A20 shows a negative ellipticity at 222 nm and a positive ellipticity at 190 nm consistent with the loss of a large portion of helical structure (Figure 3-5). The largely helical character is consistent with other reported Znf_A20[20,21] and secondary structure predictions using the APSSP2 web server. The loss of structural features

with the removal of zinc is consistent with the zinc atom having a key role in maintaining the fold of the domain.

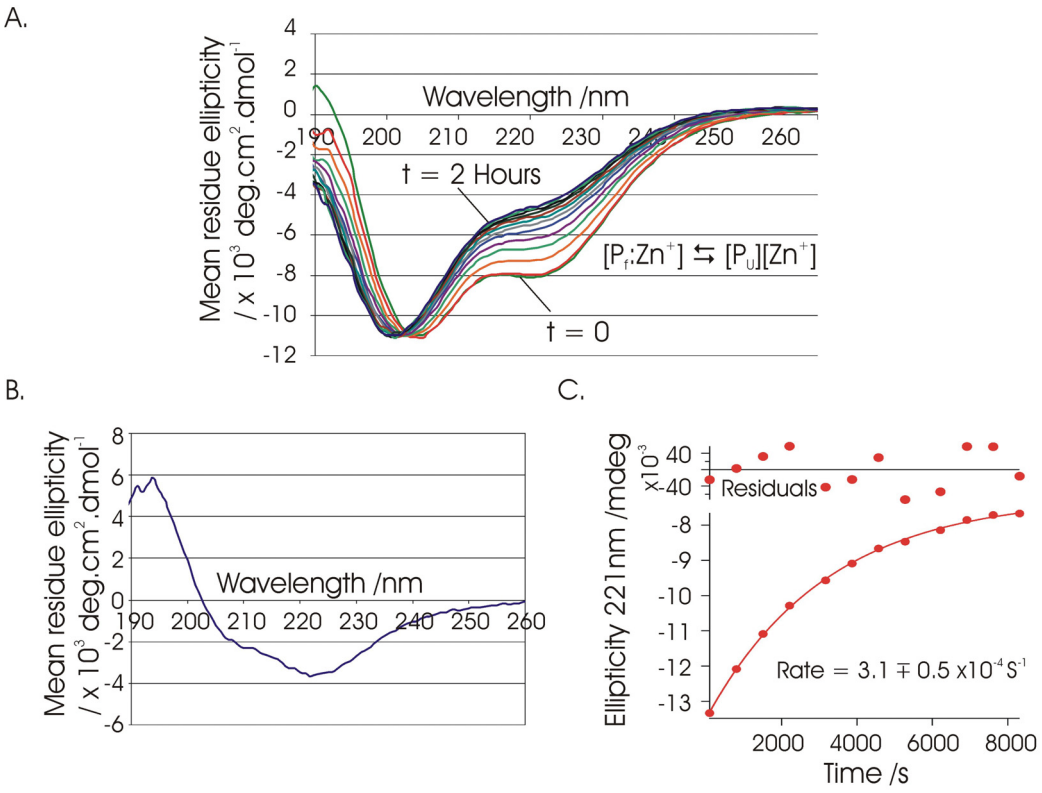



**Figure 3-5: (A) EDTA induced unfolding of the Znf_A20 monitored by CD. (B) The EDTA derived difference spectrum. (C) EDTA induced change in ellipticity as a function of time was fit to an approximate rate constant.**

### 3.2.3 Summary

From a combination of the results of SEC, ITC, AUC, ESI-MS, and CD the Znf_A20 domain, or more accurately ZNF216 residues 1-60, is monomeric in solution but possess an extended structure most likely the result of pronounced flexibility outside of the Znf_A20 domain. The domain binds a single zinc atom which is required for correct folding of the domain which appears to be largely helical in content.

## 3.3 The NMR Structure of the Znf_A20 Domain

To better characterise the ZNF216 atrogene protein a full structural characterisation of its Znf_A20 domain in solution at atomic resolution was

performed using multi-dimensional solution NMR spectroscopy techniques (Section 2.4).

### 3.3.1 The NMR Assignment of the Znf_A20

NMR spectra were recorded on unlabeled, $^{15}$N-labelled and $^{13}$C/$^{15}$N-labeled Znf_A20 domain. Assignment was achieved using a combination of homonuclear 2D NOESY and TOCSY data and 3D backbone assignment data. Assignment employed a combination of the classical approach outlined by Wuthrich *et al* in the mid 80s[131,132], combined with through bond correlations which have made the assignment of small proteins routine in modern NMR laboratories[82].

Briefly the classical approach relies on a combination of $^{1}$H-$^{1}$H 2D NMR spectra with through space (NOE) and through bond (COSY/TOCSY) correlations. Amino acid spin systems are first identified based on characteristic 'finger prints' of chemical shifts and through bond correlations. Once these 'finger prints' are used to identify resonances belonging to a given amino acid type, sequential through space connectivities between adjacent residues are used to link them. Sequential connectivities typically include correlations between $H^N_i$ and $H^N_{i\pm1}$, $H^N_i$ and $H\alpha_{i-1}$ and to a lesser extent $H^N_i$ and $H\beta_{i-1}$[82](Figure 3-7A). The linkage information builds a sequence of known direction (N to C) and composition (e.g. N-MGFSDEP-C) until a break in the expected connectivities is encountered. This builds up a series of short fragments of known sequence which can then be compared to the protein's primary sequence in order to identifying matching sequences; this allows the gradual identification of each amino acid in the protein primary sequence.

The classical method can be used with great success for small well folded proteins with good chemical shift dispersion[132]. However in larger proteins or proteins with limited chemical shift dispersion the problem of spectral overlap occurs. Essentially this is the product of an increased likelihood of two protons having near identical or overlapping chemical shifts as you increase the number of amino acids and therefore protons. This can also become a problem in proteins with large amounts of intrinsically disordered regions or repetitive sequences, both of which may result in the presence of amino acids in very similar magnetic environments and therefore increased likelihood of possessing similar chemical shifts. Where there is some degree of spectral overlap, as was the case for the Znf_A20, some resonances can be assigned using measurements recorded in $D_2O$ to

remove exchangeable resonances from the spectrum to reduce its complexity, or by recording data at different temperatures; at different temperatures chemical shifts will alter to different degrees depending on the proton which can lead to separation of overlapping resonances.

Spectral overlap can be alleviated more effectively by the use of 3D experiments. The classical assignment protocol for example, can be performed using $^{1}$H-$^{15}$N-HSQC-NOESY and $^{1}$H-$^{15}$N-HSQC-TOCSY which separates the amide chemical shifts and its associated correlations based on the chemical shift of its bound $^{15}$N giving rise to a 3D grid. This has been particularly useful in resolving potential ambiguity in 2D experiments but does restrict you to using only backbone NH correlations which can limit the ability to identify amino acid type.

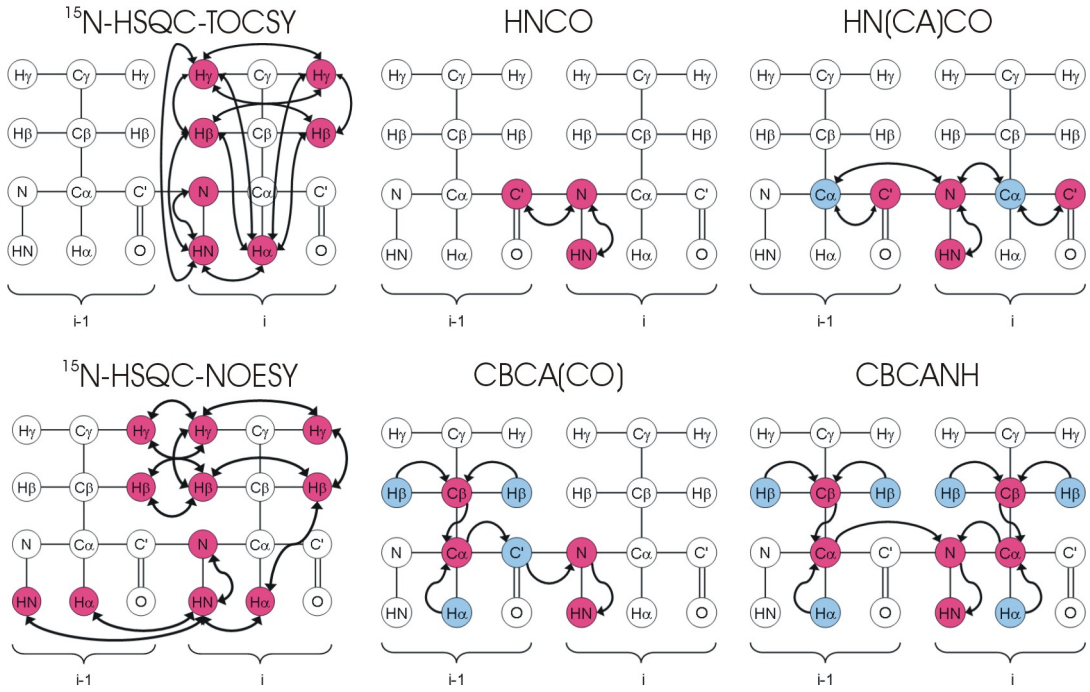



**Figure 3-6: Schematic of correlations expected in 3D backbone assignment experiments[82], Blue circles and arrows represent the magnetisation transfer pathway, red circles represent detected atoms.**

The modern approach is essentially very similar to the classical approach but relies largely on through bond correlations between adjacent residues rather than through space connectivities (Figure 3-6). This requires the use of $^{15}$N-$^{13}$C isotopic labelling of the protein to be studied and the use of 3D spectra with correlations based on the $^{1}$H-$^{15}$N-HSQC (N-H) correlated with either the Cα and Cβ atoms or the

backbone C`. The correlations used in the main 3D backbone assignment experiments are illustrated in Figure 3-6.

While the proton correlation 'finger prints' used in the classical approach may be used to identify amino acid types, this is commonly achieved instead, by comparison of the $^{13}C\alpha$ and $^{13}C\beta$ chemical shifts to a reference data base of previously assigned chemical shifts of the 20 amino acids. The reference chemical shift data set used for predicting amino acid type was obtained from the BMRB.

Assignment of the Znf_A20 was achieved using a combination of both the classical and modern approaches, utilising 2D data recorded at 298 K and 288 K, as well as 2D data in $D_2O$, in combination with modern 3D backbone assignment data at 298 K. 3D backbone assignment data yielded good correlations running throughout the 62 residue Znf_A20 construct with few breaks in connectivity. This allowed the backbone assignment including $H^N$, $N^H$, Hα, Cα, Cβ and C' chemical shifts for all non proline residues. Cα, Cβ and C' chemical shifts of proline were obtained (where the following residue was not proline) using HNCO and CBCA(CO)NH experiments (Figure 3-6). The assignment of residues Ser31 – Tyr34 is shown in Figure 3-7 demonstrating both the classical through-space sequential connectivities and modern through bond sequential connectivities. A full table of the assigned chemical shifts can be found at the BMRB under accession number 17023.

The fully assigned $^1H$-$^{15}N$-HSQC of the Znf_A20 domain is shown in Figure 3-8 demonstrating a complete backbone assignment of all residues with the exception of the N terminal glycine of the 'GS' tag. The $^1H$-$^{15}N$-HSQC shows extra resonances originating from *cis-trans* proline isomerisation which will be discussed later (Section 3.3.4).

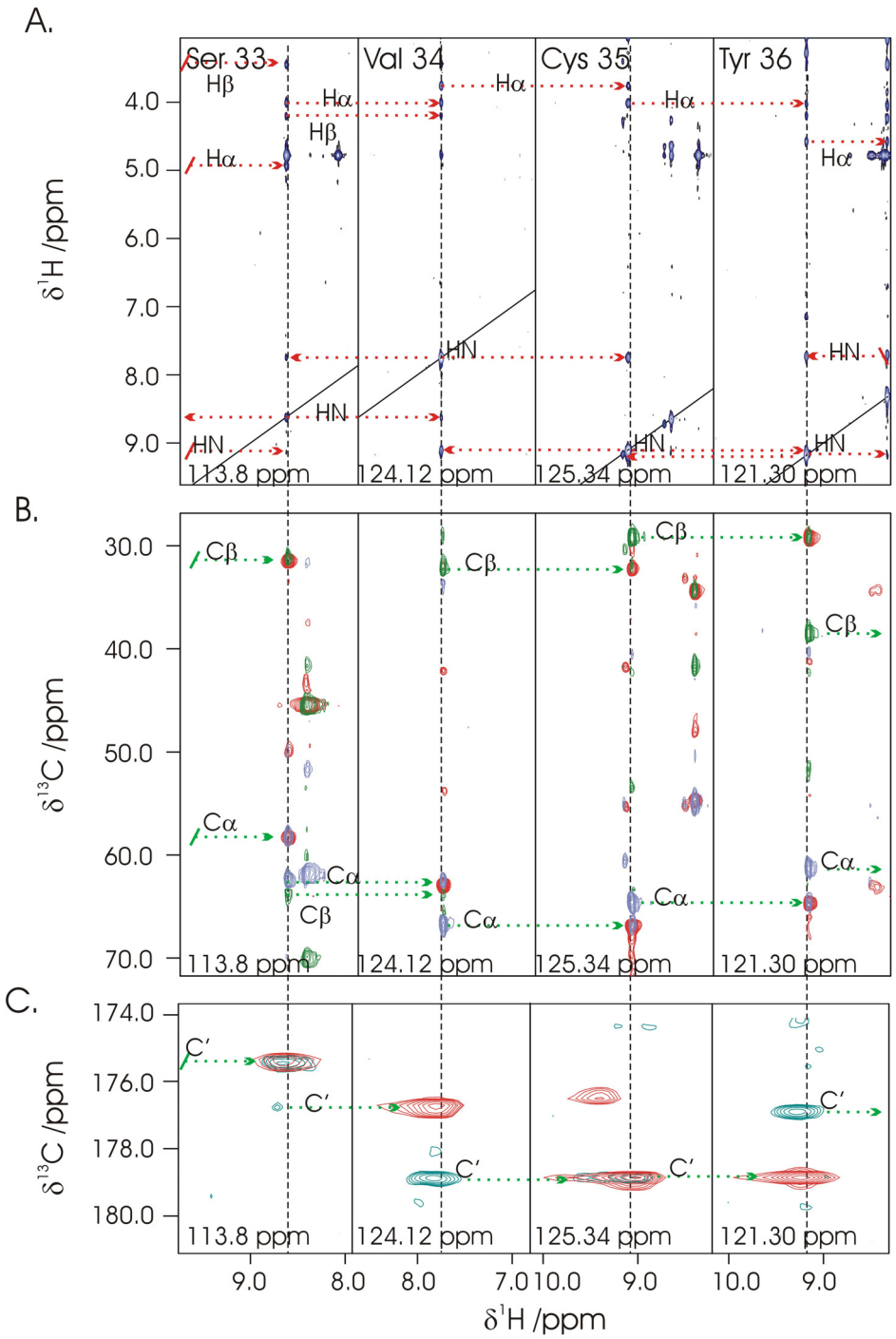



**Figure 3-7: (A) Assignment of S31 to Y34 by classical sequential NOE connectivities (Red arrows) using a $^{1}H$-$^{15}N$-HSQC-NOESY. (B) Through bond correlations (Green arrows) involving Cα and Cβ atoms using CBCANH (Blue and Green) and CBCA(CO)NH (Red). (C) C' mediated connectivities using HNCO (Red) and HN(CA)CO (Blue). All spectra are 3D spectra, vertical lines represent the $^{1}H^{N}$ chemical shift, and horizontal lines represent the different correlations with that $^{1}H^{N}$. The chemical shift of the associated $^{15}N^{H}$ is labelled.**

.

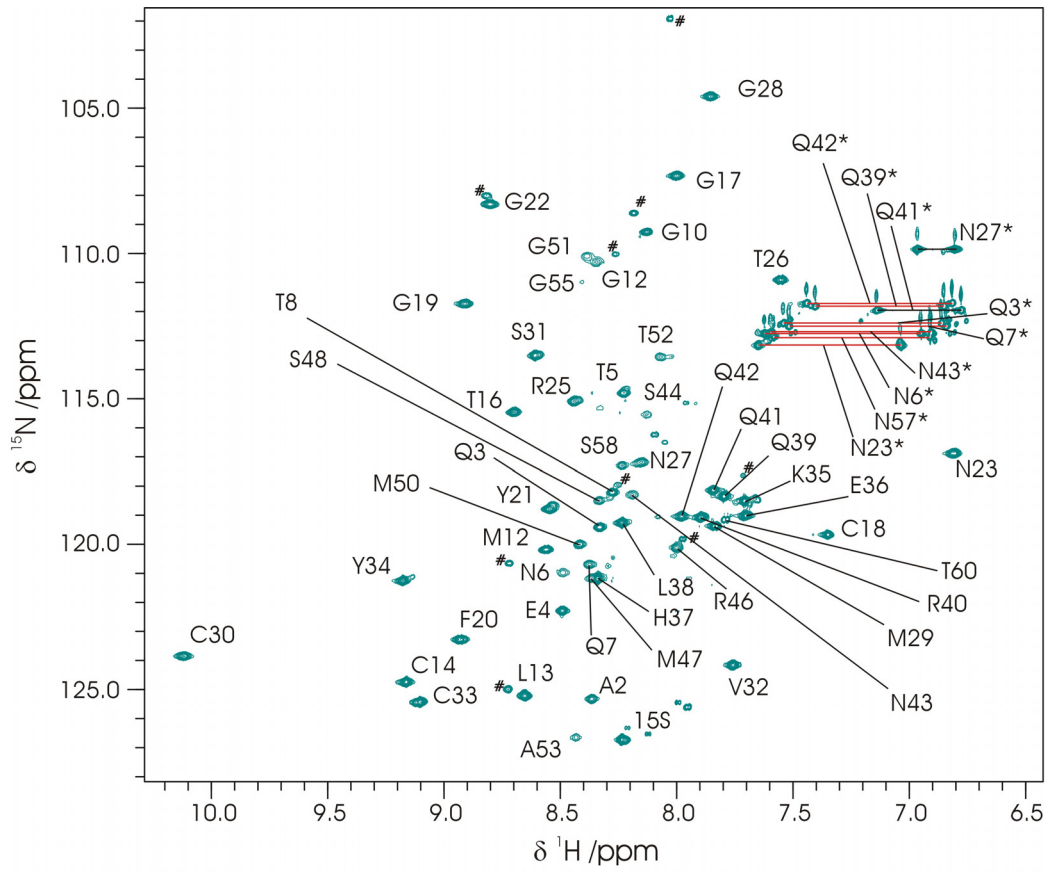



**Figure 3-8: Fully assigned $^{1}$H-$^{15}$N-HSQC of the Znf_A20 from ZNF216, red horizontal lines represent $NH_2$ of Asn and Gln side chains which are marked with a *, 'Extra' peaks from cis-trans Pro isomerisation are labelled with a '$^{\#}$'.**

The fully assigned backbone is then used as a basis to assign as many of the side chain atoms as possible. This is often hindered by spectral overlap which can be difficult to alleviate even with 3D data due to a close correlation between carbon-13 and proton chemical shifts, meaning that where there is proton chemical shift overlap there is often insufficient separation in the $^{13}$C-chemical shifts to resolve the chemical shifts unambiguously. Although in this case this prevented a full assignment of all proton and carbon chemical shifts, large portions were still assigned using $^{13}$C-separated 3D experiments including HCCH-TOCSY and $^{13}$C-HSQC-NOESY experiments. This is illustrated in Figure 3-9 showing the clear correlation between proton and carbon-13 chemical shifts which gives the $^{13}$C-HSQC an apparent 'diagonal'. A schematic of the experiments used to assign side chain resonances is also included in Figure 3-9.

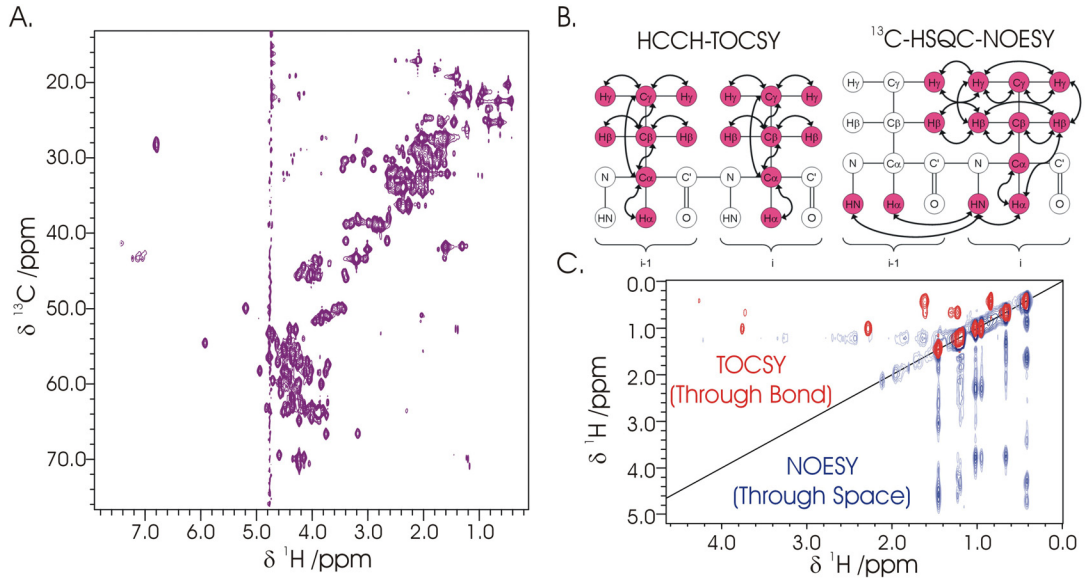



**Figure 3-9: (A) $^{13}$C-HSQC of Znf_A20. (B) schematic of correlations expected in 3D side chain assignment experiments[82]. Blue circles and arrows represent the magnetisation transfer pathway, red circles represent detected atoms. (C) Example spectra of HCCH-TOCSY (Red) and $^{13}$C-HSQC-NOESY (Blue).**

The backbone and side chain assignment is a requirement for full interpretation of the NOESY spectra; once the chemical shifts of the protons are known the through space correlations seen in NOESY spectra can be interpreted with confidence. NOESY correlations are seen for protons with a separation through space below 5 Å independent of distance in terms of primary sequence. For the Znf_A20 a total of 722 NOE correlations were unambiguously assigned from a combination of 2D $^{1}$H-$^{1}$H-NOESY, $^{1}$H-$^{15}$N-HSQC-NOESY, and $^{1}$H-$^{13}$C-HSQC-NOESY. The full set of NOE restraints can be found deposited at the BMRB under accession number 17023.

### 3.3.2 The Solution Structure of the Znf_A20 domain

Chemical shifts were used to generate dihedral restraints using the programs TALOS[84] and PREDITOR[83]. A total of 177 dihedral angles, 722 NOE distance restraints and 30 backbone $^{1}D_{NH}$ RDC were used in a simulated annealing protocol in XPLOR-NIH[133] (Section 2.5.3) to generate 200 structures of the Znf_A20 domain. χ1 angles predicted using the program PREDITOR are inaccurate in residues which are solvent exposed due to reduced restrictions of the Cα-Cβ bond rotation in these cases[83]. Therefore during initial rounds of structure generation dihedral restraints for

χ1 angles were omitted and reintroduced for residues which are likely to be buried. Previous biophysical data and sequence alignments suggested a single zinc coordination site between the four cysteines. Given the strong evidence for the cysteines as the zinc-ligating residues and the lack of clear alternatives for zinc coordination, the simulated annealing protocol was performed with the four cysteines fixed in a tetrahedral geometry around the zinc atom coordinated by the cysteine sulphurs. In order to avoid inaccuracies caused by motional averaging backbone $^{1}D_{NH}$ RDC were only used for residues possessing large values of $^{15}N(^{1}H)$ NOE as an indicator of low internal dynamics[134,135].

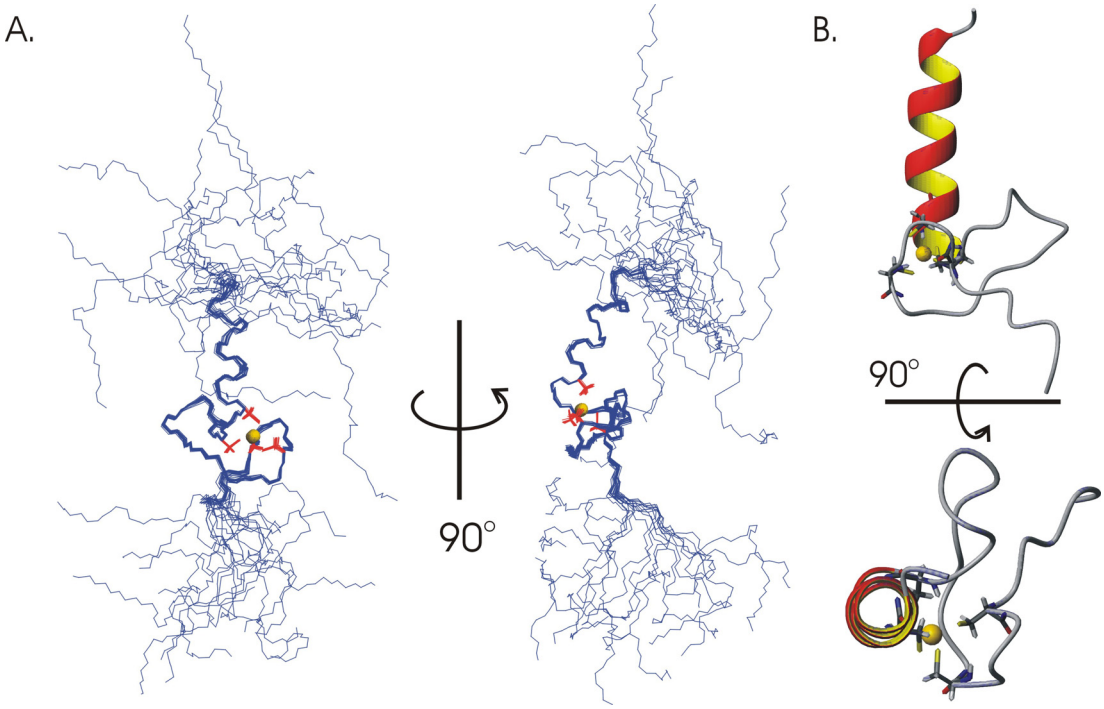



**Figure 3-10:(A) Backbone trace of the ensemble of the 20 lowest energy structures of the Znf_A20 domain, the side chains of the cysteines involved in zinc coordination are shown in red while all other side chains are omitted, the zinc atom is shown in gold. (B) Ribbon diagram of the lowest energy structure of Znf_A20 demonstrating the so called 'treble-clef' zinc finger motif.**

Of the 200 generated structures, a total of 42 had no NOE violations greater than 0.3 Å, no dihedral violations greater than 5°, or RDC violations greater than 3 Hz. From those 42 the 20 with the lowest energy were selected to represent the structural ensemble. The structural ensemble of the 20 lowest energy structures is shown in Figure 3-10. Both the N and C terminus show poor structural definition

while residues Met12 to Asn42 have a consistent fold, with a backbone and heavy atom RMSD of 0.309 and 0.845 respectively.

The fold of the Znf_A20 is characteristic of those reported previously forming an A20 type Cys2/Cys2 zinc finger. As predicted cysteines 14 and 18 form a zinc knuckle with cysteines 30 and 33 at the N-terminus of the helix, to complete the four coordinate zinc cluster with an 11 residue loop between the second and third cysteine to give a so called treble clef zinc finger motif (Figure 3-10)[127].

The 11 residue loop forms a region of stable non-regular secondary structure stabilized by an apparent network of hydrogen bonds including (backbone NH to backbone O), Cys14 with Phe20, Ser15 with Gly28, Gly17 with Cys14, Gly22 with Met12, and Asn23 with Met29 (Figure 3-11). The overall fold is further stabilized by contacts of both the loop and the zinc knuckle with the helix via CH-π interactions between the Hβ and γ methyl protons of Thr26 in the loop with the aromatic ring of Tyr34 in the helix and by the burial of Met29 against the helix and other side chains of the knuckle. This combined with the four coordinate zinc cluster gives rise to a very compact core, forming a stable structure with low backbone RMSD. The helix extends from Cys30 to Asn42, which although is considerably shorter than that seen for Rabex-5, is around 1.5 turns longer than other A20 domain structures in the RPDB[20,21].

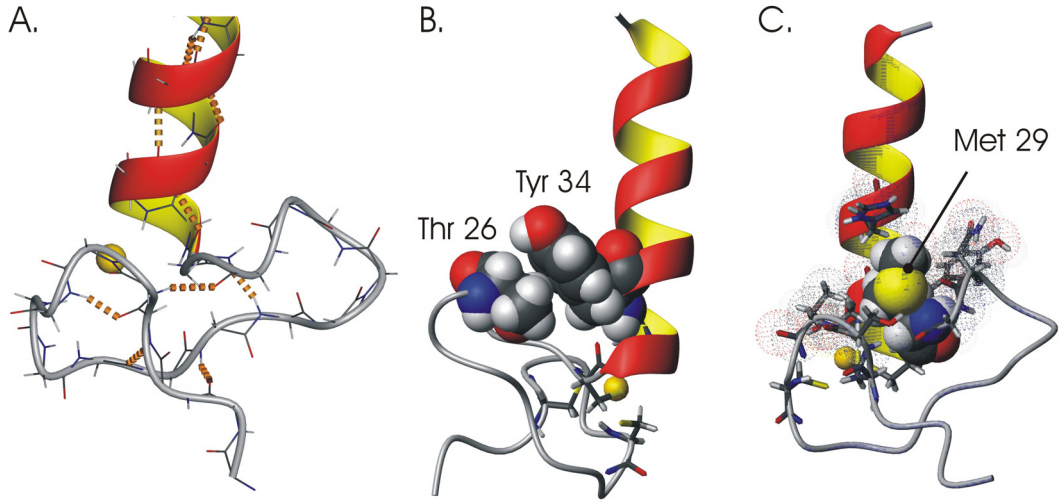



**Figure 3-11: Stabilising interactions around the 'treble clef'. (A) Proposed network of hydrogen bonds. (B) The packing of Tyr34 against Thr26 (C) The burial of Met29.**

The solved structure of the Znf_A20 is in good agreement with the previously solved Znf_A20 domains available in the RPDB, showing an average backbone RMSD of below 1.5 Å consistent with considerable structural conservation. The fold most closely resembles that of Rabex-5 with a backbone RMSD around the treble clef of < 0.8 Å.

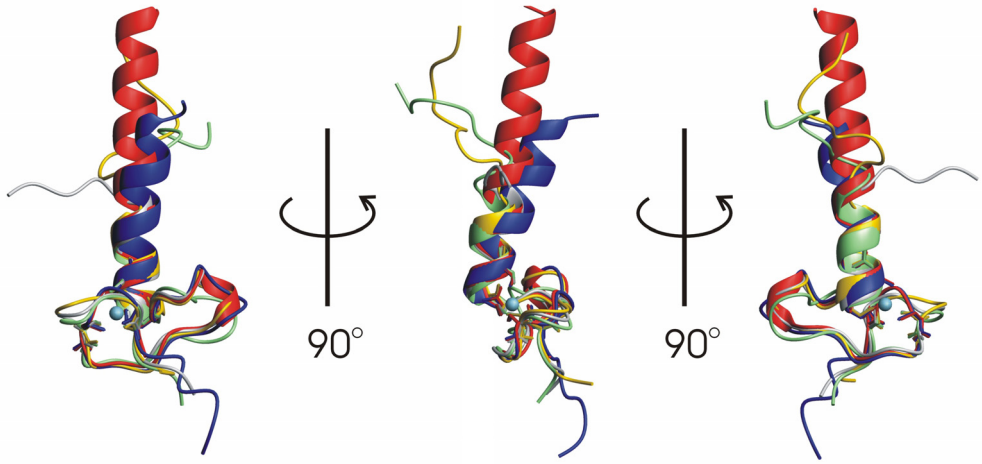



**Figure 3-12: Comparison of the Znf_A20 domain of ZNF216 (Blue) to previously solved structures including 2FIF (Red) 2EQG (Green), 2EQE (Grey) and 2EQF (Yellow) the flexible termini of the Znf_A20 and the helix of rabex-5 (2FIF) have been truncated for clarity.**

Coordinates and structure details for the Znf_A20 were deposited under the accession code 2KZY at the RPDB and at the BMRB under the accession numbers 17023.

## 3.3.3 Analysis of the Structural Model

### 3.3.3.1 Structure Validation

The 20 lowest energy structure were analysed using the protein structure validation software suite (PSVS)[103]. The quality of the structure was assessed in terms of physical properties such as bond lengths and angles, van der Waals clashes and distribution of dihedral angles (Ramachandran plot), as well as analysing the information content of the distance and dihedral restraints and the number of violations (Table 3-1).

**Table 3-1: Output from structure validation by the PSVS[103].**

| ***Summary of Conformationally restricting Restraints*** [a] | |
|---|---|
| *NOE Based Constraints:* | |
| Total | 446 |
| Intra-residue [i=j] | 54 |
| Sequential [i-j =1] | 152 |
| Medium range [i-j =<5] | 108 |
| Long range [i-j = ≥ 5] | 132 |
| NOE Constraints per restrained residue [b] | 7.2 |
| *Dihedral-angle constraints:* | 177 |
| Total number of restricting constraints [b] | 623 |
| Total number of restricting constraints per restrained residue [b] | 10.0 |
| Restricting long-range constraints per restrained residue [b] | 2.1 |
| **Total Structures Computed** | 200 |
| **Number of Structures Used** | 20 |
| **Residual constraint Violations** [a,c] | |
| *Distance violations /structure* | |
| 0.1-0.2 Å | 3.4 |
| >0.2 Å | 0 |
| RMS of distance violations / constraints | 0.02 Å |
| Maximum distance violation [d] | 0.20 Å |
| *Dihedral angle violations / structure* | 0 |
| 1-10° | 7.6 |
| >10° | 0 |
| RMS of dihedral angle violation / constraint | 0.46 ° |
| Maximum dihedral angle violation[d] | 4.60 ° |
| **RMSD Values** | |
| All backbone atoms | 7.7 Å |
| All heavy atoms | 7.6 Å |
| Ordered backbone atoms[f] | 0.3 Å |
| Ordered heavy atoms[f] | 0.9 Å |
| **Structure Quality Factors – overall statistics**[f] | Mean score |
| Procheck G-factor [e] (Phi /Psi only) | -0.15 |
| Procheck G-factor [e] (all dihedral angles) | -0.31 |
| Verify3D | 0.24 |

| | |
|---|---|
| MolProbity clash score | 3.11 |
| **Close contacts and deviation from ideal geometry[f]** | |
| Number of close contacts | 0 |
| RMS deviation for bond angles | 1.4 ° |
| RMS deviation for bond lengths | 0.011 Å |
| **[a] Analysed residues '1 to 62' (GS-1-60)** | |
| **[b] There are 62 residues with conformationally restricting restraints** | |
| **[c] Calculated for all constraints over the given residues, using sum over $r^{-6}$** | |
| **[d] Largest constraint violation among all the reported structures** | |
| **[e] Residues with sum of phi and psi order parameters > 1.8** | |
| **[f] Residues selected based on user defined residues (11-43)** | |
| **Generated using PSVS 1.4** | |

One of the key statistics on which protein structures are judged is the Ramachandran plot which analyses the correlation of the phi and psi angles of a given residue, compared to those seen in high resolution crystal structures, and to sterically-allowed combinations. The Ramachandran plot for the Znf_A20 shows that all restrained residues have combinations of Phi and Psi angles in regions deemed to be favoured or allowed (Figure 3-13). Outliers are restricted to only a few models and to residues outside of the core domain including Gly10, Pro11 and Gly45.

The NOE distance restraints used in structure generation are arguably the most important factor in NMR based structure generation, but the information content of a supplied distance restraint varies greatly. Not all of the supplied NOE restraints will impact the outcome of the structure, for example they could be NOE between two atoms of a fixed distance such as those between the protons of an aromatic ring, alternatively where a weak NOE was observed the outer boundaries of the distance restraint may be satisfied by the majority of conformations adopted by the molecule. Analysis of the distance restraints used for the generation of the Znf_A20 structure shows that of the 722 NOE restraints used in structure generation 446 were conformationally restricting, which combined with 177 dihedral angle restraints, gives an average of 10 restraints per residue, with an average of 2.1 long range restraints per residue. A summary of the distance and dihedral analysis from PSVS of the restraints used in the generation of the Znf_A20 structure is shown in Table 3-1 and Figure 3-14.

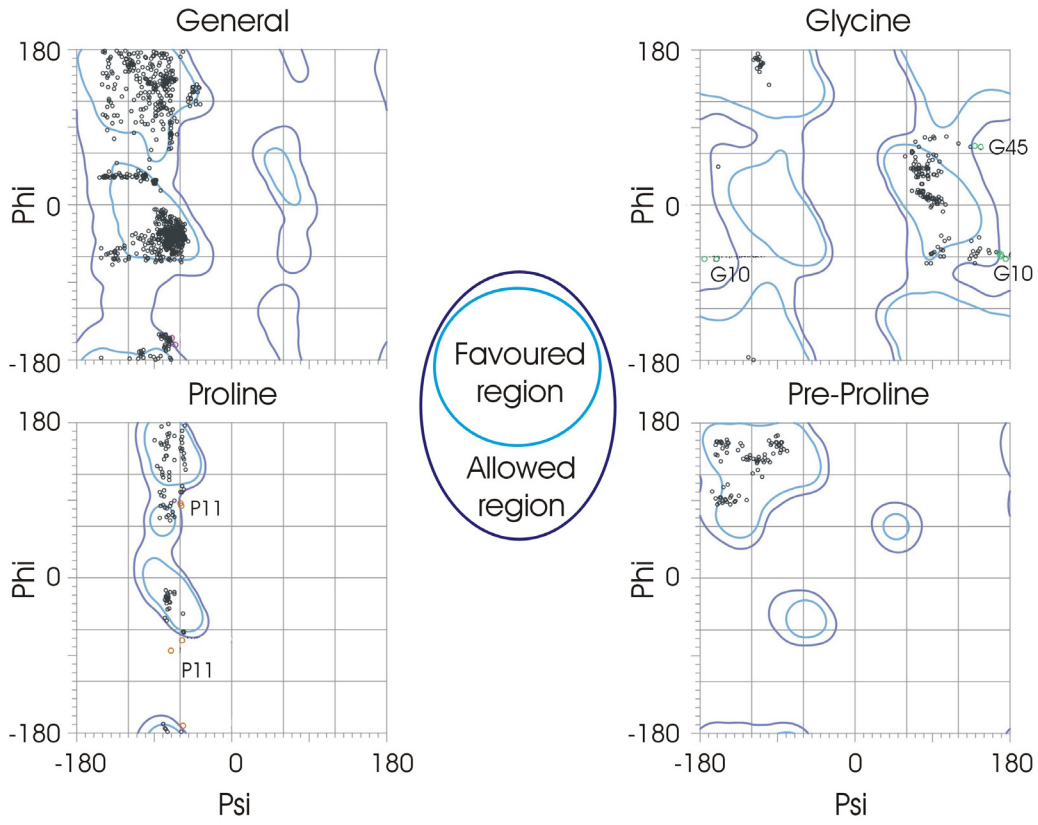



**Figure 3-13: Molprobity Ramachandran plot for the Znf_A20 structural ensemble, dark blue line represents favoured regions, light blue represents additionally allowed regions, severe outliers are labelled.**

**Table 3-2: Ramachandran statistics for the Znf_A20 structural ensemble.**

| **Procheck** | Favoured / % | Additionally Allowed / % | Generously Allowed / % | Disallowed / % |
|---|---|---|---|---|
| All Residues | 73.7 | 26.2 | 0.1 | 0.0 |
| Ordered Residues | 81.7 | 18.3 | 0.0 | 0.0 |
| **Molprobity** | | | | |
| All Residues | 83.7 | 15.1 | NA | 1.3 |
| Ordered Residues | 88.7 | 10.9 | NA | 0.3 |

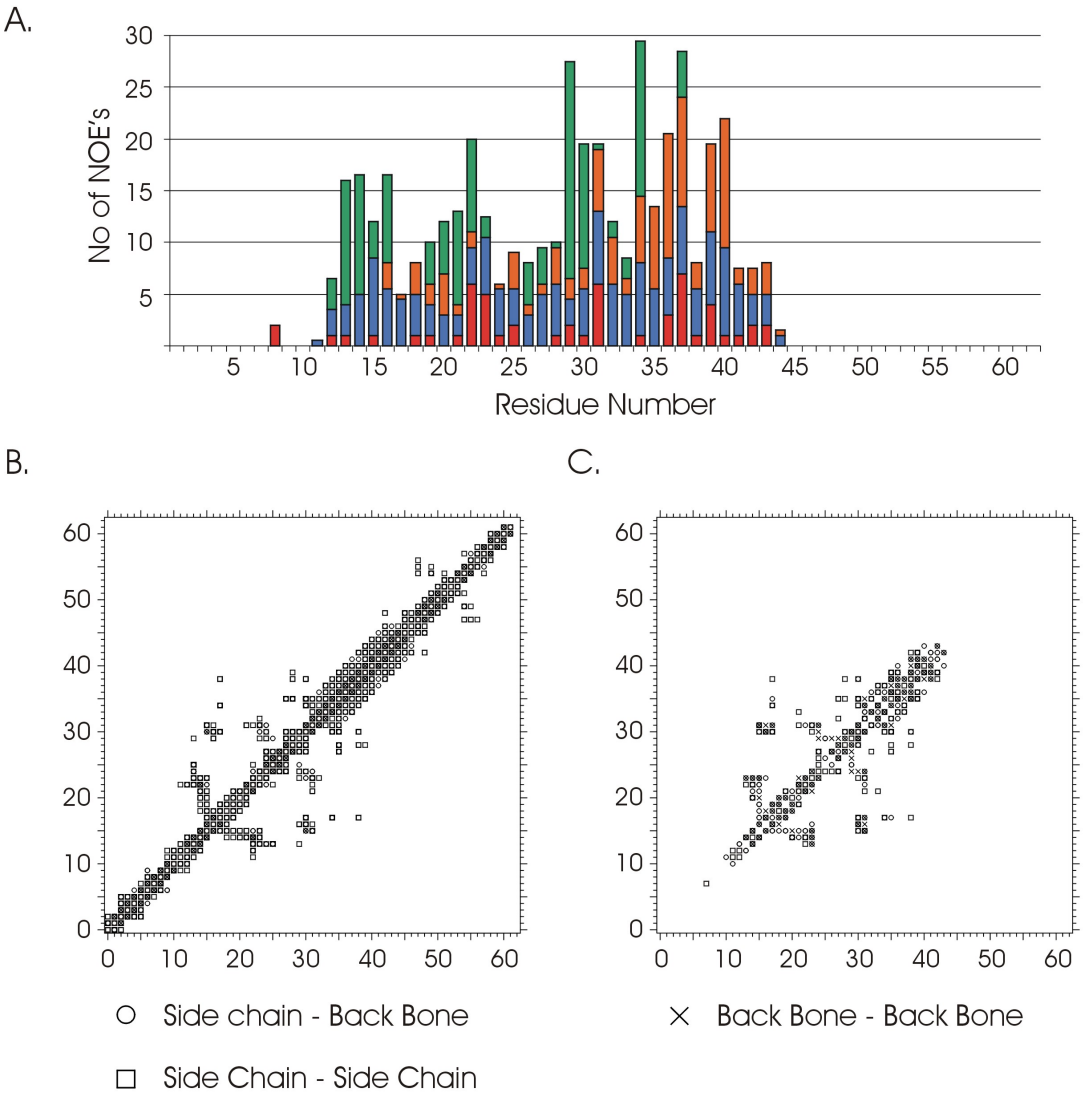



**Figure 3-14: (A) Number of conformationally restricting distance restraints including intra-residue (Red), sequential (Blue), medium distance (Orange) and long range (Green) on a per residue basis. (B) Contact map determined from the structural ensemble (B) and from the distance restraints. All were determined using PSVS[103].**

### 3.3.3.2 Evidence for the Accuracy of the Structure

The lack of significant NOE and dihedral angle violations and good agreement between the restraint and structure-determined contact maps (Figure 3-14), is a strong indicator of a structure which is consistent with the data from which it was created. The determined structure also shows an excellent fit to the recorded RDC data used to guide the structure generation (Figure 3-15), giving high confidence in the quality of the structure as a fit to the available data.

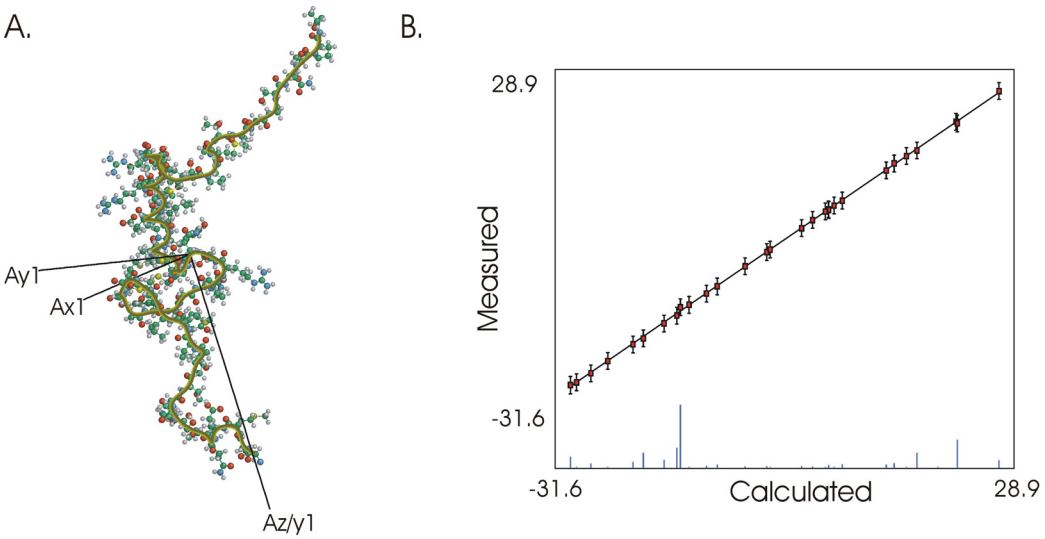



**Figure 3-15: (A) Fit of the Znf_A20 to the recorded RDC data using Pales[102]. (B), Fit of the measured RDC against those calculated for the determined alignment tensor and structure.**

One of the true strengths of NMR lies in the sensitivity of the chemical shift to the environment of the given nucleus. In proteins the different structural folds and variety in sequence gives relatively large chemical shift dispersion for what is essentially a repeating unit. Efforts to accurately predict the chemical shifts of a protein from a solved structure or vice versa have been the aim of several groups over the years. The group of Wishart has had good success with both determining structures from chemical shifts[98] and calculating chemical shifts from a solved structure[109]. The program SHIFTX has shown a good correlation between predicted and measured chemical shifts in test data sets [109] opening the prospect of validating structures based on chemical shifts. A plot of the observed chemical shifts for the Znf_A20 to those predicted using SHIFTX from the structure presented here gives a correlation with a gradient close to 1 in all cases. However the quality of the correlation varies greatly (Figure 3-16). The correlation is excellent for both $^{13}C\alpha$ and $^{13}C\beta$ shifts with $R^2$ values >0.9. While this is considerably worse for other atom types it is within the ranges for other NMR structures reported by Neal *et al*[109].

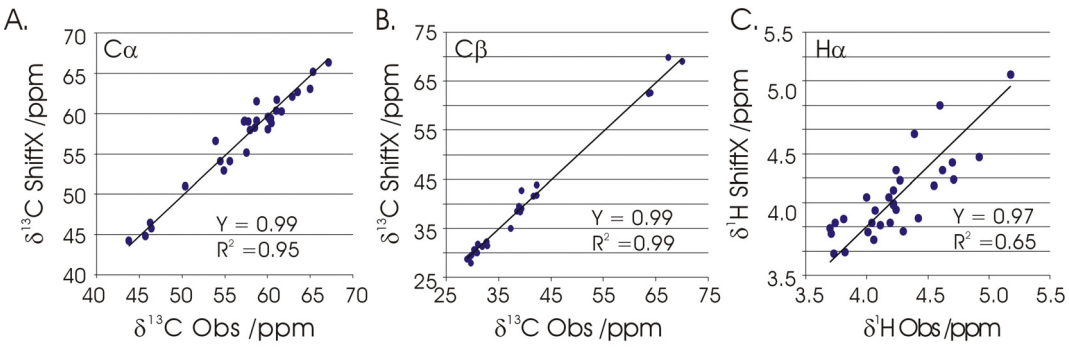



**Figure 3-16: Plots of observed chemical shifts against values predicted for the solved Znf_A20 structure using SHIFTX for (A) Cα. (B) Cβ. (C) Hα.**

The observed CH-π interaction between the Hβ and γ methyl protons of Thr26 in the loop with the aromatic ring of Tyr34 is also strongly supported by the chemical shifts which are some of those best predicted by SHIFTX. Both the Hβ and Hγ are shifted up field by around 1 ppm compared to average and random coil values for Thr Hβ and Hγ[136,137], a trend which is repeated for the majority of the resonances for Thr26 (Table 3-3). This is consistent with the close proximity of those atoms to the face of the Tyr ring, and the subsequent up field shift caused by the aromatic-ring magnetic-anisotropy.

**Table 3-3: Up field shifted resonances for Thr26 compared with those predicted for the structure[109], average BMRB values[136], and random coil (RC) values[109,137]. Δδ in all cases are compared to the measured δ.**

| Atom | δ Measured | δ SHIFTX | δ Average(±) | δ (RC) | Δδ SHIFTX | Δδ Average | Δδ RC |
|---|---|---|---|---|---|---|---|
| N | 110.92 | 108.4 | 115.45 (4.74) | 113.6 | -2.48 | -4.53 | -2.68 |
| C | 175.3 | 174.5 | 174.56 (1.76) | 175.2 | -0.77 | 0.74 | 0.1 |
| Cα | 58.79 | 61.5 | 62.23 (2.62) | 63.1 | 2.76 | -3.44 | -4.31 |
| Cβ | 67.12 | 69.8 | 69.72 (1.68) | 68.1 | 2.72 | -2.6 | -0.98 |
| Cγ | 20.3 | - | 21.56 (1.1) | - | - | -1.26 | - |
| $H^N$ | 7.54 | 6.83 | 8.26 (0.62) | 8.15 | -0.71 | -0.72 | -0.61 |
| Hα | 3.74 | 3.93 | 4.46 (0.48) | 4.35 | 0.19 | -0.72 | -0.61 |
| Hβ | 3.18 | 3.19 | 4.17 (0.33) | 4.43 | 0.01 | -0.99 | -1.25 |
| Hγ* | 0.48 | 0.87 | 1.14 (0.23) | 1.2 | 0.39 | -0.66 | -0.72 |

A study by Kornhaber *et al* in 2006[138] demonstrated that zinc-ligated cysteine residues could be distinguished with reasonable accuracy from non zinc-ligated reduced or oxidised cysteines using NMR chemical shifts. Although the chemical shifts for the Cα and Cβ overlap significantly for all cysteine types, a principal component analysis of Cα and Cβ shifts is sufficiently grouped to allow accurate determination of the cysteine state. This is improved further if the secondary structure surrounding the residue is known[138]. For the determined secondary structure and calculated chemical shifts all four cysteine residues have a 90 % probability of being zinc ligated[138].

#### 3.3.3.3 The N and C Terminal Portions are Highly Dynamic

The NMR structure presented here shows very low backbone RMSD across residues Met12-Asn43 but poor definition at both the N and C termini. This could indicate that those regions are intrinsically disordered or highly dynamic, however more accurately they should be described as regions which lack restraints during modelling. However experimental evidence suggests that this is not simply an artefact of modelling and that these regions do indeed represent highly dynamic regions.

The first is the relatively high Stoke's radius calculated for the Znf_A20 by gel filtration, highly dynamic termini would give rise to a higher Stoke's radius than for a globular protein of similar size. The program HYDROPRO[139] (Version 7 C) was used to estimate the hydrodynamic properties of the Znf_A20 from the proposed structure including Stoke's radius and sedimentation coefficient. A translational Stoke's radius of around 19 ± 3 Å was determined from the Znf_A20 structure, this is actually larger than the 16 ± 1 Å calculated by gel filtration supporting the notion that the apparent high molecular weight can be attributed to a high Stoke's radius but that the solved structure perhaps under-represents the degree of compactness in the protein's termini.

Similarly the solvent exposed surface area for the calculated structure is 6385 $Å^2$ 40% larger than that calculated from the average charge-state observed in ESI-MS. This could suggest, as with the measured Stoke's radius, that the determined structure over exaggerates the degree of disorder in the termini and that there is residual structure within these regions. However the calibration curve from Kaltashov *et al*[46] which was used for the estimation of solvent exposed surface area

contained only globular proteins and disordered regions may well behave differently. Although the solvent exposed surface area is thought to be the major contributor to the charge of the observed ions in the ESI-MS, other factors can play a role including buffer and the number of high affinity proton sites[46,140]. It is also worth noting that during the ionisation process the protein experiences a highly charged droplet environment as the droplet reaches the Rayleigh limit[46]. This change in environment may result in chain condensation at the termini which contain a number of hydrophobic residues including three Met, two Ala, four Pro, and four Thr residues. In the high ionic strength of the droplet one can imagine an increased clustering of these hydrophobic side chains resulting in a reduction in the apparent solvent exposed surface area, this could also be facilitated by the protonation of acidic sites under the high proton conditions[46,140]. Such structural collapse may be exaggerated in disordered regions where there is no ordered structure present to prevent it; however this would need to be investigated further.

On the other hand HYDROPRO calculated a sedimentation coefficient of 0.9 ± 0.1 S, this is in good agreement with the AUC determined sedimentation coefficient of 1.05 ± 0.5 S supporting the extended nature observed in the structure.

Evidence for the disordered nature of the terminal regions can also be provided by a simple examination of the chemical shifts; deviation of $^{1}H\alpha$ chemical shifts from a set of reference values for random coil (RC) proteins has long been used as indicators of secondary structure[141]. In later years this chemical shift deviation (CSD) was expanded to include the chemical shifts of all the backbone carbon, nitrogen and proton atoms using the so called chemical shift index (CSI)[98,142]. However this can be used both ways and more recently this same approach has been used to predict regions of protein which are random-coil-like using the so called random coil index (RCI)[143,144]. The program PREDITOR, which was used for generating dihedral angle restraints, also analyses chemical shifts by using RCI as an indicator of the reliability of its dihedral angle predictions[83,143,144]. The RCI for the Znf_A20 is in very good agreement with the backbone atom mean deviation from the average for the solved structure ensemble, giving low values of RCI (<0.1) from residue Leu13 to Asn43, which gradually increase towards the termini (Figure 3-17).

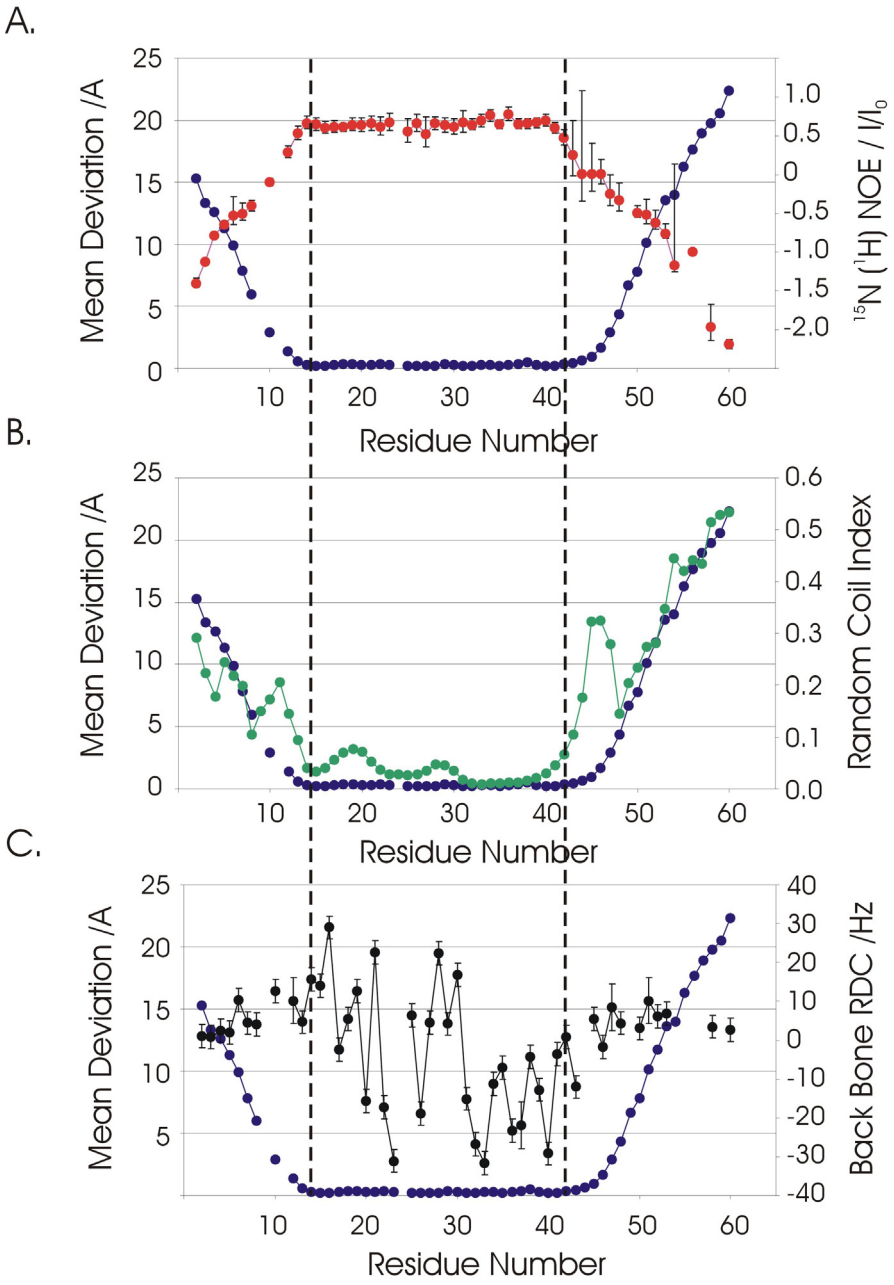



**Figure 3-17: Local mean deviation from the average structure (Blue) compared to: (A) The $^{15}$N($^{1}$H) NOE (Red). (B) The chemical shift derived RCI (Green). (C) The recorded backbone RDC values (Black). The domain boundaries of Met12 to Asn43 are illustrated with a dashed line.**

The $^{15}$N($^{1}$H) NOE is also a good indicator of local dynamics. Essentially a $^{15}$N($^{1}$H) NOE ratio close to 1 suggests that the correlation time ($\tau_c$) of the NH bond vector is close to the $\tau_c$ of the molecule as a whole and therefore will undergo little or no isolated motions. Conversely very low values of $^{15}$N($^{1}$H) NOE (<0.4) suggest that the $\tau_c$ of the NH bond vector is much shorter than the overall molecular $\tau_c$ as would

be expected for highly dynamic regions. The Znf_A20 shows very high values of $^{15}N(^{1}H)$ NOE between Leu13 and Asn43 with a maximum and average value of 0.79 and 0.65. However like the RCI, the values of $^{15}N(^{1}H)$ NOE drop off sharply outside of this range, turning large and negative within 3 residues of the domain boundaries indicated by the structure, with an average value outside of the ordered region of -0.63.

As alluded to previously the flexibility within the polypeptide chain may result in motional averaging of RDCs causing a reduction in the recorded value[134,135]. While this can lead to inaccuracies when using RDCs in structure generation protocols (where motional averaging is difficult to account for), it can be a useful indicator of dynamics. The RDC values for the Znf_A20 are relatively large, giving a max and average magnitude in 7% PAG of 31.6 Hz and 14.6 Hz over the ordered region (Figure 3-17). Such large RDC values, for what is essentially a small protein, suggest a high degree of partial alignment in the PAG. This could be explained by the relatively large Stoke's radius determined by SEC combined with the low motional averaging suggested by the structure and by the large values of $^{15}N(^{1}H)$NOE. As seen with the values of $^{15}N(^{1}H)$ NOE the magnitude of the RDC is seen to visibly reduce towards the termini (Figure 3-17). This is not quite as pronounced as the changes observed for the values of $^{15}N(^{1}H)$ NOE but this is to be expected as simulations have shown that the small restrictions in the conformational properties of disordered polypeptide-chains is sufficient to record a small RDC value[135].

### 3.3.4 Evidence for *cis-trans* Isomerisation

During the assignment it became clear that the number of correlations seen in the $^{1}H$-$^{15}N$-HSQC is greater than that expected for the number of amino acids in the protein's primary sequence. This could be caused by a number of things; degradation of a protein during the period of an NMR experiment can lead to the build up of extra peaks in the $^{1}H$-$^{15}N$-HSQC originating from the NH groups of the fragments of that breakdown; these peaks are usually very narrow and sharp due to their very rapid $\tau_c$ and have little chemical shift dispersion. This may well explain some of the small very sharp peaks seen towards the bottom half of the Znf_A20 $^{1}H$-$^{15}N$-HSQC close to the peak for Ser15 in Figure 3-8. However this is unlikely to explain all of these

extra peaks some of which have good chemical shift dispersion and line shapes consistent with something larger than a fragment.

The remaining peaks may well be the result of conformational exchange, in principle the observed structure at equilibrium represents that which possess the lowest free energy under the conditions at which the measurements were made. However such a static image of a complex molecule such as a protein is inadequate and in reality the structure can be considered as sampling an ensemble of different conformations which all possess a similar free energy close to the lowest energy state. If these alternative conformations are sufficiently close in energy to the lowest energy state (< 10 kJ/mol) they will become significantly populated at equilibrium. The population of alternative states will be reflected in the NMR spectrum in the form of extra peaks or subtle changes in chemical shift or resonance line-shapes depending on the rate constant of the equilibrium[145].

Though there is no clear evidence for chemical exchange correlations in NOESY or TOCSY spectra of the Znf_A20, the extra peaks do seem to share common through bond correlations with assigned peaks from the main species. This is consistent with chemical exchange where the backbone assignment 'path' temporarily splits into two 'paths' through residues which occupy different conformations. While this may be the result of a population of Znf_A20 which has lost its zinc atom resulting in unfolding, no evidence was seen for apo-Znf_A20 in the ESI-MS recorded prior to the NMR experiment and the affected residues are also inconsistent with this suggestion.

Alternative conformations of proteins can arise as a result of *cis-trans* isomerisation of the ω bond of proline residues. The overwhelming majority of amino acids exist in the *trans* isomer (>99.9%) under normal conditions[146]. This is primarily because the amide hydrogen offers less steric repulsion to the preceding Cα atom than to the following Cα atom[146]. The *cis* and *trans* isomers of the X-Pro peptide bond (where X represents any amino acid) both experience steric clashes with the neighbouring substitution and the energetic difference between the two is dramatically reduced[146]. In free unrestrained proline residues populations of *cis*-proline between 10-40 % have been reported[146]. The extra peaks in the $^{1}H$-$^{15}N$-HSQC of the Znf_A20 could be explained by the presence of *cis-trans* isomerisation and this would explain the lack of chemical exchange peaks in the NMR spectra as

*cis-trans* proline isomerisation may be too slow for this. There is a variety of other evidence to support this notion.

Based on inter-residue through-bond correlations the extra $^{1}$H-$^{15}$N-HSQC correlations (that do not resemble degradation fragments) were assigned as alternative resonances belonging to residues Thr5, Thr8, Gly10, Met12, Leu13, Tyr21, Gly22, Asn23, Arg25, Ser58 and Thr60. Though not all of these residues are adjacent to Proline residues, when these residues are mapped onto the structure of the (all trans), Znf_A20 there is a clear correlation between the positions of the identified residues and the position of proline residues (Figure 3-18). The only example which is not obviously close to Pro is Thr5 which sits in the N-terminal flexible tail. This is reasonably close to Pro9 and given the sensitivity of the chemical shift it is feasible for the isomerisation of Pro9 to be affecting Thr5, possibly through an element of transient structure.

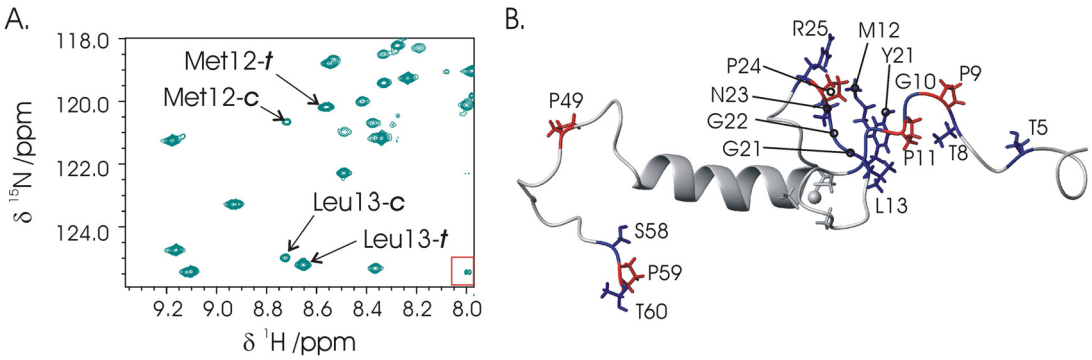



**Figure 3-18: (A) Section of $^{1}$H-$^{15}$N-HSQC residues showing peaks resulting from nearby *trans* (t) or *cis* (c) proline residues. (B) Proximity of residues which show two species by NMR (Blue) and the position of prolines (Red) plotted onto the structure of the Znf_A20.**

Of the assigned extra peaks a number of them follow proline, this allows the determination of the Cβ chemical shifts of the adjacent proline using the CACB(CO)NH experiment. The chemical shift of the Cβ of proline residues in the main species are centred around 32 ppm consistent with *trans*-proline[146], on the other hand in the second species the Cβ shifts is centred on 34 ppm consistent with *cis*-proline[146], which adds particularly strong evidence for the notion of *cis-trans* proline isomerisation in the Znf_A20 domain.

The assignment of the residues of the *cis*-proline species allows the estimation of the population of each *cis*-proline using the relative NMR intensities of

different species of the same residue. The average population is calculated to be around 16 ± 6 % although many of the extra peaks significantly overlap with their corresponding peak from the main species giving large errors. A value of 16 ± 6 %, represents a $\Delta G_{298}$ of around 4.5 ± 1 kJ/mol consistent with previously reported values of 3.9 kJ/mol for unrestrained proline residues[147]. The only proline residue which resides in a well defined region of the protein spectra is Pro24, while there is evidence of the *cis*-isomer of Pro24 from both Arg25 and the side chain resonances of Asn23, in both cases the peaks are particularly overlapped precluding an accurate population determination. However the population of *cis*-Pro24 does not appear to be significantly different to those of the other potential prolines despite being conformationally restricted by the protein fold.

It is worth noting at this point that there is no suggestion that these residues are cooperatively isomerising to the *cis* isomer. The suggestion is simply that each of the prolines are, in isolation from the other proline residues, in equilibrium between the *cis* and *trans* isomer creating alternative local conformations which are observed by NMR.

The structural impact of these observed *cis*-prolines does not appear to extend beyond their immediate vicinity, although there is insufficient information to warrant a full structural investigation incorporating *cis*-prolines at different positions. The chemical shifts of the majority of residues in the Znf_A20 are seemingly un-changed by the isomerisation. Those which do show CSP show only small changes with the average being around 0.2, with those in the ordered region showing even smaller CSP closer to 0.08. This is supported by the absence of large changes in the recorded RDC values between the main and alternative species showing a maximum change in RDC value of 3 Hz, similarly there is very little change in the values of $^{15}N(^{1}H)$ NOE with an average change of only -0.07. This further suggests that these local conformational changes do not significantly change the global structure or dynamics of the Znf_A20 domain.

## 3.4 Conclusions

ZNF216 residues 1-60 has been expressed and purified to a high purity. Biophysical measurements utilising SEC, AUC, ITC, ESI-MS, and CD have revealed that the protein is monomeric in solution with a relatively extended conformation and has confirmed the presence of a zinc finger domain which requires the presence of a

single zinc ion to maintain its structure. The solution structure of this zinc finger domain has been solved by standard NMR techniques using the XPLOR-NIH protocol; this confirmed the presence of an A20-type zinc finger domain running from residues Met12 to Asn43 as previously proposed by bioinformatics approaches.

The solved structure for the Znf_A20 domain is in good agreement with previously solved structures suggesting a high degree of structural conservation across the domain family. Sequence alignments have demonstrated that the Znf_A20 domains of TNFaIP3 show a lower sequence similarity to Znf_A20 domains found in other proteins. All of the available structures of Znf_A20 domains in the absence of a binding partner originate from TNFaIP3, the structure presented here therefore represents the first solution structure of a Znf_A20 domain which shows high sequence identity with the majority of the domain family. Despite the difference in sequence the structural conservation remains high, although those Znf_A20 from TNFaIP3 include an extra residue in the loop between the first two cysteines which reduces the compactness of the fold around the first loop.

Several features of the Znf_A20 primary sequence have common characteristics; the conservation of the four cysteines is absolutely required for the fold of the domain due to the requirement of the structural zinc atom for correct folding, however the structure also gives potential explanations for other features of the primary sequence. The sequence contains three common glycines, in ZNF216 these are glycine 17, 19 and 22, in the structure presented here and in the crystal structure of Rabex-5, Gly17 and Gly19 (or corresponding positions) possess phi and psi angles which would be unfavourable for other residues, occupying regions of Ramachandran space favoured by Glycines only. In the case of Gly17, the adopted dihedral angles allow it to form a hydrogen bond with Cys14 adding to the overall stability of the protein. Gly22 on the other hand represents one of the few fully buried residues and it may well be the case that the lack of a Cβ at this position aids the correct fold of the domain. The small size and flexibility of glycine may explain the conservation at these positions where the unique properties of glycine allow the domain to form its compact core and aid the positioning of the four cysteines for stable zinc coordination. This may be contradicted by those domains which have replaced these glycines with larger side chains, however this could be explained by the insertion of an extra residue in the loop between the first two cysteines which could allow the core to adapt to compensate for more bulky side chains.

Several aromatic and hydrophobic residues also show a degree of conservation including the aromatic pair of Phe20 and Tyr21 which along with Met12 and Leu14 form a hydrophobic 'strip' which, from a structural point of view is not obviously rationalised, however the conservation of the aromatic residue at Tyr34 and the hydrophobic chain at Met29 could be rationalised by key stabilising roles. The contacts made between Tyr34 and Thr26 'docks' the long loop of the zinc knuckle against the helix, while Met29 is central to the hydrophobic core formed between the zinc knuckle and the helix, unlike some of the other features these, particularly the hydrophobic chain at Met29, are well conserved across all Znf_A20 domains, supporting the notion of them being key to the fold.

Understanding the structure of a protein is key to explaining the molecular detail behind its function. The ZNF216 protein has been implicated as a shuttle protein, facilitating transport of ubiquitinated proteins in muscle to the proteasome for degradation during muscle atrophy. ZNF216 has also demonstrated an ability to modulate NF-κB activity indirectly via RIP and TRAF6. These two important cellular process are dependent on both of the zinc finger domains of ZNF216[113,117] but the Znf_A20 domain and its proposed ability to bind to Ub appear to be vital[113,117]. The structure of the ZNF216 Znf_A20 domain presented here represents a step towards a full understanding of the molecular detail of ZNF216 function in order to better understand its roles in muscle atrophy and NF-κB regulation.

# Chapter 4

# 4 Investigating the Interaction of the Znf_A20 of ZNF216 with Ubiquitin

## 4.1 Introduction

### 4.1.1 The Znf_A20 Interaction with Ub

The interaction between Ub and Znf_A20 domains has been reported in a number of studies, many of which are associated with regulation of NF-κB including that of AWP1, TEX27, Rabex-5, and TNFaIP3. Interestingly it has been reported that the Znf_A20 of both ZNF216 and AWP1 bind poly-Ub chains, while little or no poly Ub binding could be detected for the Znf_A20 domains of Rabex-5, TNFaIP3 or TEX27[113,123]. This adds additional ambiguity to the function of Znf_A20 domains and the possibility that different Znf_A20 domains, like other UBDs, may show different poly-Ub chain linkage specificities[2,148]. It is not yet known whether all Znf_A20 domains interact with Ub, indeed of the 7 Znf_A20 domains of TNFiAP3 only the fourth Znf_A20 domain is thought to bind Ub[32] and of the large family of SAPs, which possess both Znf_A20 and Znf_AN1, no Ub interaction has yet been reported although links with the UPS have been discussed[120,126,149].

The role of ZNF216 as a proteasomal shuttling protein is dependent on the interaction of its Znf_A20 domain with Ub. This interaction would be central to the proposed recognition of ubiquitinated muscle proteins and their subsequent transportation to the proteasome for degradation, an event which is thought to be central to the atrophy process[113]. It is also possible that the proposed interactions of the Znf_A20 with RIP and IKKγ may be mediated via Ub as both of these proteins are ubiquitinated *in vivo*[150,151].

### 4.1.2 Rabex-5 Znf_A20 Interaction with Ub

The protein Rabex-5 is a Znf_A20-containing exchange factor for the protein Rab5 and functions as a regulator of endosomal trafficking[20,21]. The interaction of its Znf_A20 domain with Ub is key to its function as it is required for its ligase

activity[20,21]. A crystal structure of Rabex-5 in complex with Ub revealed that the Znf_A20 domain of Rabex-5 is a fusion with a MIU domain, another UBD, and that the Znf_A20-MIU fusion can bind two molecules of Ub, one via its MIU binding the hydrophobic β-sheet binding surface of Ub centred on Ile44, and a second bound to the Znf_A20 domain using a novel Asp58 centred binding surface[20,21]. The Asp58 binding site, identified in the Rabex-5-Znf_A20 Ub crystal structure, comprises of a variety of polar and hydrophobic contacts centred on the formation of two hydrogen bonds involving the two oxygen atoms of the carboxylate side chain of Ub-Asp58 with the side chain alcohol and backbone NH of a key Ser residue at the base of the Znf_A20 helix and the highly conserved di-aromatic motif of the Znf_A20 domain (Section 3.2.1) which make key contacts with the side chains of Ub-Arg54 and Asn60. Other key Ub residues contributing to binding include Thr55, Ser57 and Tyr59[20,21]. The Rabex-5-Znf_A20 Ub interaction is reported to be relatively high affinity for a UBD interaction with Ub, with a $K_d$ of around 20 μM[21].

This new Ub binding surface represents one of a range of new binding sites located on Ub other than the classically utilised Ile44 hydrophobic patch. In addition to the Asp58 binding patch, new interaction surfaces have recently been identified centred on Leu8 and the C-terminal di-glycine[2,16,17].

### 4.1.3 Summary and Aims

The interaction of the Znf_A20 domain with Ub is central to the function of ZNF216 and warrants molecular insight into the nature of its interaction. Moreover the general interaction of Znf_A20 domains with Ub has a number of questions surrounding it, the first being, that given the presence of only a single crystal structure of such a complex in the RPDB, and the unusual context of the Znf_A20 involved (as an MIU fusion), does the observed binding interface hold true for other Znf_A20 domains? Or is the site occupied by Rabex-5 perturbed by the presence of a second Ub molecule bound to the MIU domain? Is there variation between different Znf_A20 or between the complexes in solution compared to that observed in the crystal?

The second question is, do all Znf_A20 domains carry the characteristics required to bind to Ub? This may be difficult based on a single structure of a Znf_A20 complex with Ub. Predicting the features which facilitate the interaction

with Ub and assessing other Znf_A20 for the presence of those features would be aided by the presence of other Znf_A20 structures in complex with Ub.

Thirdly, given the reports that some Znf_A20 domains possess the ability to bind poly-Ub while others do not, and that Znf_A20 domains are linked to a variety of different cellular processes, is there poly-Ub linkage specificity within the Znf_A20 domain family which may be associated with the different cellular processes. If this is the case what are the structural features which determine the linkage specificity in some cases and prevent poly-Ub binding in others?

The final question, which again surrounds specificity, is given the similarity between Ub and related Ub-like (UBL) proteins, are Znf_A20 domains, like other UBDs, specific for Ub, and if so what structural features confer this selectivity.

Towards answering these questions and to improve the understanding of the ZNF216 interaction with Ub, the interaction has been characterised using biophysical techniques to determine the affinity constants and thermodynamics for the interaction. The interaction has been studied in molecular detail using NMR based approaches to generate a solution structure of the non-covalent complex for the Znf_A20 with Ub.

## 4.2 The Znf_A20 Domain of ZNF216 Forms a Non-Covalent Complex with Ub

The Znf_A20 domain of ZNF216 has been proposed to interact non-covalently with Ub, in order to confirm and characterise this interaction, membrane-based binding assays, isothermal titration calorimetry (ITC), and native electrospray ionisation mass spectrometry (ESI-MS) were performed.

### 4.2.1 Membrane-Based Binding Assays

Membrane-based binding assays were performed by the group of Dr Robert Layfield in the School of Biomedical Sciences. Ub and surface mutants designed to disrupt known binding patches, including I44A, D58A, and ΔG75G76 were spotted onto nitrocellulose and after blocking the membrane, used to capture recombinant wt-GST-ZNF216 protein or a C30A/C33A mutant. Binding was confirmed by immunodetection using anti-GST. Binding was detected between Ub and GST-ZNF216 in all cases with the exception of the Ub D58A mutant, however the C30A/C33A mutant was unable to bind to Ub under any of the conditions used

suggesting that the zinc ligation is a requirement for the interaction. The majority of Ub binding proteins recognise a hydrophobic patch centred on Ile44, the Rabex-5 Znf_A20 domain however recognises an alternative binding patch centred on Asp58. The attenuation of Znf_A20 binding by the D58A mutants but not the I44A mutant in the membrane-based binding assay presented here suggest that the ZNF216 Znf-A20 similarly does not bind to the classical hydrophobic patch, instead utilising the alternative Asp58 centred binding patch.

Membrane-based binding assays were also used to assess the specificity of the ZNF216 for poly Ub linkages including linear di-Ub (Ub-Ub), Lys48-linked tetra-Ub (K48-$Ub_4$), and Lys63-linked tetra Ub (K63-$Ub_4$) and for Ub versus the Ub like (UBL) proteins SUMO1 and NEDD8 (Figure 4-1 B).

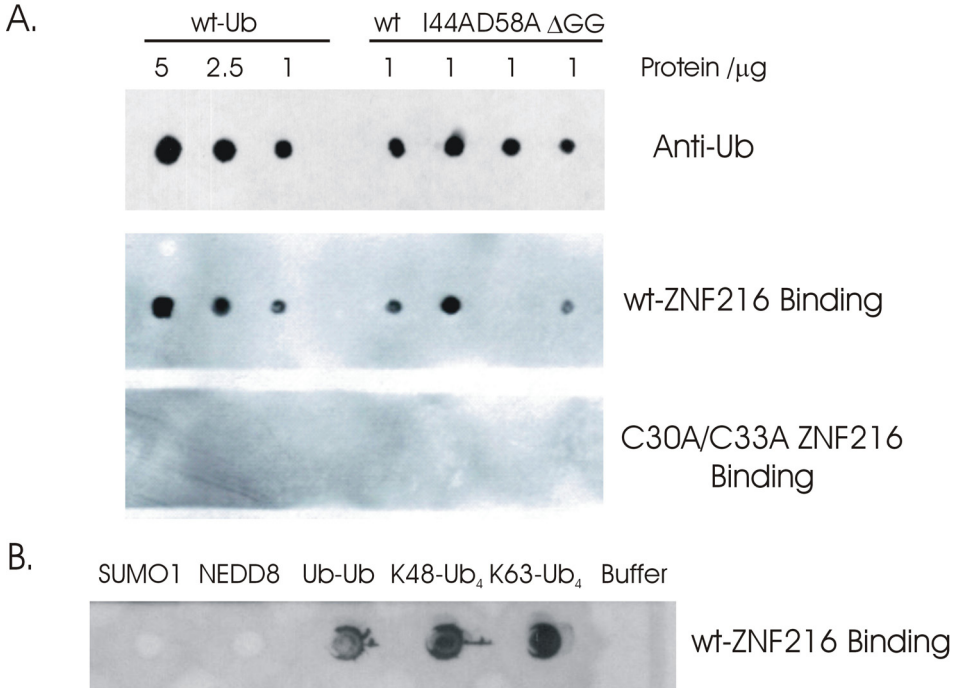



**Figure 4-1: Membrane-based binding assays of; (A) membrane layered Ub and Ub mutants with GST-ZNF216 and GST-ZNF216-C30A/C33A. (B) Examination of the specificity of the Ub interaction with ZNF216 for Ub chain linkages and UBL domains.**

The ZNF216 binds to the three linkage types of poly Ub used in membrane-based binding assays with no clear preference for the two 'classical' linkage types namely Lys48- and Lys63-linked poly Ub. Linear Ub appeared marginally weaker but this is within the error of the technique and is not significant.

The ZNF216 interaction with Ub does not appear to extend to the UBL proteins SUMO1 or NEDD8, which share structure and sequence characteristics with Ub[2,14] yet do not bind ZNF216 in membrane-based assays (Figure 4-1 B). Interestingly both proteins share relatively high sequence characteristics around the proposed binding site for the Znf_A20 with Asp58 being conserved in NEDD8 and mutated to a Glu in SUMO1. This suggests that although Asp58 is required for the interaction, it is not sufficient for the interaction alone, highlighting the need for a more detailed picture of the interaction surface.

### 4.2.2 Biophysical Characterisation

ITC was performed in order to characterise the thermodynamic properties of the proposed interaction between the Znf_A20 and Ub. Titrations were performed whereby Ub at high concentration (~300 μM) was titrated into a calorimetric cell containing Znf_A20 at around 20 μM and the resulting exothermic pulses used to calculate the thermodynamic parameters of binding (Section 2.3.2.2). This was repeated with Znf_A20 at 300 μM being titrated into Ub at 20 μM.

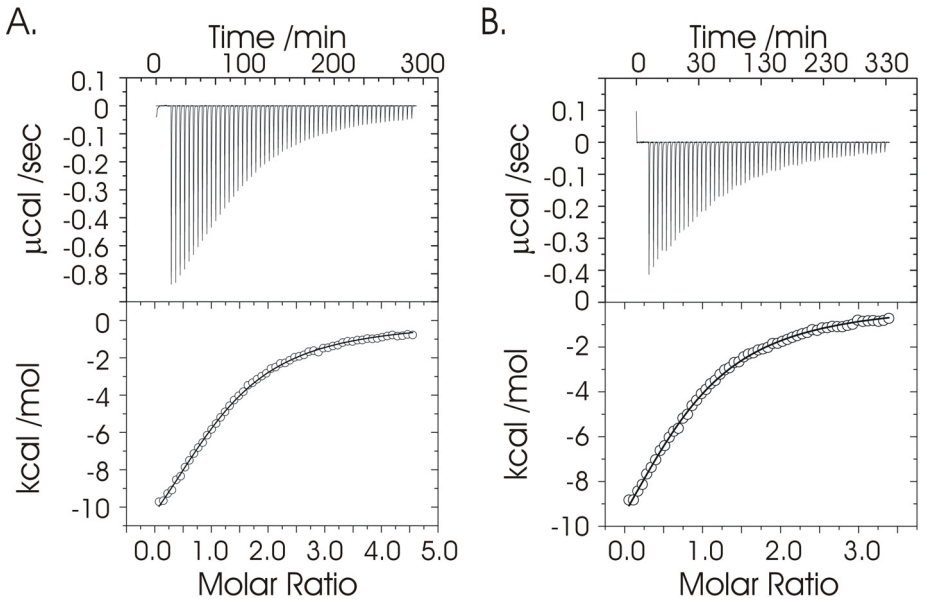



**Figure 4-2: ITC titrations of (A) Mono-Ub into Znf_A20. (B) Znf_A20 into mono-Ub.**

Both titrations show the expected exothermic heat pulses consistent with a binding event between the two proteins. The integrated heat responses was analysed using the origin software; the data was fit to a single site model to extract

stoichiometry, $K_d$, ΔH, ΔS, and ΔG. All titrations were within good agreement across all ITC experiments giving a stoichiometry of 0.9 ±0.11, a $K_d$ of 12 ± 3 μM, and an enthalpy change of -64 ± 5 kJ/mol consistent with the formation of a relatively high affinity 1:1 complex between the Znf_A20 and mono-Ub which is enthalpically driven (Figure 4-2). The thermodynamic parameters derived from ITC titrations can be seen in Table 4-1. ITC data for the Znf_A20 interaction with mono-Ub clearly suggests a one-to-one complex, in order to confirm this and the ITC observed $K_d$, native ESI-MS was performed. Although ESI-MS is recorded in the gas phase, 'soft' ionisation techniques such as ESI, are shown to give accurate reflections of solution phase populations with little or no gas phase dissociation of non-covalent complexes in favourable cases[45-47]. ESI-MS was performed on samples containing mono-Ub and Znf_A20 at a range of concentrations with the maximum protein concentration maintained below 6 μM.

**Table 4-1: Thermodynamic parameters derived from ITC titrations of Ub and Znf_A20.**

| Protein in the Syringe | Mono-Ub | Znf_A20 |
|---|---|---|
| Protein in the Cell | Znf_A20 | Mono-Ub |
| Stoichiometry | 1.03 (0.10) | 0.85 (0.21) |
| $K_a$ (±) /M | 84800 (6400) | 86900 (6500) |
| $K_d$ (±) /μM | 12 (1.7) | 11 (1.5) |
| ΔH (±) / kJ/mol | -63.9 (1.7) | -69.7 (5) |
| ΔS (±) / J/mol/K | -120 (6) | -142 (2) |
| ΔG (±) / kJ/mol | -28.1 (6) | -27.8 (5) |

While the observed mass peaks readily identify the binding stoichiometry, the relative intensities of the observed mass ions can be used to determine the binding affinity. In this case peaks were observed corresponding to free mono-Ub, free Znf_A20, and a 1:1 complex between Znf_A20 and mono-Ub. No evidence for higher order complexes or homo-oligomers was observed consistent with previous biophysical data suggesting that both proteins are monomeric in solution and form a one to one complex. The relative peak intensities in ESI-MS can be assumed to vary

linearly with the solution populations allowing determination of the concentrations of the various species observed in the mass spectrum (Section 2.3.3.2)[45]. The $K_d$ can be calculated from the gradient of a plot of [P:L]/[P] against [L]; where [P] represents the concentration of free protein (in this case Ub), [L] represents the concentration of free ligand (in this case Znf_A20) and [P:L] represents the concentration of the one to one complex of the two. For a titration of Znf_A20 with mono-Ub a $K_d$ of 6.8 ±1 µM, consistent with that seen by ITC.

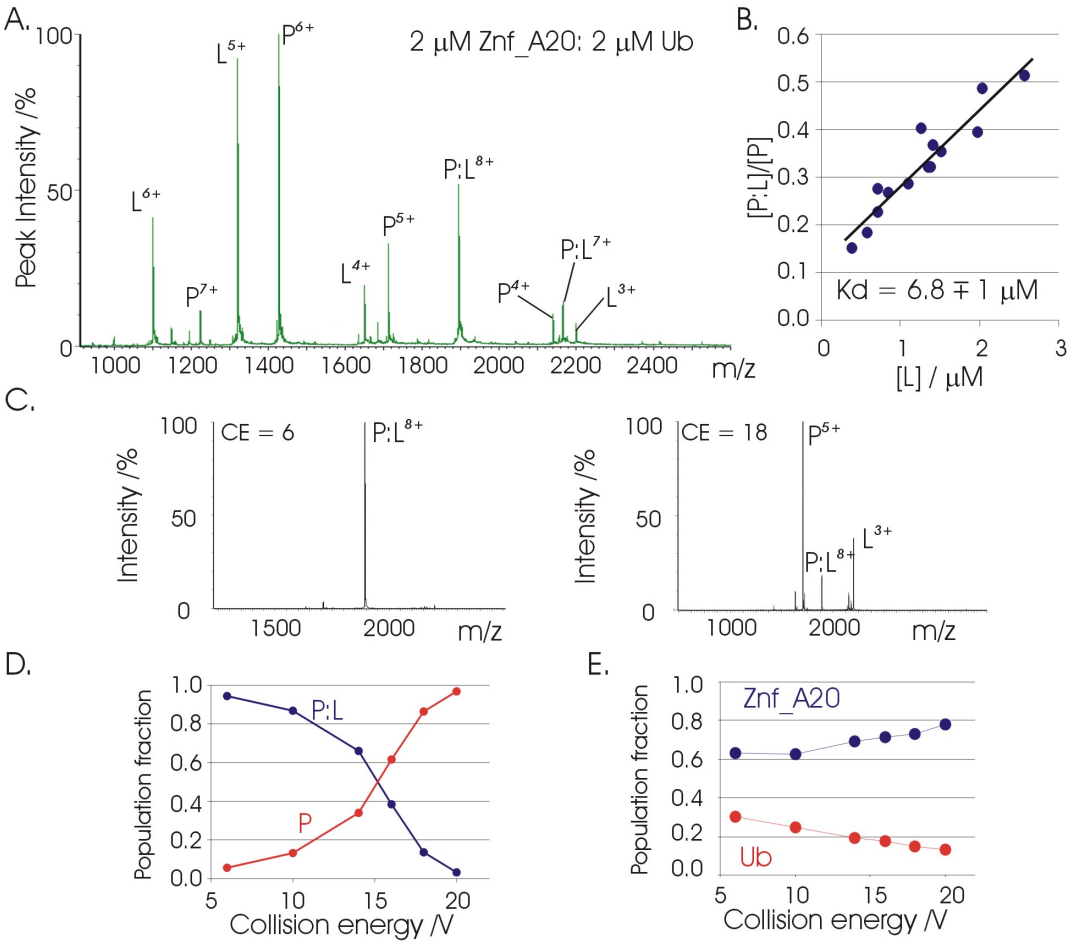



**Figure 4-3: (A) ESI-MS of 2 µM Znf_A20 mixed with 2 µM mono-Ub. (B) Plots of [P:L/P] Vs. [L] derived from the ESI-MS titration of Znf_A20 and mono-Ub. (C) CID of the Znf_A20-Ub complex 8+ ion. (D) The observed populations as a function of transfer energy for the CID of the complex 8+ ion. (E) Distribution of the zinc ion in CID fragments (E).**

In order to confirm the nature of the presumed Znf_A20 complex with Ub mass ion, collision induced dissociation (CID) was performed. Mass ions corresponding to the proposed complex 8+ ion were isolated using a quadrupole and

accelerated through a transfer cell containing argon gas in order to produce high-energy collisions between mass ions and Argon. The acceleration voltage, and therefore the kinetic energy of the collisions with argon, was gradually increased until dissociation of the complex ion was observed (Figure 4-3). The complex mass ion dissociated into the expected components of Znf_A20 and Ub confirming the composition of the complex mass ion.

Interestingly the dissociation of the complex ion was split further into a subset of fragments corresponding to different partitioning of the zinc ion in the two fragments; in most cases the zinc ion is associated with the Znf_A20 ion, however around 15-30 % of the zinc is found associated with the Ub fragment (Figure 4-3). Although it can not be assumed that the structure of complex is the same in the gas phase as in solution, the appearance of zinc ions associated with Ub would suggest a close contact between Ub and the Znf_A20-associated zinc ion leading to association of the zinc with Ub during CID.

As discussed previously, the average charge-state observed for mass ions in the ESI-MS correlates well with the solvent exposed surface area. In principle this can be used to estimate the surface area buried by the binding of the two proteins by a simple calculation of the surface areas of the two free proteins and the complex based on the average charge-state. The average charge-states for mono-Ub, Znf_A20, and their complex are +5.76, +5.01, and +7.8 corresponding to predicted solvent exposed surface areas of 4574 ± 150, 3765 ± 150 and 7215 ± 200 $Å^2$, giving an estimated buried surface are of 1124.6 ± 300 $Å^2$. It is worth noting that the calculated solvent exposed surface area for the free Znf_A20 is considerable smaller than that calculated from its structure, and that many factors can potentially influence the average charge-state giving rise to large errors. The correlation of average charge-state as a function of surface area is still under discussion in the literature[46,140].

The lack of poly-Ub chain-linkage specificity observed in membrane-based binding assays was investigated using ESI-MS. Lys48- and Lys63-linked di-Ub was purchased from Enzo life sciences and titrated at various concentrations with Znf_A20 with the total protein concentration maintained below 6 μM to prevent non-specific binding. Titrations involving both Lys48- and Lys63-linked Di-Ub showed the presence of both 1:1 and 2:1 Znf_A20:di-Ub complexes (Figure 4-4). The low concentrations of the purchased di-Ub were too low to measure an absorbance using the NanoDrop making them difficult to accurately quantify, therefore Di-Ub

concentrations were estimated indirectly by recording ESI-MS on mixtures of di-Ub and known concentrations of mono Ub or Znf_A20 and the concentrations calculated based on their relative intensities in the ESI-MS spectrum. This resulted in large errors in the concentration measurements (up to 30 %) and large corresponding errors in the derived equilibrium constants. However it is worth noting that although the error values for the equilibrium constants appear large, the differences in terms of the free energy of binding remain small (<1 kJ/mol)(Table 4-2).

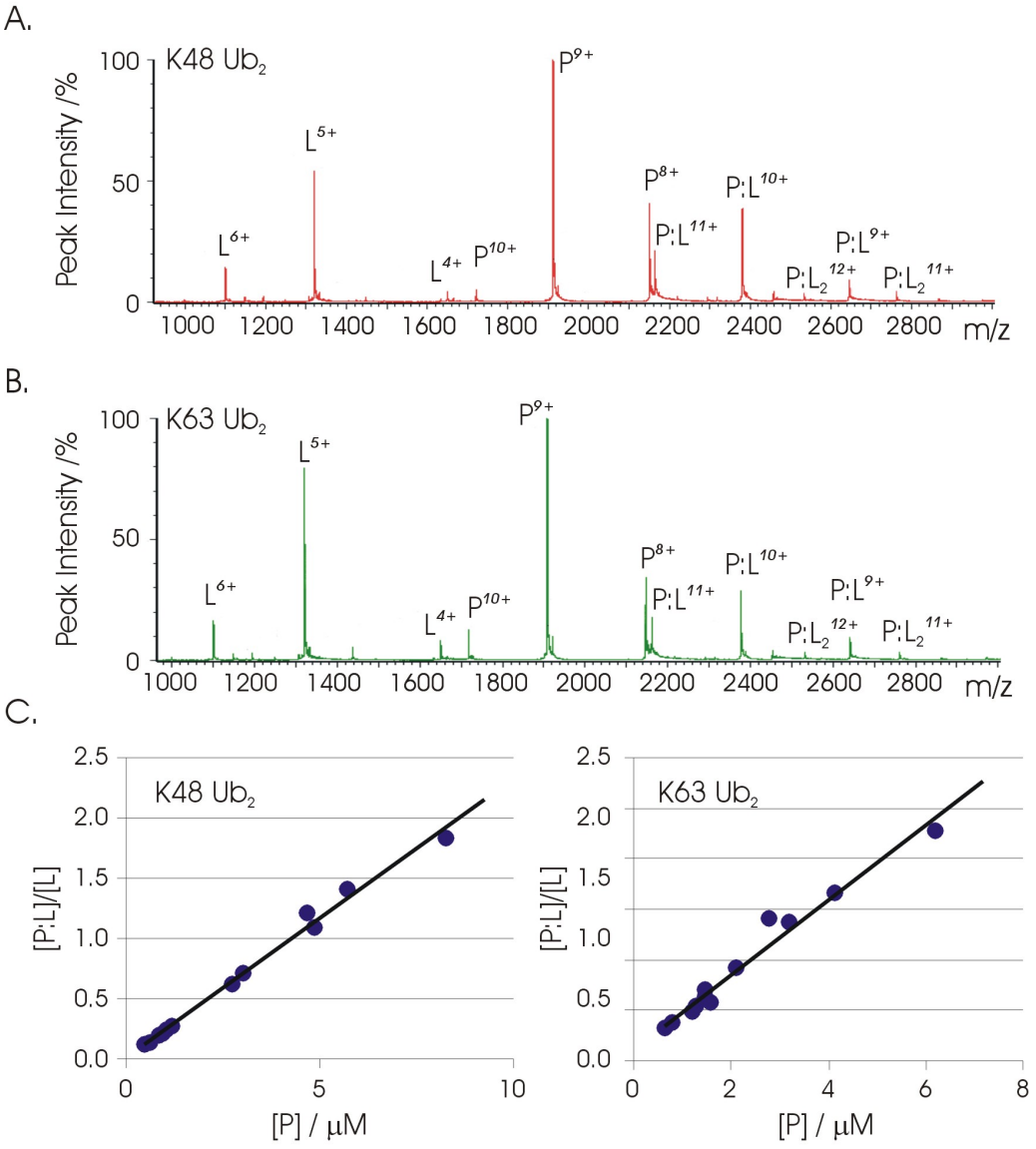



**Figure 4-4: (A-B) ESI-MS of 2 µm Znf_A20 and 2 µM K48-Ub2 or K63-Ub2. (C) Plots of [PL]/[L] Vs. [P] for K48-Ub2 and K63-Ub2. Where [P] and [L] represent the concentrations of the free binding sites on di-Ub and Znf_A20 respectively and [PL] equals the concentration of bound Znf_A20.**

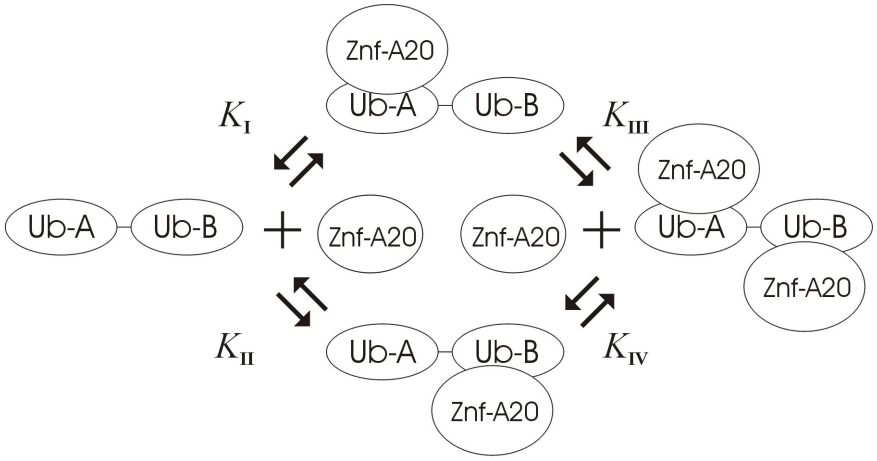



**Figure 4-5: Model of the binding equilibrium of Znf_A20 to $Ub_2$.**

**Equation 4-1**[47]

**Equation 4-1.1** $\boldsymbol{Ka_1 = [PL]/[P][L]}$

**Equation 4-1.2** $\boldsymbol{Ka_2 = [PL_2]/[PL][L]}$

**Equation 4-1.3** $\boldsymbol{Ka_1 = K_I + K_{II}}$

**Equation 4-1.4** $\boldsymbol{1/Ka_2 = (1/K_{III}) + (1/K_{IV})}$

Where [P] represents the concentration of free di-Ub, [L] the concentration of free Znf_A20, [PL] the concentration of their 1:1 complex, and [$PL_2$] the concentration of their 1:2 complex[47].

The binding of Znf_A20 to di-Ub can be described by the equilibrium shown in Figure 4-5, the measured populations in the ESI-MS can not distinguish which site is occupied and therefore the measured equilibrium constants are described as products of the equilibrium constants that make them described by Equation 2-15. If the two sites are considered to be equal an apparent K for the two sites can be calculated based on the concentration of available sites. In this case the equilibrium constant K' is the same as that in Equation 4-1.1 where in this instance [P] represents the number of free binding sites available on di-Ub (2 x [P] + [PL]), [L] represents the concentration of free Znf_A20 and [PL] represents the sum of occupied binding sites ([PL] + 2 x [$PL_2$]). The equilibrium constants derived from ESI-MS titrations with di-Ub and Znf_A20 are described in Table 4-2.

ESI-MS titrations of Znf_A20 with mono and both linkage types of di-Ub give equilibrium constants that are within good approximation of one another. This supports the suggestion from membrane-based binding assays that the Znf_A20 has

little or no specificity for Ub chain linkage type and that the two sites are independent. However for two equal sites $Ka_1 / Ka_2 = 4$ (Equation 2-15), while this is the case for Lys63-$Ub_2$, for Lys48-$Ub_2$ $Ka_1 / Ka_2 = 7.9$, suggesting negative cooperativity between the two sites. However the errors associated with this value are large, and while there is a suggestion of negative cooperativity the data is of insufficient quality and would need to be assessed by an alternative method.

**Table 4-2: Table of ESI-MS determined equilibrium constants for Znf_A20 titrations with mono- and di- Ub.**

| | Mono-Ub | K48-$Ub_2$ | K63-$Ub_2$ |
|---|---|---|---|
| $Ka_1$ /M (±) | - | 498451 (202971) | 282509 (28917) |
| $Ka_2$ /M (±) | - | 63023 (6879) | 65443 (6879) |
| $Ka$' /M (±) | 147550 (8971) | 231367 (90459) | 157291 (18459) |
| $Ka_1 / Ka_2$ (±) | - | 7.9 (3.8) | 4.3 (0.5) |
| $K_{d1}$ /μM (±) | - | 2.0 (1) | 3.6 (1) |
| $K_{d2}$ /μM (±) | - | 15.6 (4.8) | 15.5 (1.5) |
| $K_d$' /μM (±) | 6.8 (1) | 4.3 (1.4) | 6.5 (1) |

### 4.2.3 Summary

From a combination of Biochemical and Biophysical techniques the formation of a high affinity complex between Ub and the Znf_A20 domain of ZNF216 has been confirmed. The interaction utilises the Asp58 centred patch although this residue is not sufficient for the interaction alone, the interaction shows no clear specificity for poly-Ub chain linkages in membrane-based binding assays or by ESI-MS.

## 4.3 Structural Investigation of the Znf_A20 Interaction with Ub

### 4.3.1 Chemical Shift Mapping of the Znf_A20-Ub Interaction

In order to characterise the interaction between the Znf_A20 and Ub on a per-residue basis $^{1}H$-$^{15}N$-HSQC based NMR titrations were performed. As discussed previously the sensitivity of the backbone amide proton and nitrogen chemical shifts

to their environment make them ideal probes for studying protein-protein interactions on a per residue basis. $^{1}$H-$^{15}$N-HSQC-based titrations were performed on $^{15}$N-labelled Znf_A20 with unlabelled Ub and *vice-versa*; in both cases the $^{15}$N-labelled protein was maintained at 1 mM while the unlabeled protein was titrated in from 0 up to 4 mM.

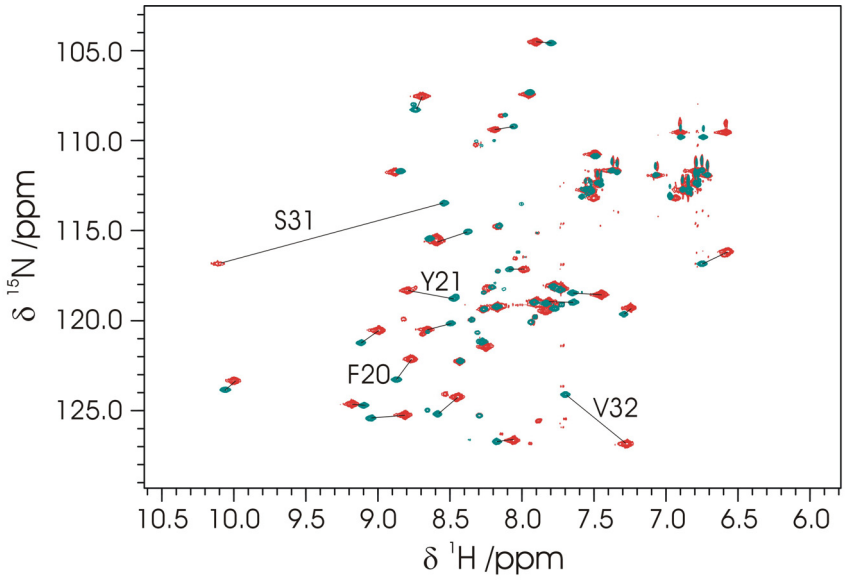



**Figure 4-6: $^{1}$H-$^{15}$N-HSQC of $^{15}$N-Znf_A20 in the presence (Red) and absence (Blue) of 4 mM Ub, the four large chemical shift changes for F20, Y21, S31 and V32 are labelled.**

The addition of Ub to $^{15}$N-Znf_A20 results in large chemical shift changes consistent with a high affinity binding event (Figure 4-6). Many peaks of the $^{1}$H-$^{15}$N-HSQC of $^{15}$N-Znf_A20 were seen to broaden in response to the addition of Ub followed by a return to sharp line shapes at high Ub concentration (Figure 4-7). This is consistent with a binding event in an exchange regime which is intermediate on the NMR time scale, with residues with large chemical shift changes disappearing completely during the titration in an intermediate to slow exchange regime.

The majority of resonances were traced from the free pre-assigned $^{1}$H-$^{15}$N-HSQC to that of the Ub saturated form. Where this was not possible due to extensive exchange broadening, assignment of the bound form was confirmed as described previously using $^{1}$H-$^{15}$N-HSQC-NOESY correlations (Section 3.3.1).

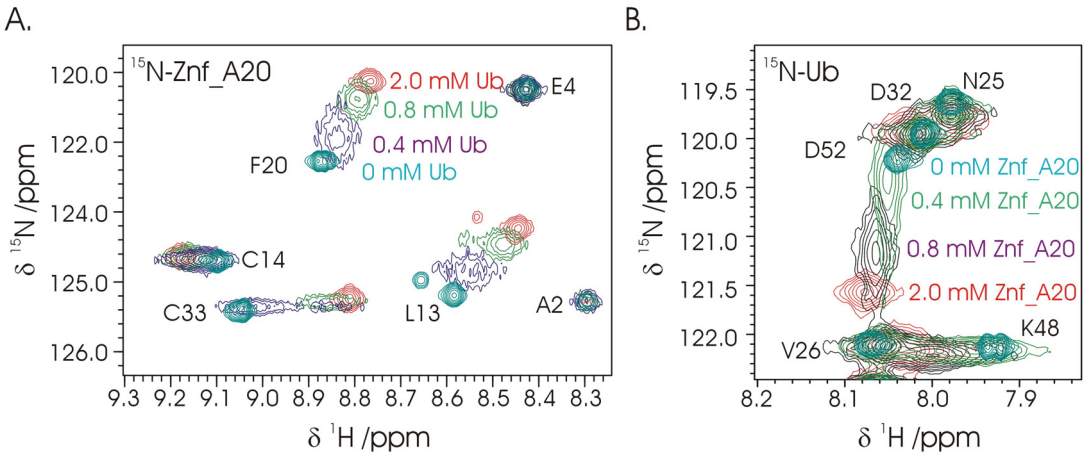



**Figure 4-7: Portion of the $^{1}$H-$^{15}$N-HSQC of; (A) Znf_A20. (B) Ub. In both cases in the presence of 0, 0.4, 0.8 and 4 mM partner protein. Assignments for the unbound peaks are included.**

In particular, large CSPs of 1.7 and 0.7 ppm are seen for Znf_A20 residues Ser31 and Val32, respectively (Figure 4-7 and Figure 4-8), with additional significant effects (>0.2) for residues Leu13, Phe20, Tyr21, Asn23, Arg25, Cys33, and Lys35 which cluster to form a discrete binding patch involving residues at the base of the helix and from the loops of the zinc knuckle (Figure 4-8).

The $^{1}$H-$^{15}$N-HSQC of $^{15}$N-Ub showed the same large chemical shift perturbations in response to the addition of Znf_A20 as a clear indication of a binding event (Figure 4-9). As seen for the $^{15}$N-Znf_A20 titration several peaks broadened in response to the addition of Znf_A20 which then return to sharp line shapes at high Znf_A20 concentrations (Figure 4-7).

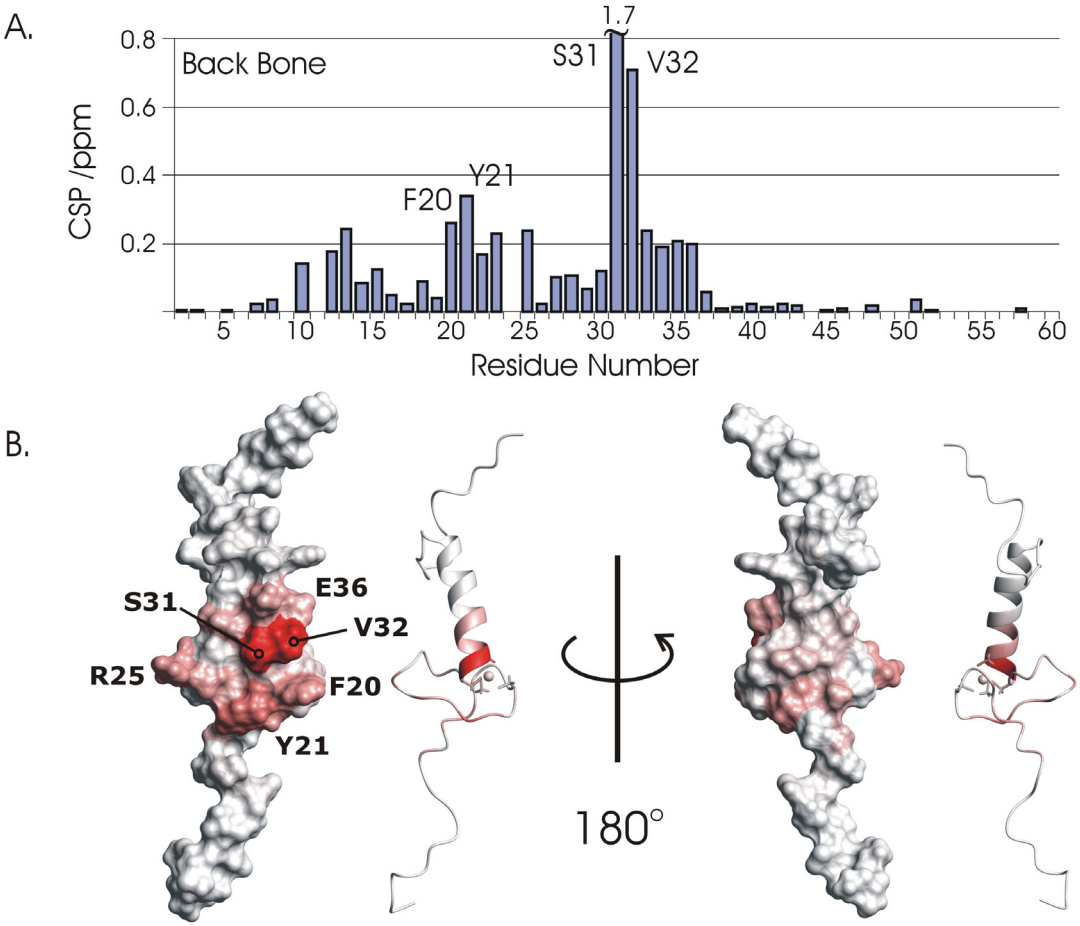



**Figure 4-8: (A) Plot of residue specific chemical shift perturbations for the $^{15}$N-Znf_A20 binding to Ub. (B) High chemical shift perturbations were plotted onto the surface of the Znf_A20 in red.**

Significant CSPs were seen for residues Ile23, Asp24, and Lys48–Ser65. A cluster of highly perturbed residues (CSP >0.35) were seen in and around the loop between β-strands 4 and 5, including Thr55, Asp58, and the side chain amino group of Asn60 and Asp24 at the N-terminus of the main α-helix. Interestingly the side chain NH of Arg54 is not visible in the $^{1}$H-$^{15}$N-HSQC of free Ub, however in the saturated spectra it shows good peak intensity with a relatively narrow line width. Although the backbone NH of Arg54 shows only a relatively small CSP the sharpening of the side chain NH may well indicate that Arg54 is involved in binding to the Znf_A20 (Figure 4-9 and Figure 4-10).

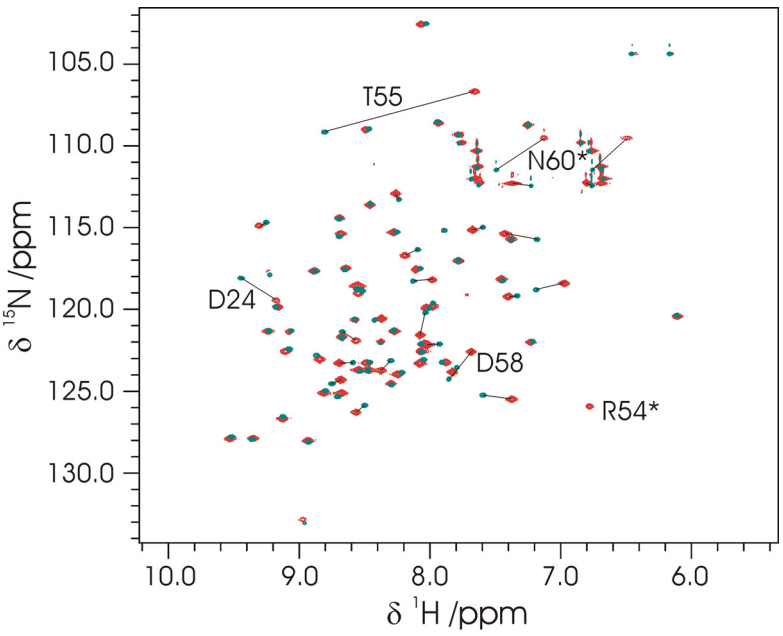



**Figure 4-9: $^{1}$H-$^{15}$N-HSQC of $^{15}$N-Ub in the presence (Red) and absence (Blue) of 4 mM Znf_A20, large chemical shift changes are labelled, side chains are indicated by '*'. The side chain NH of R54 visible in only the Znf_A20 saturated spectrum is labelled.**

No chemical shift changes were observed to indicate binding to the hydrophobic patch centred on Ile44 suggesting that the canonical binding surface is not directly utilised by the Znf_A20 (Figure 4-10). Residues showing large values of CSP were mapped onto the structure of the Ub crystal structure (RPDB ID: 1UBQ), forming a contiguous interaction surface centred at Asp58. The Asp58 centred binding site is adjacent to the Ile44 centred canonical binding site but forms a unique binding site (Figure 4-10C) which is consistent with the D58A but not the I44A mutant showing an impaired interaction by membrane-based binding assays (Figure 4-1).

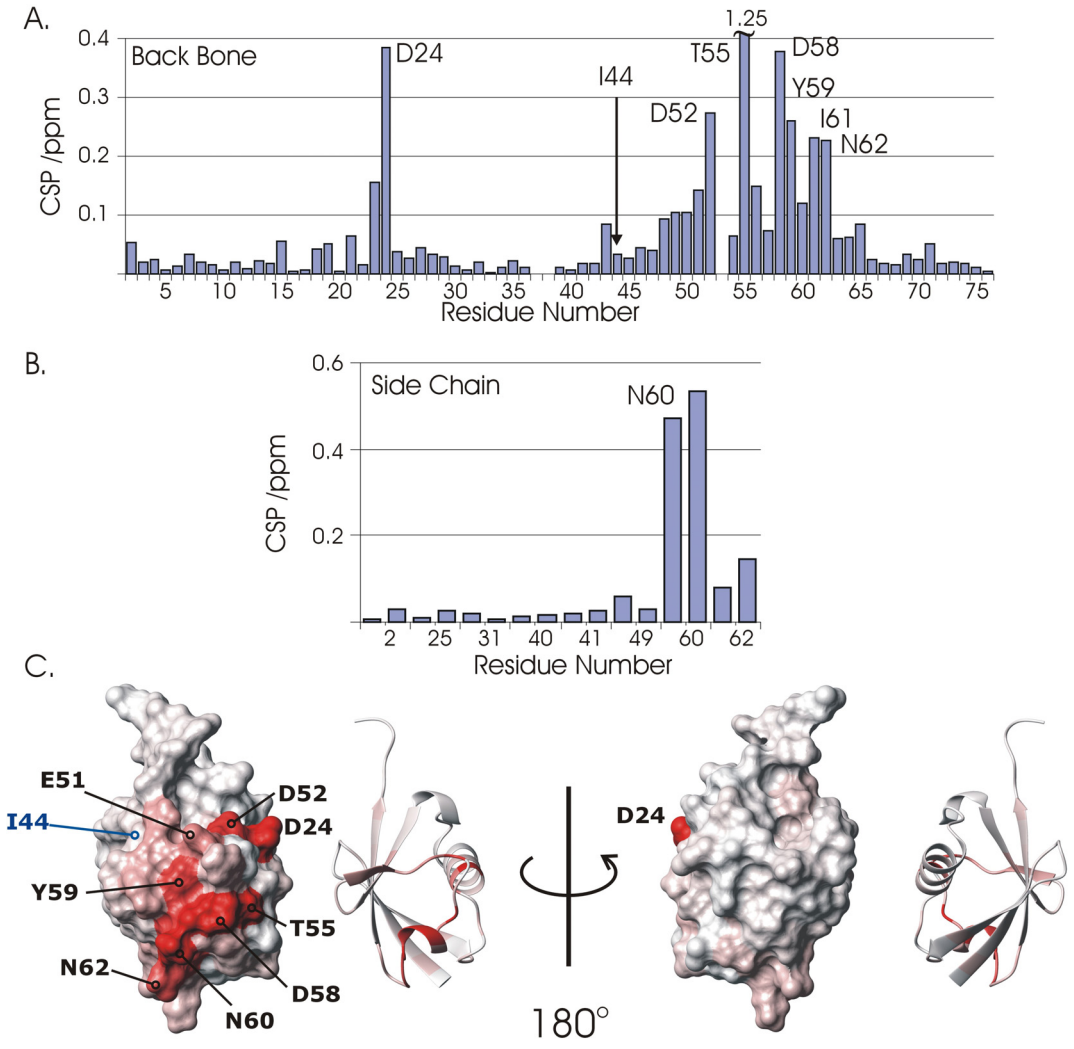



**Figure 4-10: Plot of residue specific chemical shift perturbations observed for $^{15}$N-Ub binding to Znf_A20 for (A) Backbone NH (B) and side chain $NH_2$. (C) High chemical shift perturbations were plotted onto the surface of Ub in red. Residues with large CSP and the canonical Ile44 site are annotated.**

Using the interaction surface map determined from $^{1}$H-$^{15}$N-HSQC titrations the binding surface properties of the two proteins were examined using the program MOLMOL[64]. The binding surface of the Znf_A20 appears largely hydrophobic in nature (Figure 4-11) made up of Val32, Phe20, and Tyr21 which surround Ser31, with only the two positively charged residues of Arg25 and Lys35 at the periphery of the proposed binding site (Figure 4-10).

The proposed binding surface on Ub on the other hand, appears largely polar in portions, with Asp58 sat at its centre and Asp24, Glu51, Asp52, and Arg54 at the periphery. However the surface is varied and appears to have a mix of hydrophobic and polar residues running down the middle including of Thr55, Tyr59, Asn60, and Asn62.

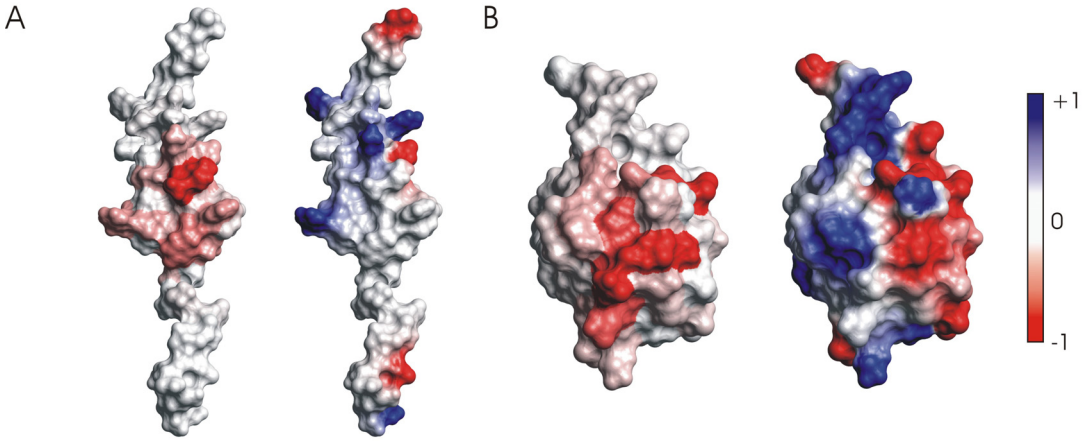



**Figure 4-11: Comparison of the CSP interaction surface map (Left) to the electrostatic surface potential (Right). (A) for Znf_A20 (B) for Ub. Electrostatic surface potentials were calculated using MOLMOL[64].**

### 4.3.2 Docking Restraints

The identification of a binding site by chemical shift mapping is a prerequisite for NMR structural studies of a protein-protein complex using NMR. However the lack of obvious surface-complementarity would make the generation of an accurate NMR-derived model of the complex difficult without further information, as the standard protocols are based largely on surface-complementarity[72]. Extra information derived from NMR can come from a variety of sources; the most useful would be the detection of intermolecular NOEs between the two proteins. This can be achieved using 'half-filtered' NOE where $^{15}$N-$^{13}$C-labeled protein is mixed with unlabeled protein and $^{1}$H-$^{1}$H NOE are filtered based on the labelling of the atoms bonded to the two coupled nuclei[56]. These 'half-filtered' NOE were used as two types; the first based on $^{15}$N-based filtering, where NOE are detected between protons bonded to $^{15}$N and those bonded to any other isotope ($^{14}$N, $^{13}$C and $^{12}$C). The second involve $^{13}$C-based filtering, where NOE are detected between protons bonded to $^{13}$C and those attached to any other nuclei ($^{13}$C, $^{15}$N and

$^{14}N$)[56]. 'Half-filtered' NOE were recorded on samples of both $^{15}N$-$^{13}C$-Znf_A20 with unlabeled Ub and vice versa; in both cases the data was of very poor quality with only four unambiguous intermolecular NOEs being detected. The potential reasons for this are numerous, but the intermediate exchange regime of the binding event is likely to have had a great effect on the ability to detect intermolecular NOE.

### 4.3.3 Intermolecular Paramagnetic Relaxation Enhancement

The lack of detectable intermolecular NOE between Znf_A20 and Ub necessitated alternative methods to generate intermolecular restraints with which to guide complex-model generation. Paramagnetic relaxation enhancement (PRE) experiments can be used to generate intermolecular restraints by monitoring the enhanced relaxation of nuclei in $^{15}N$-labeled protein caused by the presence of an MTSL spin label covalently attached to its unlabeled partner[68,99,152]. The spin label MTSL was covalently coupled to a cysteine engineered into the K48C mutant of Ub, over 90% incorporation of the MTSL spin label was conformed by ESI-MS (Appendix D). 2 mM MTSL-K48C-Ub was mixed with 1 mM $^{15}N$-Znf_A20 in order to measure intermolecular PRE. $^{1}H$-$^{15}N$-HSQC was recorded before and after reduction of the spin label using 10mM sodium ascorbate, a long relaxation delay of 5 s was used to ensure proper relaxation. PRE was measured as the ratio of the peak intensities in the paramagnetic and diamagnetic states.

A wide range of intermolecular PRE was observed, with complete signal attenuation of residues Leu13, Cys14, Phe20, and Tyr21 with many of their surrounding residues showing considerable attenuation which appeared to reduce in a distance dependent manner (Figure 4-12).

PRE intensities were used to generate approximate distance restraints for initial complex modelling similar to previous studies[67,68]: PRE values of < 0.1 were given a range of 2–15 Å, between 0.15 and 0.4 a range of 10 to 19.5 Å , 0.4 to 0.6 a range of 15 – 22 Å, and 0.6 -0.8 were given a range of 15 – 25 Å: values over 0.8 were not used. Accurate distance restraints were not calculated from PRE measurements for a number of reasons, the first is that accurate distance calculations require a knowledge of the protein $\tau_c$ which is an unknown in this instance. The second is that we are not dealing with a fixed distance; based on chemical shift mapping the Lys48 side chain (and therefore the MTSL-Cys side chain) points out into solution away from the binding site, it would therefore be expected to be

dynamic and not occupy a fixed position. Also the measured PRE are between two proteins which exist for a period of time free in solution, therefore again the measured PRE is not over a fixed distance. Finally the concentration of MTSL-K8C-Ub is relatively high (2 mM) so that there may be additional effects from free spin-labelled protein in solution. All of these factors would make accurate determination of distance restraints from PRE measurements difficult and wide distance restraints boundaries necessary.

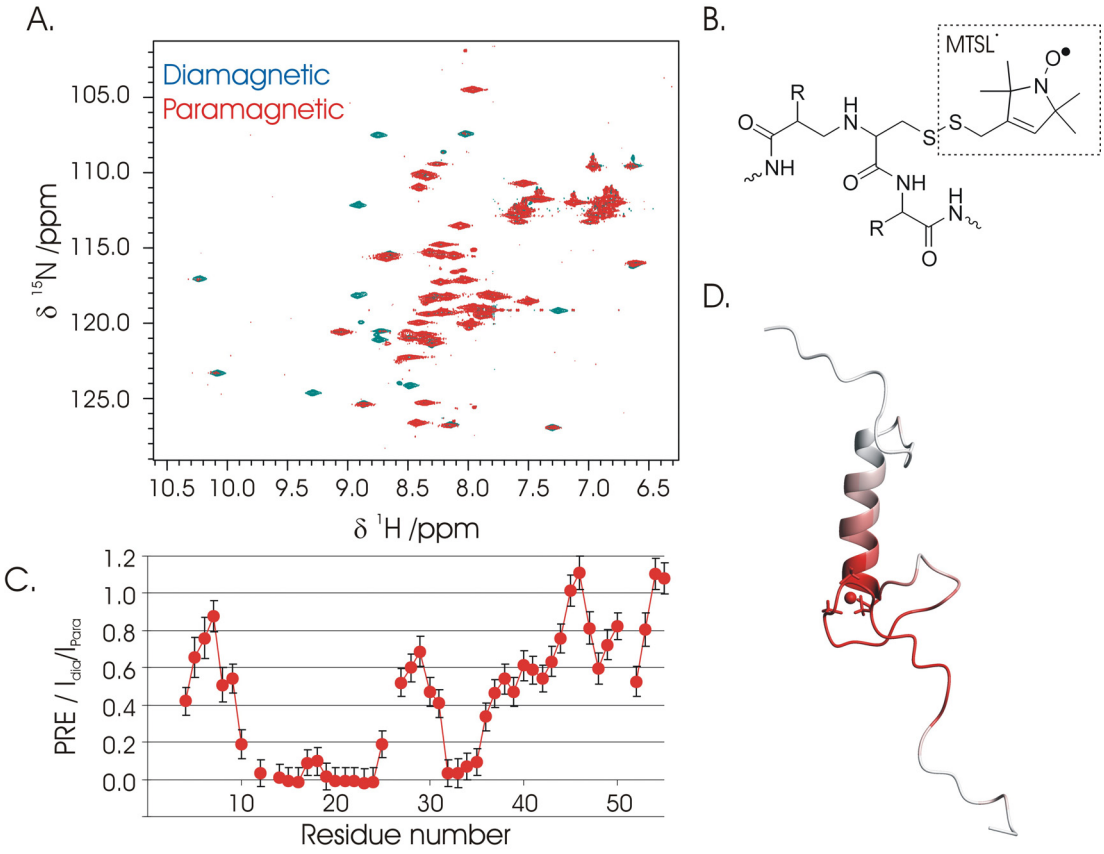



**Figure 4-12: (A) $^{15}$N-HQSC of $^{15}$N-Znf_A20 in the presence of paramagnetic (Red) and diamagnetic (Teal) MTSL-K48C-Ub. (B) The structure of the MTSL spin label coupled to a cysteine containing poly peptide. (C) Plot of residue specific PRE values for the Znf_A20. (D) Residues showing large PREs were plotted onto the ribbon diagram of the Znf_A20 in red demonstrating the localisation of intermolecular PRE.**

The K48C-Ub mutation with MTSL attachment also resulted in cross interface CSPs seen in the $^{1}$H-$^{15}$N-HSQC of Znf_A20 when compared to saturation

with wt-Ub. The observed CSPs were localised almost exclusively to Phe20 and surrounding areas (Figure 4-13). These localised CSPs may suggest that the K48C site (or the MTSL group) is close in space to Phe20. CSPs were used to generate an inter-residue distance restraint for complex modelling with a maximum distance boundary of 5.5 Å between Phe20 in the Znf_A20 and Lys48 in Ub.

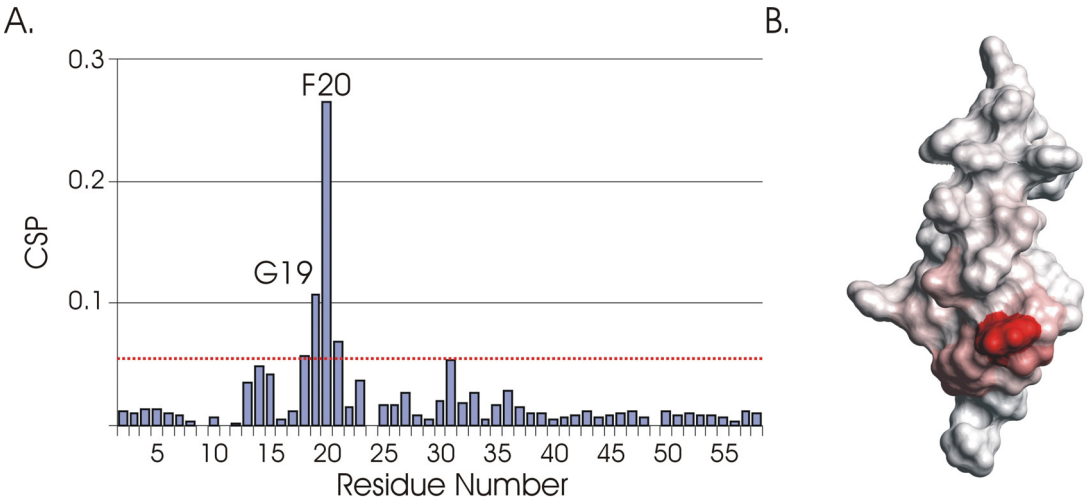



**Figure 4-13: (A) plot of residue specific CSP caused by the MTSL-K48C-Ub site. The vertical red line represents the threshold value for significant CSP. (B) Large CSP plotted onto the surface of the Znf_A20 in red.**

Intermolecular PRE provided a total of 29 distance restraints which were used along side CSP and RDC data as the basis for the generation of a structure between Znf_A20 and Ub.

### 4.3.4 Complex RDC Measurements

In the past 10 years the use of RDCs recorded on partially aligned biomolecules to generate long-distance information for NMR based structure generation has become routine[70,102,153,154]. More recently the usefulness of RDC measurements for the relative alignment of individual proteins in multi-protein systems has become increasingly clear[102,154]. This can be particularly useful in systems where standard restraints such as NOE restraints are difficult to obtain or ambiguous[154]. Docking software such as HADDOCK,[72] which is used routinely for NMR based protein complexes, typically use RDC restraints in one of two ways. The first involves fitting the RDC data of the two monomeric units to the same alignment tensor, which is then used to orientate the two proteins prior to docking[102,154]. This

method is particularly good for systems where the structures of the two units are accurately defined giving accurate well defined alignment tensors and in symmetric homo-oligomers where other sources of restraints may be unreliable. The second is the use of pairwise intervector projection angles, which are independent of the orientation of the alignment tensor with respect to the molecule[70,71]. This has the benefit of not requiring a previous knowledge of the alignment tensor or very accurately defined monomeric units. Moreover where docking is performed using structures of the free monomeric proteins both intra- and intermolecular intervector projection angles restraints can be used in the docking protocol; while the intermolecular restraints guide the alignment during complex docking, intramolecular restraints will refine the monomeric structure to account for subtle conformational changes[71].

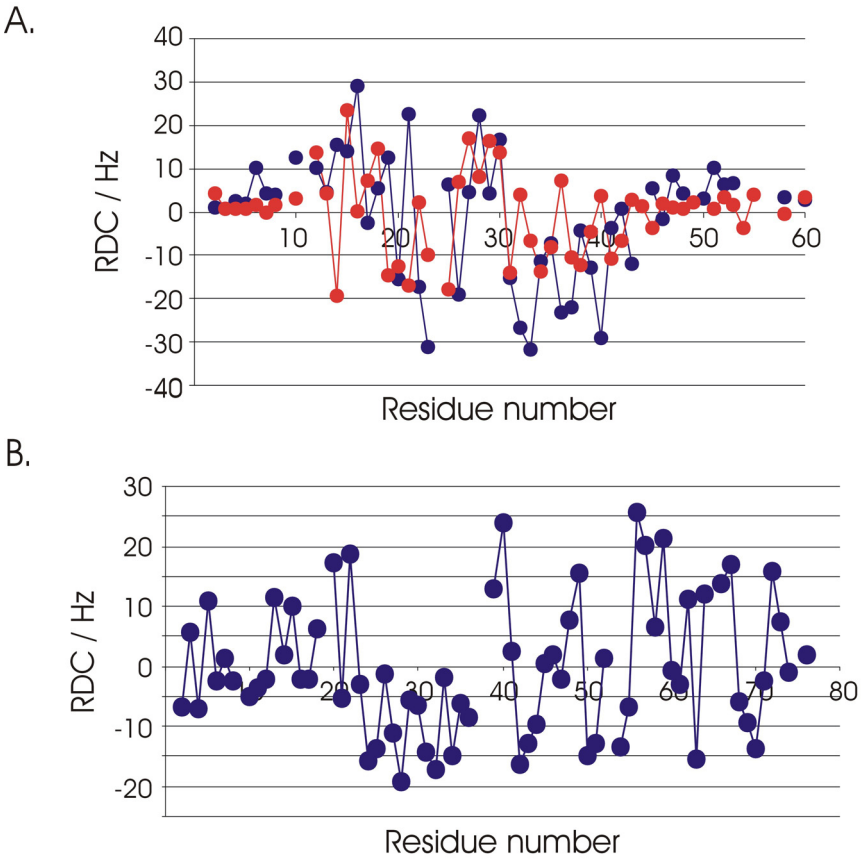



**Figure 4-14: Backbone $^{1}D_{NH}$ RDC measurements recorded on (A) $^{15}$N-Znf_A20 recorded in the presence (in 5% PAG) (Blue) and absence (in 7% PAG) (Red) of 2 mM Ub. (B) on $^{15}$N-Ub in the presence of 2 mM Znf_A20 (in 5% PAG).**

RDCs were recorded for both $^{15}$N-Znf_A20 saturated with 2 mM Ub and vice versa. Alignment was initially achieved using 7% PAG, however this resulted in severe line broadening which made accurate RDC measurement difficult. In order to reduce the observed line broadening the 7% PAG was replaced with 5% PAG in order to reduce the degree of alignment to reduce line broadening. This approach proved successful and a total of 28 and 64 $^{1}D_{NH}$ RDCs for Znf_A20 and Ub respectively were measured and used to generate a total of 284 intermolecular and 390 intramolecular vector angle projection restraints using the programs PALES[153] and HADDOCK[71,72].

For both proteins large RDC values were observed (-19 to 26 Hz) suggesting good partial alignment in the 5% PAG. The measured RDC values for the free and Ub bound Znf_A20 do not correlate well, this could be caused by subtle changes in either the conformation or the alignment tensor between the free and bound state. The relatively small CSP for the majority of the Znf_A20 would suggest that no considerable conformational change has occurred.

### 4.3.5 Mutation Based Intermolecular CSP Restraints

The dependence on relatively long-range distance restraints from PRE and RDC measurements may be aided by more close-contact based restraints. In an attempt to gain more information for the complex structure generation, mutations were made in the Znf_A20 domain and $^{1}$H-$^{15}$N-HSQC titrations repeated (Using $^{15}$N-UB) in order to observe intermolecular CSPs corresponding to contacts involving the mutation site, or its surrounding residues. In order to guarantee the chosen mutations would not disrupt the Znf_A20 fold or its binding to Ub, two separate mutations were made to solvent exposed residues, in both cases being mutated to their corresponding residues in the Znf_A20 domain of Rabex-5 which binds to Ub with similar affinity to ZNF216[20,21]. The chosen mutation were V32K and F20Y both of which are located at or close to the binding patch, 2D $^{1}$H-$^{1}$H-NOESY and $^{1}$H-$^{1}$H-TOCSY spectra were recorded on both mutants; neither showed large CSP for residues other than the mutation site and residues which make immediate contact with the mutation site suggesting a conservation of the domains fold.

$^{1}$H-$^{15}$N-HSQC titration of $^{15}$N-Ub with mutant-Znf_A20 gave HSQC which overlaid well with that of Ub saturated with Wt-Znf_A20 with only localised CSPs observed in both cases.

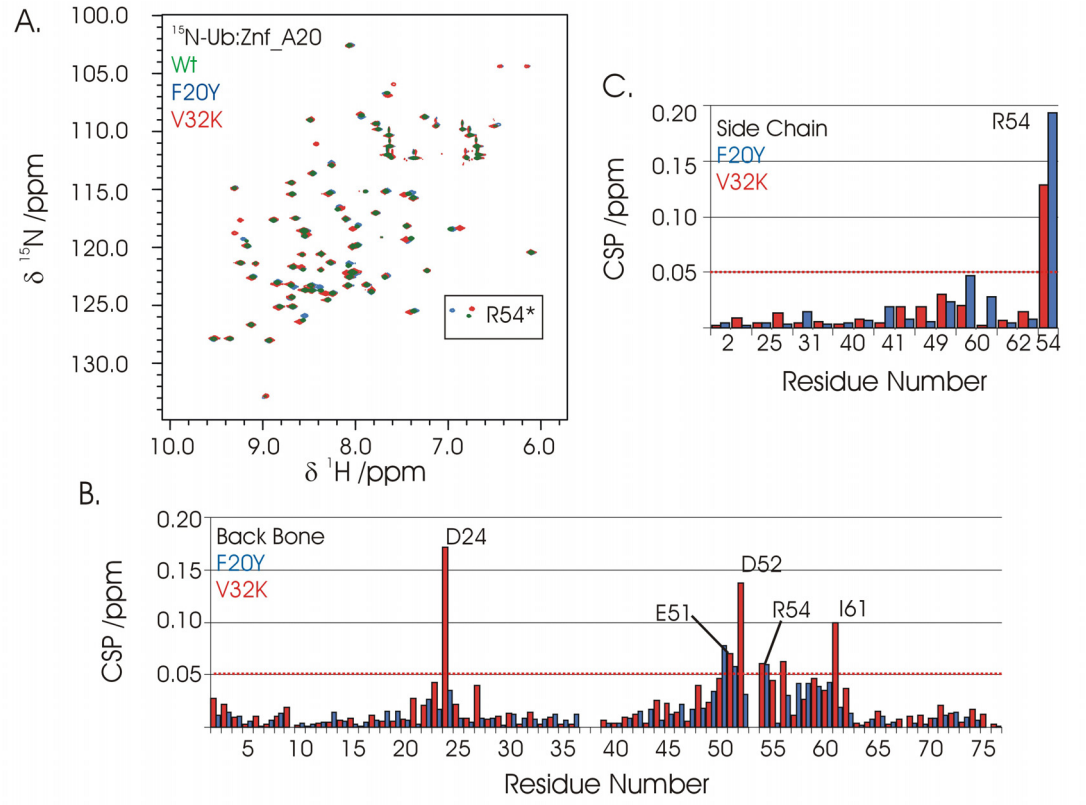



**Figure 4-15: (A) $^{1}$H-$^{15}$N-HSQC of $^{15}$N-Ub saturated with Znf_A20 (Green), F20Y-Znf_A20 (Blue) or V32K-Znf_A20 (Red). (B and C) Plot of residue specific intermolecular CSP caused by the F20Y (Blue) or the V32K (Red) mutants for backbone NH and side chain NH/NH$_2$. The horizontal red line represents the threshold value for significant CSP.**

The intermolecular CSP caused by the F20Y mutant were almost exclusively localised to the side chain NH of Arg54 (Figure 4-15) and to a lesser extent its backbone NH and the backbone NH of Leu50. The V32K mutant similarly causes localised CSP involving the side chain NH of Arg54, the backbone NH of Asp24, and Asp52 and to a lesser extent Leu50 and Ile61 (Figure 4-15). Although the Arg54 backbone NH does not experience a large CSP during titrations it is sat at the centre of the proposed binding patch and its side chain NH is only visible in the presence of saturating amounts of Znf_A20. Combined with large intermolecular CSP from the two studied mutants, this supports the suggestion that Arg54 may be involved in the binding patch (Figure 4-10).

### 4.3.6 Complex Structure Generation

The structure of the Znf_A20 was refined for the bound state using 345 NOESY-distance restraints derived from $^{1}$H-$^{15}$N-HSQC -NOESY and $^{13}$C- and $^{15}$N-half-filtered NOESY experiments, 30 $^{1}D_{NH}$ RDC restraints and 161 dihedral restraints predicted from backbone chemical shifts. The XPLOR-NIH protocol was performed as described above (Section 2.5.3) but with the free Znf_A20 structure used as an initial structure. The 10 lowest energy structures were selected based on the absence of NOE restraint violations above 0.3 Å or dihedral violations above 5°. The 10 lowest structures gave a backbone RMSD over the core domain of 0.6 Å, and a low RMSD to the free structure of 0.7 Å. Complex structure generation was performed using both the free and bound calculated structures for the Znf_A20 domain with little or no difference in the final structures obtained.

A combination of CSP mapping, intermolecular distance restraints (NOE, mutation-based intermolecular CSP, and PRE restraints) and RDC restraints were used to generate a family of structures for the Znf_A20-Ub non-covalent complex using the program HADDOCK[72] (Version 2.1). Due to the non-complementary nature of the two surfaces the electrostatic energy terms were not used during the flexible docking protocol, with this exception, complex structures were generated using the standard HADDOCK protocol with the 'ZN ion' patch included[72,73].

During initial modelling the Ub-Asp58 side chain was observed to be consistently in close proximity to Ser31 of Znf_A20. In the structure of the Rabex-5 Znf_A20 structure the presence of a hydrogen bond between Ub-Asp58 and the Ser corresponding to ZNF216-Ser31 is also observed. The proton chemical shift for the backbone NH proposed to form this hydrogen bond, shifts downfield by ~1.5 ppm from around 8.5ppm to 10 ppm as a result of binding to Ub. Such a large chemical shift change can best be explained by the formation of the proposed hydrogen bond to Ub-Asp58. In light of this, the appropriate hydrogen bond restraint was incorporated and the HADDOCK protocol repeated. A full list of the restraints used in generation of the complex structures can be found deposited at the BMRB under accession code 17024.

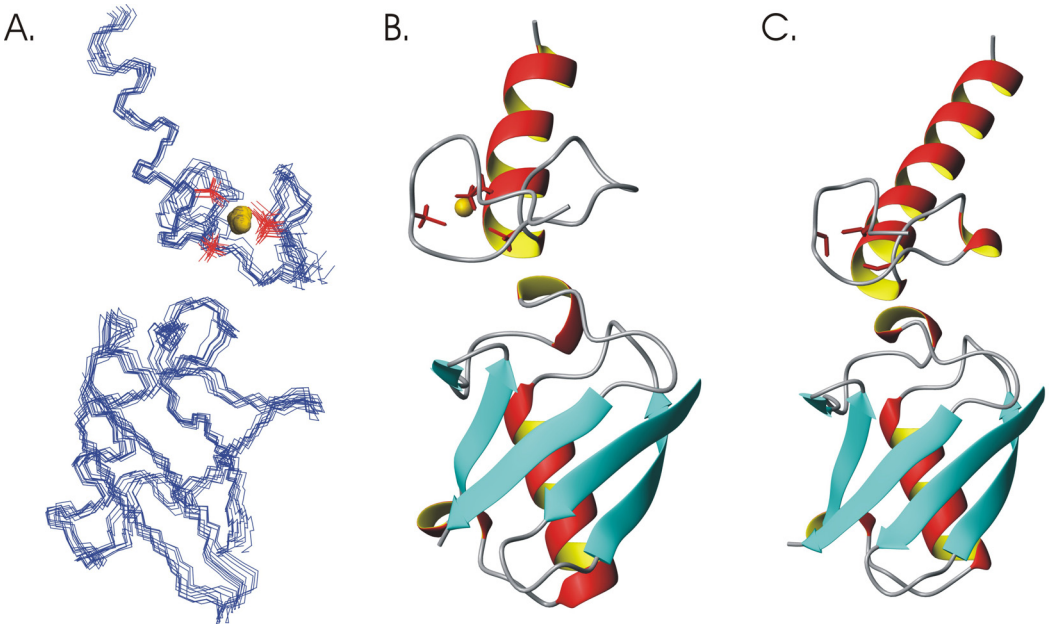



**Figure 4-16: (A) Backbone trace of the ensemble of 10 lowest energy structures for the non-covalent complex between Znf_A20 and Ub, the four Cys side chains are drawn in red, the zinc atom in gold. (B) Ribbon diagram of the lowest energy structure. (C) Ribbon diagram of the Rabex-5 Znf_A20 complex with Ub (RPDB ID: 2C7N).**

A total of 500 structures were generated from the rigid body docking protocol with 50 selected for the 'semi-flexible' and 'fully-flexible' docking protocol. Of the 50 generated structures 40 showed no distance restraint violations and violated less than 10% of the intervector projection angle restraints, all 40 structures clustered into the same HADDOCK defined cluster with a 1.5 Å cut off value.

Of the 40 acceptable structures the 10 with the lowest HADDOCK energies were selected to represent the structural ensemble giving an average HADDOCK energy of -35 ± 1.5. The lowest 10 structures have a low backbone RMSD of 0.5Å across the full residue range (Figure 4-16). The structure was analysed using PSVS as before. The structural statistics can be found in Table 4-3 along with the Ramachandran statistics Table 4-4.

**Table 4-3: Output from the structure validation by the PSVS for the 10 lowest structures.**

| ***Summary from psvs*** [a] | |
|---|---|
| *Distance Constraints:* | |
| Ambiguous (AIR) restraints | 13 |
| Unambiguous total | 37 |
| NOE | 4 |
| PRE | 29 |
| mCSP | 4 |
| **Total Structures Computed** | 500 |
| **Number of Structures Used** | 10 |
| **Residual constraint Violations** [a,b] | |
| *Distance violations /structure* | |
| 0.1-0.2 Å | 3 |
| >0.2 Å | 2 |
| RMS of distance violations / constraints | 0.04 Å |
| Maximum distance violation [c] | 0.50 Å |
| **RMSD Values** | |
| All backbone atoms | 0.5 Å |
| All heavy atoms | 0.5 Å |
| **Structure Quality Factors – overall statistics**[e] | Mean score |
| Procheck G-factor [d] (Phi /Psi only) | -0.17 |
| Procheck G-factor [d] (all dihedral angles) | -0.13 |
| Verify3D | 0.41 |
| MolProbity clash score | 13.66 |
| **Close contacts and deviation from ideal geometry**[e] | |
| Number of close contacts | 10 |
| RMS deviation for bond angles | 1.5 ° |
| RMS deviation for bond lengths | 0.011 Å |
| **[a] Analysed all residues** | |
| **[b] Calculated for all constraints over the given residues, using sum over $r^{-6}$** | |
| **[c] Largest constraint violation among all the reported structures** | |
| **[d] Residues with sum of phi and psi order parameters > 1.8** | |
| **Generated using PSVS 1.4** | |

**Table 4-4: Ramachandran statistics for the structural ensemble.**

| **Procheck** | Favoured / % | Additionally allowed / % | Generously allowed / % | Disallowed / % |
|---|---|---|---|---|
| All Residues | 88.7 | 11.3 | 0.0 | 0.0 |
| **Molprobity** | | | | |
| All Residues | 94.3 | 5.7 | NA | 1.3 |

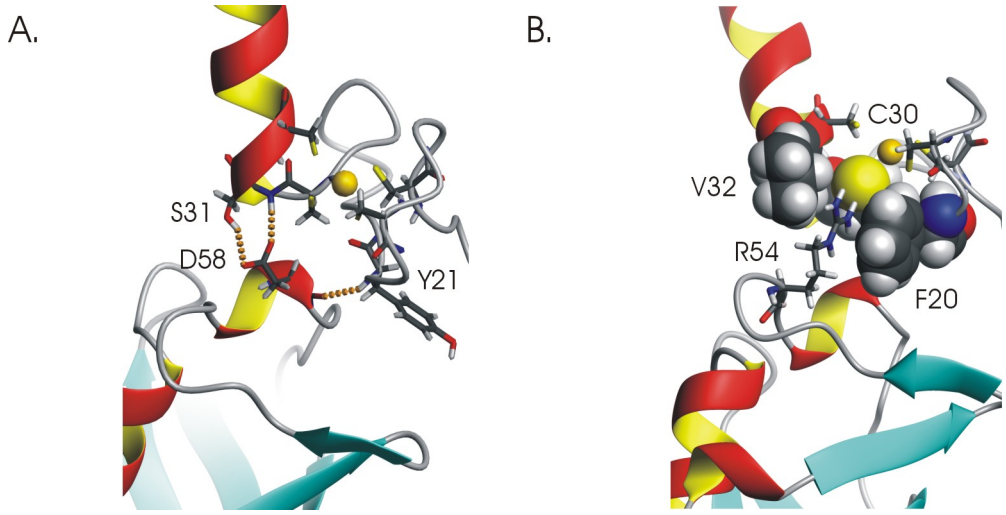



**Figure 4-17: Key intermolecular contacts between Ub and Znf_A20 including; (A) Hydrogen bonds between Ub-Asp58 with Znf_A20-Ser31 and Tyr23. (B) The interaction of Ub-Arg54 with Znf_A20-Phe20, Val32 and Cys30.**

The solution structure presented here is in good agreement with that observed in the crystal structure of the Rabex-5 Znf_A20-Ub complex, the interaction buries an average of 771 ± 65 Å$^2$ of solvent exposed surface area compared to ~800 Å$^2$ for the Rabex-5 complex. The key residues within the zinc knuckle are involved in similar interactions where Ub-Asp58 forms hydrogen bonds to Znf_A20-Ser31 and the burial of Ub-Arg54 in a cleft formed from Znf_A20-Phe20, Val32, and Cys30 resulting in favourable stacking interactions between the phenyl group of Znf_A20-Phe20 with the guanidinium group of Ub-Arg54 which also forms favourable polar interactions with the thiolate sulphur of Znf_A20-Cys30 (Figure 4-17)[155]. Additional favourable interactions between Znf_A20 and Ub may include a hydrogen bond between the backbone NH of Znf_A20-Tyr21 and the backbone carbonyl of Ub-Asp58, an edge-to-face aromatic interaction between Znf_A20-Phe20 and Ub-Tyr59

[156], and packing of Ub-Asn60 against Znf_A20- Tyr21 and Asn23. A summary of the observed contacts can be seen in Figure 4-19.

Although the intermolecular contacts are essentially the same the two structures are visibly different in several respects; the zinc knuckle of the structure presented here is sat closer to the Ub loop between strand β4 and the $3_{10}$ helix resulting in both Znf_A20-Phe20 and Tyr21 displaced by up to 1.5 Å compared to the corresponding di-Tyr in the crystal structure of Rabex-5 (RPDB ID: 1FIF and 2CN7). Whether this represents genuine differences between the two complexes or is simply the result of inaccuracies in the structure presented here as a result of over reliance on long distance restraints and the absence of reliable close contact restraints is difficult to judge. The two structures make near identical intermolecular contacts and the structure presented here is consistent with all of the available data.

### 4.3.7 Evaluation of the Observed Chemical Shift Perturbations

$^{1}$H-$^{15}$N-HSQC titrations using both $^{15}$N-Znf_A20 with unlabeled Ub and vice versa, showed large CSPs for both proteins. The complex structure presented here is consistent with the observed CSPs and can provide justification for the majority of the observed features of the $^{1}$H-$^{15}$N-HSQC titrations. The measured CSP are all located close or adjacent to the binding interface of the structure proposed here (Figure 4-18).

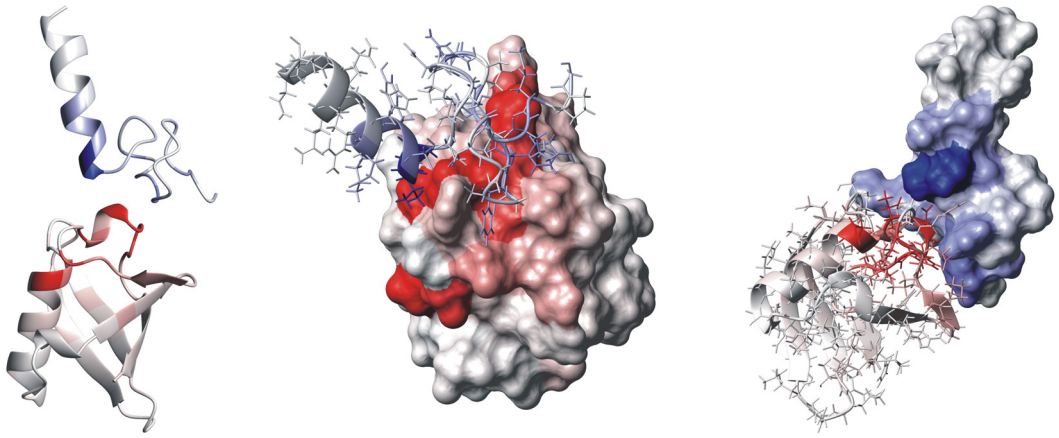


**Figure 4-18: Large chemical shift perturbations plotted onto the surface or onto ribbon diagram of Znf_A20 (Blue) and Ub (Red).**

The first CSP of interest is the ~1.5 ppm downfield shift of the backbone amide proton of Ser31 of Znf_A20. Such large chemical shift perturbations could be

explained by the introduction of an aromatic ring in close proximity to Ser31 however the magnitude and sign of the chemical shift perturbation is better explained by the formation of a hydrogen bond which is consistent with the reported Rabex-5 complex with Ub and early modelling[63].

The appearance of the side chain NH of Ub-Arg54 during the titration is wholly consistent with the structure reported here. As discussed above, the interaction of Znf_A20 with Ub results in the burial of the Ub-Arg54 side chain including the measured NH (Figure 4-17). The resulting loss of solvent exposure would result in reduced solvent exchange and a subsequent reduction in exchange broadening allowing detection of this NH in the $^{1}H$-$^{15}N$-HSQC spectrum[157].

Indeed the majority of the CSPs can be justified by direct intermolecular contacts; in particular the large shifts for Znf_A20-Phe20, Tyr21, Ser31, and Val32 and for Ub- Ser57, Asp58, Tyr59, and Asn60 can all be justified by direct contacts with the interface.

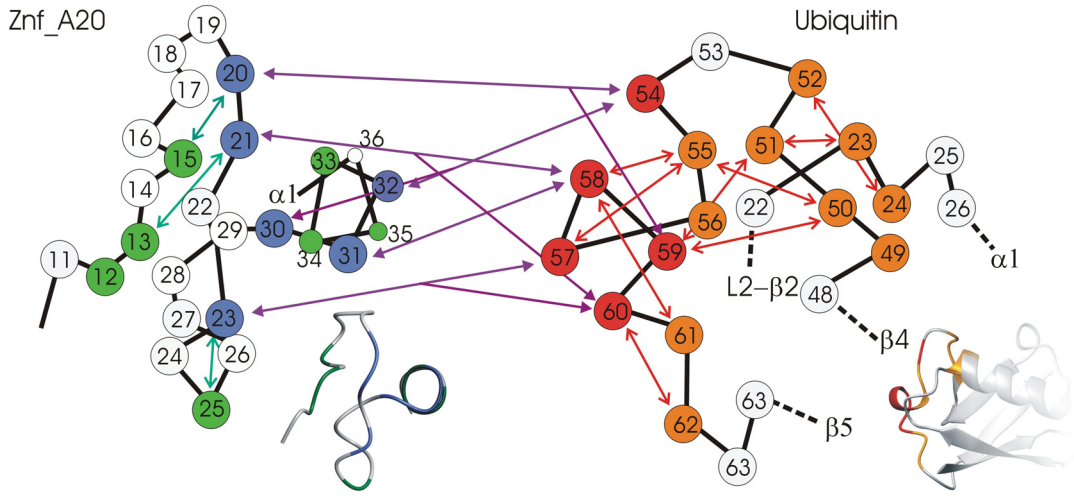



**Figure 4-19: Diagram of the observed contacts between the Znf_A20 and Ub, Interface residues with large CSPs are indicated for Znf_A20 (Blue) and Ub (Red), residues which do not make direct interface contacts but show large CSP are indicated for the Znf_A20 (Green) and Ub (Orange). Solid lines represent inter-residue contacts.**

Although the large CSP for Znf_A20-Val32 may be explained by direct contacts with Ub it is worth noting that Val32 undergoes a large $^{15}N$ chemical shift change of ~2.5 ppm, the backbone amide $^{15}N$-chemical shift has been shown to be heavily dependent on the $\chi_1$ angle of the preceding residue[83,109], the preceding residue

in this case is Ser31 which forms a proposed hydrogen bond to Ub-Asp58. The large observed CSP may well be contributed to by the restriction of rotation around the Ser31$\chi_1$ bond in order to form the proposed hydrogen bond, as well as by direct interface contacts.

### 4.3.8 Non-interface Chemical Shift Perturbations

There are however, several CSP which can not be explained based on direct interface contacts, but can be justified on the basis of the observed complex structure by indirect-intramolecular contact with interface residues (Figure 4-19). Examples of this include Znf_A20-Leu13 which does not make direct contact with Ub, however the movement of the Znf_A20-Tyr21 side chain to make contacts with Ub results in it now making contacts with Leu13. In the free structure the Leu13 side chain sat next to that of Tyr21, in the complex it sits above the plane of the ring potentially forming CH-$\pi$ interactions[158]. This is consistent with not only the Leu13 backbone CSP but also a large side chain CSP for Leu13-H$\delta$ and -C$\delta$ atoms. Similar intramolecular contacts with the interface can also explain the chemical shifts of Znf_A20- Lys35 and Glu36 which is positioned adjacent to the interface residues Ser31 and Val32.

In the case of Ub there are many CSP corresponding to residues distant from the interface. Many can be explained by intramolecular interface contacts, including Ub-Thr55 which is not involved in direct contact with the Znf_A20 but gives the largest recorded CSP for Ub. Thr55 packs against the $3_{10}$ helix containing Ub- Asp58 and Tyr59, in the crystal structure of free human Ub (RPDB ID: 1UBQ) the Thr55 backbone NH forms a hydrogen bond to the Asp58 side chain which is supported by reports that Asp58 shows very little side chain motions despite being solvent exposed[159], in the bound form the Asp58 side chain appears to form a hydrogen bond to Znf_A20-Ser31 which may disrupt the Thr55 hydrogen bond. This would be consistent with the observed CSP which measure an up field shift of ~1.1 ppm for the Thr55 amide proton.

Other CSP on the other hand are more distant from the interface including a cluster of associated residues around the C-terminal end of Ub-$\beta$4, the N-terminal end of the $\beta$4 -$3_{10}$ helix loop and the N-terminal end of the $\alpha$-helix. This patch includes some particularly large CSPs for Asp24, Asp52 and Glu51 (Figure 4-10). The structure does provide insight into these CSPs also; it is possible that subtle

changes in the conformation around the β4 -$3_{10}$ helix loop result from binding contacts at the $3_{10}$ helix through subtle backbone conformation changes (Figure 4-19). However a more direct contact is also visible, in the crystal structure of human Ub (RPDB ID: 1UBQ) the phenol oxygen of Ub-Tyr59 forms a hydrogen bond with the backbone NH of Ub-Glu51 providing a physical link between the binding site and the Asp52-Glu51 pair. In turn Asp52 sits adjacent to Asp24 in the Ub structure, it is therefore possible for binding at the $3_{10}$ helix to cause potential alterations of conformation and/or dynamics across the entire β4 -$3_{10}$ helix loop and the N-terminal end of the α-helix despite being distant from the binding site (Figure 4-19). It is also worth noting that many crystal structures of Ub show a number of water molecules around the interface between Ub and their bound UBD, this may well be the case for the Znf_A20 interaction with Ub which could be contributed by water molecules between this cluster of polar residues and Znf_A20[20,21,159].

### 4.3.9 Evaluation of the Mutation Based Restraints

The observed mutation based CSP, measured as a result of titrations using wt-Znf_A20 compared to those with the Znf_A20 mutants F20Y and V32K, gave rise to discrete patches of CSP. The observed CSP are wholly consistent with the presented structure; the F20Y mutant causes only small intra- and intermolecular CSP, intramolecular CSP were restricted to Phe20 and Tyr21, while the intermolecular CSP were localised to Ub- Arg54 and Leu50. In the case of the V32K mutant, larger CSPs were observed, but were similarly localised; the intramolecular CSP localised to residues adjacent to the mutation site including Cys33 and Lys35 while the intermolecular CSP are localised to Ub- Asp24, Asp52, Arg54, and Ile61. In both cases mapping the intra- and intermolecular CSP onto the structure of the Znf_A20 complex shows good correlation between the mutation site, the intramolecular CSP, and the intermolecular CSP, which are fully consistent with the observed structure (Figure 4-20).

Mutation based CSP resulting from the Ub mutant K48C covalently attached to the spin label MTSL also demonstrated localised intermolecular CSP which again are consistent with the solved structure (Figure 4-21).

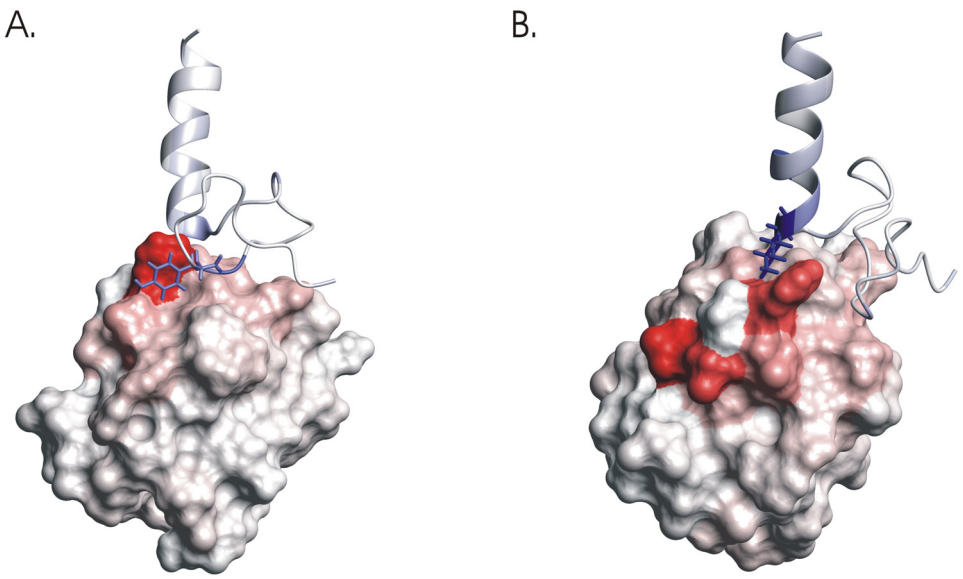



**Figure 4-20: Mutation based CSP mapped onto the complex structure of; (A) F20Y (B) and V32K. Intramolecular CSP (Blue) were mapped onto a ribbon diagram of the Znf_A20, intermolecular CSP (Red) were mapped onto the surface of Ub. In all cases the side chain of the mutated residue is shown.**

The large CSP observed for Asp24 and Asp52 with the Znf_A20-V32K mutant, are in the right vicinity of the protein in terms of the orientation of the two proteins relative to one another, however their distance from the mutation site is too great for a direct contact (at around 11 Å). However given the sensitivity of chemical shifts to their environment, these longer range effects may be understandable considering the introduction of a new surface positive charge in the V32K mutation as both of the perturbed aspartate residues carry a negative charge and may experience long range charge-charge interactions with the new lysine residue. The 'point' charge of Lys32 would be around 15 Å from that of the two Asp residues, although this is much too long for a salt bridge a weak charge-charge interaction remains which is worth around 1-2 kJ/mol which may sufficient to cause a CSP in these residues[160,161]. The replacing of a hydrophobic Val side chain with a positively charged lysine would also be expected to alter the protein-solvent interface affecting water structure and retention time at the protein surface which may have a knock on effect on the chemical shifts of the other solvent exposed residues[159,162].

### 4.3.10 Evaluation of the PRE Restraints

Paramagnetic relaxation enhancement (PRE) measurements were made using the spin label MTSL covalently attached to the K48C Ub mutant. The distance restraints derived from PREs made up the majority of the intermolecular restraints during modelling. However large distance boundaries were imposed on the restraints to account for inaccuracies in the measurements and the quality of the data. Despite this the PRE data was invaluable in orientating the two molecules during docking. The MTSL-K48C-Ub bound to Znf_A20 resulted in localised attenuation of residues corresponding to those close in space to the approximate position of the spin label which drop off rapidly with distance (Figure 4-21). A plot of PRE measurement against structure measured distance gives a correlation close to those seen in other systems. However despite the data being in good agreement with the calculated structure the correlation is poorly defined due to the inherent inaccuracies in the measurements.

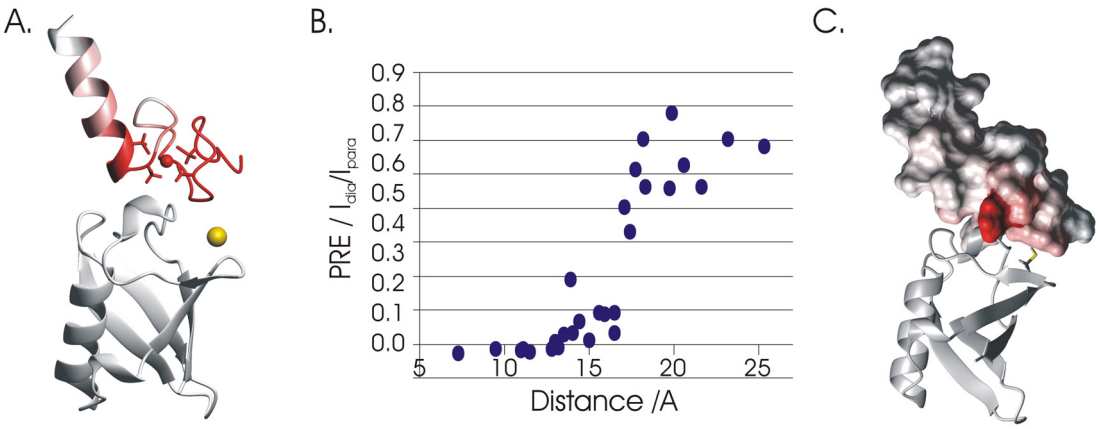



**Figure 4-21: (A) large PRE effects plotted onto the ribbon diagram (Red) of the Znf_A20 in complex with Ub, the approximate position of the spin label is illustrated as a yellow sphere[152]. (B) Plot of Measured PRE against mean distance in the structural ensemble. (C) Intermolecular CSP caused by the MTSL-K48C-Ub mutant, plotted onto the surface of Znf_A20 in complex with Ub, the side chain of the mutated residue is shown.**

### 4.3.11 Predicting Ubiquitin Binding Znf_A20 Domains

The consistency of the observed ubiquitin binding residues in the Znf_A20 domains of Rabex-5 and ZNF216 is a good basis on which to predict whether other Znf_A20 domains of currently unknown functions have the ability to bind to Ub.

Several of those residues identified as being common characteristics of Znf_A20 domains are residues which form key contacts with Ub for example Znf_A20-Ser31, which forms the key interactions with Ub-Asp58 is highly conserved across Znf_A20 domains and the di-aromatic motif of Phe20 and Tyr21 are similarly highly conserved. Other conserved residues identified as key for the interaction with Ub include Cys30, Asn23, and Val32. Cys30 is also required for coordination of the structural zinc ion and is therefore completely conserved. Asn23 is conserved in around half of the identified Znf_A20 being replaced in the remaining cases by Ser or Thr. Val34 on the other hand is not conserved, however as has already been demonstrated here, mutations at this position are tolerated.

**Table 4-5: Conservation of key Znf_A20 residues required for binding to Ub through a selection of identified Znf_A20 domains from UniProt. Identical (Red) and chemically similar (Blue) residues are highlighted.**

| (ZNF216) | Phe20 | Tyr21 | Asn23 | Ser31 |
|---|---|---|---|---|
| **Rabex-5** | Tyr | Tyr | Asn | Ser |
| **Tex27** | Phe | Trp | Ser | Ser |
| **PEM6** | Phe | Tyr | Asn | Ser |
| **SAP3** | Phe | Tyr | Asn | Ser |
| **ZNF123** | Phe | Tyr | Asn | Ser |
| **SAP6** | Phe | Tyr | Asn | Ser |
| **SAP7** | Phe | Phe | Ser | Ser |
| **SAP11** | Phe | Pro | Asn | Gln |
| **SAP12 \| A20 2** | Phe | Phe | Ser | Ser |
| **TNFaIP3 \| A20 3** | Gln | Asp | Thr | Ser |
| **TNFaIP3 \| A20 4** | Tyr | Phe | Thr | Thr |
| **TNFaIP3 \| A20 6** | Asn | Ile | Ala | Met |

Znf_A20 domains were examined for the presence of key binding residues including conservation of (ZNF216 residues) Ser31, Phe20, Tyr21, and Asn23. On the basis of these residues the majority of examined Znf_A20 domains carry the key residues for Ub binding, with all but SAP11 and the 3$^{rd}$ and 6$^{th}$ Znf_A20 domains of TNFaIP3 containing the di-aromatic motif. While all positions show variation, the

properties are largely maintained and the Znf_A20 would be predicted to bind to Ub. For example previous reports have stated that the Znf_A20 domain of TEX27 has the ability to bind ubiquitin[123], in TEX27 Asn23 is replaced by a Ser, this is the case in a number of Znf_A20 domains supporting the notion that it is the polar property of Asn23 which is important.

Given that the three residues which are key to ubiquitin binding are also highly conserved among other Znf_A20 domain, it can be suggested that Ub binding may well be a general property of Znf_A20 domains. This includes those for a family of stress associated proteins (SAP) found in plants, which like ZNF216 also contains a Znf_AN1 domain and currently have no identified function, but are linked to stress tolerance in plants[120,163]. The only Znf_A20 which appear unlikely to interact with Ub are those from TNFaIP3 with the exception of Znf_A20 number 4, which has previously been reported to bind to Ub and has sequence characteristics consistent with binding to Ub.

### 4.3.12 Lack of Chain Linkage Specificity in Poly-Ub Binding

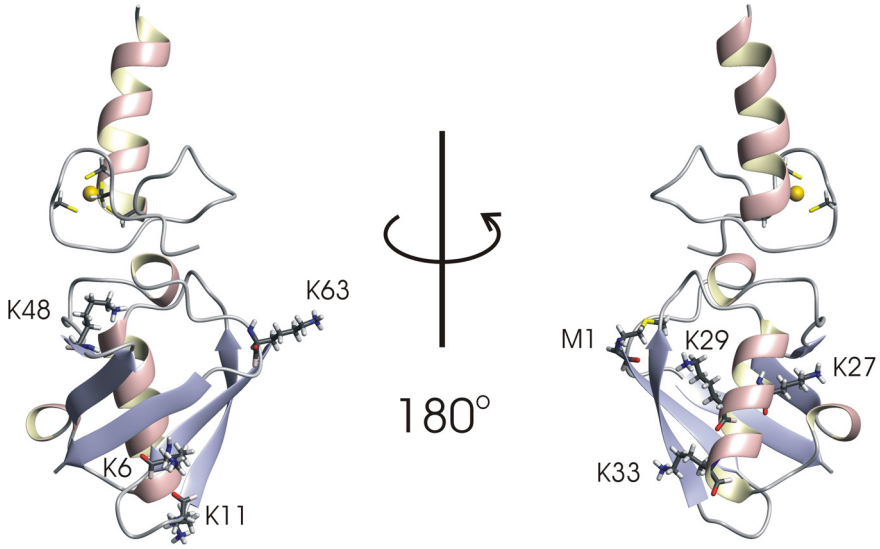



**Figure 4-22: Proximity of the seven Lysines, and N-terminus of Ub to the Znf_A20 binding site.**

The structure of the Znf_A20 in complex with mono-Ub can be used to examine the possibility of interaction with any of the seven lysines of Ub used to construct poly-Ub chains *in vivo* which is key to the versatile function of Ub. Initial examination of the structure shows only Lys48 in close proximity to the Znf_A20

binding patch suggesting that the Znf_A20 is unlikely to show specificity for poly-Ub linkages (except Lys48) based on direct interaction with the isopeptide linkages of the poly-Ub chain (Figure 4-23). However as has been demonstrated in other systems[17], that does not rule out the existence of secondary contacts away from the main binding site when interacting with poly-Ub units.

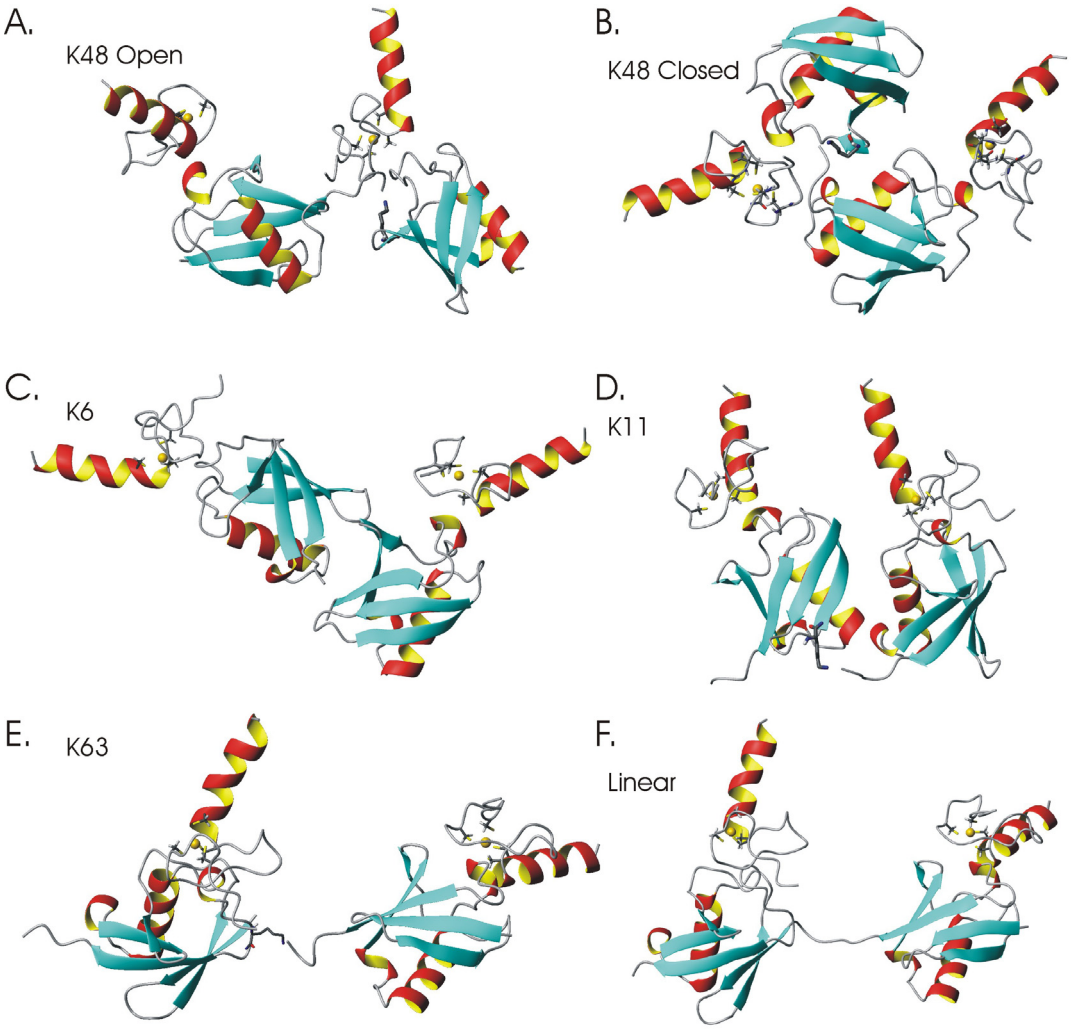



**Figure 4-23: Model of the Znf_A20 in complex with di-Ub linked via; (A and B) Lys-48 in its open and closed conformation (C) Lys6 (D) Lys11 (E) Lys63 (F) linear di-Ub. Models are based on the complex presented here and RPDB structures 1F9JK, 2PEA, 2XK5, 2XEW, 2JF5and 2W9N.**

In the case of Lys48 the C-terminal di-glycine tail of the distal Ub attached to Lys48 is predicted to be in close proximity to the Tyr21 side chain, opening the possibility for clashes or for favourable contacts between the distal Ub and the Znf_A20 bound to the proximal Ub. Alternatively this may simply restrict the available conformations of the poly-Ub unit when Znf_A20 is bound to the proximal

Ub. Simply replacing the mono-Ub unit in the Znf_A20 complex with di K48-Ub in the open and closed conformations of di-Ub or with tetra-Ub (RPDB IDs: 2PEA, 1F9JK and 2O6V) appears to support this showing significant clashes between the distal-Ub and Znf_A20 in the open structure with only a few potential minor clashes in the closed conformations of di- and tetra- Ub (Figure 4-23).

Similar analysis involving the available RPDB structures of Lys6, Lys11, and Lys63 linked and linear di-Ub (RPDB ID: 2XK5, 2XEW, 2JF5 and 2W9N) reveal no clashes or clear interaction with either Ub unit, the isopeptide linkage or between the two Znf_A20 molecules occupying sites on adjacent Ub units (Figure 4-23).

Examining static structures such as those deposited in the RPDB may be misleading and should be treated with care. This is highlighted when examining di-Lys63-Ub; in the available free structures it is hard to imagine restrictive clashes which may prevent binding of Znf_A20 to either Ub moiety. However examination of the structure of di- Lys63-Ub in complex with the NZF domain of TAB2 (RPDB ID: 2WWZ and 3A9K) it is clear that although the two Asp58 centred binding sites are exposed and can be occupied in isolation by Znf_A20 without steric clashes, when Znf_A20 is bound at sites on adjacent Ub moieties the two Znf_A20 clash presumably preventing occupation of both sites simultaneously. In contrast the Lys48 di-Ub structure in complex with the second UBA domain of hHR23A appears to maintain the Lys48 di-Ub moiety in a conformation which minimises clashes with the Znf_A20. The concept of UBD interactions directly modulating the affinity of another binding to a separate site is a novel one, an idea which could expand further the complexity of Ub signalling [2,15]. The concept of UBDs modulating binding by other UBDs is discussed in detail in later chapters (Chapter 9).

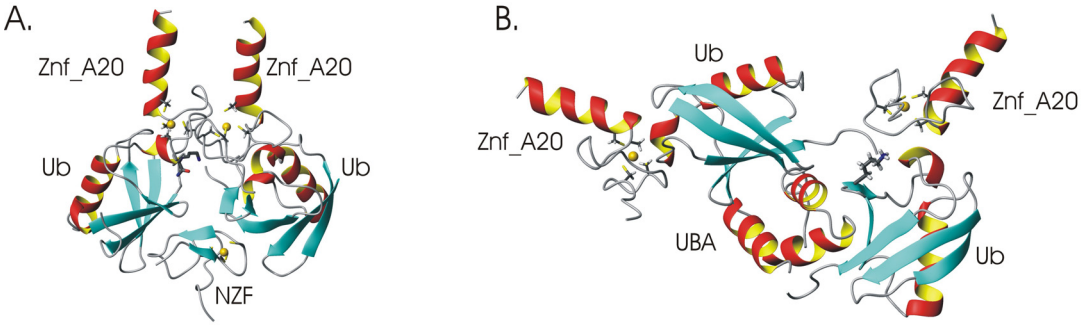



**Figure 4-24: (A) Znf_A20 complex with Lys63 linked di-Ub in complex with the NZF domain of TAB2 (B) Lys48 linked di-Ub in complex with RAD23. (RPDB ID:2A9K and 1Z06).**

Interestingly a number of Znf_A20 domains, including Rabex-5, TEX27 and TNFaIP3, are reported to show little or no binding to Lys63 or Lys48 poly-Ub chains, the distance of Lys63 from the Znf_A20 binding patch makes this difficult to comprehend[113]. The closest residues to Lys63 in ZNF216 which has been shown here and by others to bind both Lys63 and Lys48 poly-Ub[113] are Arg25 and Pro24 at a minimum of 10 Å away, these are replaced by smaller residues in Rabex-5, TEX27 and TNFaIP3 making steric clashes unlikely. Interference with the Lys48 linked poly-Ub linkage may be more easily explained by considering the proximity of the di-aromatic motif to the Lys48 site. In TEX27 Tyr21 of ZNF216 is replaced by the considerably larger Trp residue, given the proximity of the Tyr21 side chain to the C-terminal di-Gly of the distal Ub, the larger Trp would likely interfere considerably and prevent binding to the proximal-Ub in Lys48 linked molecules. However this is not the case for Rabex-5 or TNFaIP3 suggesting further insight is required to properly understand poly-Ub binding by Znf_A20 domains.

### 4.3.13 Specificity for Ub Versus UBL Domains

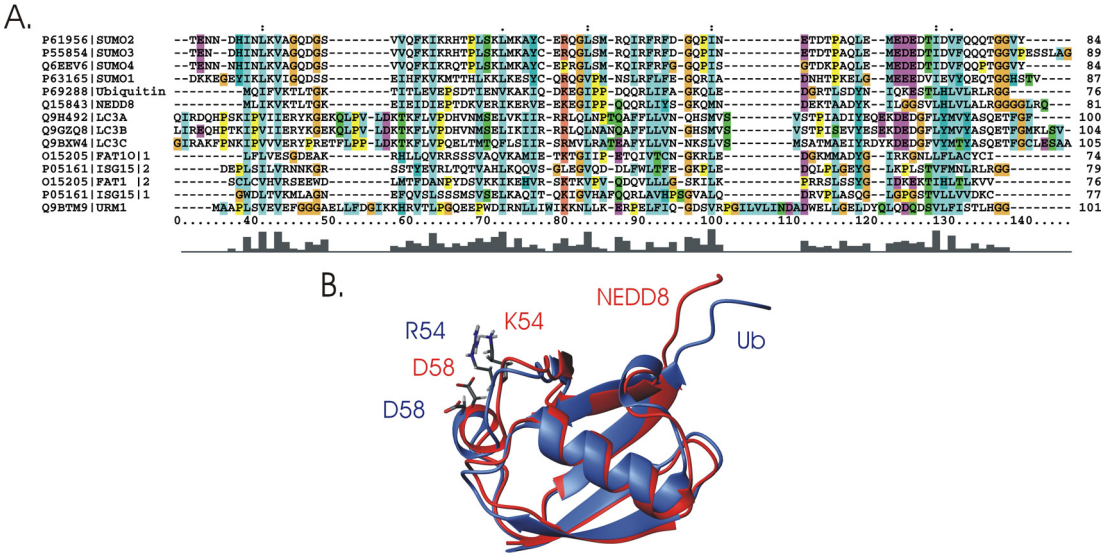



**Figure 4-25: (A) Sequence alignment of Ub with UBL proteins (B) Structural overlay of Ub with the UBL NEDD8.**

Ub is a highly conserved molecule with only 4 residues out of 76 showing any species variation[14,15], however there are around 100 proteins in humans which are related to Ub. These ubiquitin like (UBL) proteins share high structural and sequence similarity with Ub (Figure 4-25). Despite this UBDs generally show high

specificity for Ub[14,15] and the membrane-based binding assays reported here suggest that the Znf_A20 is no exception. In order to investigate the molecular basis for this, Ub residues showing high CSP during NMR titrations with Znf_A20 were examined for their conservation in UBL proteins (Table 4-6).

**Table 4-6: Conservation of key Znf_A20 binding residues in UBL proteins. Residues that are directly involved in binding are shaded green, identical (RED) and similar (Blue) residues are highlighted. Key: N8 = NEDD8, S1/2/3/4 = SUMO 1/2/3/4, Ia/b = ISG15 (UBL domains 1 and 2), F1 a/b = FAT10 (UBL domains 1 and 2) UM = Ubiquitin-related modifier 1, L3 A/B/C = LC3A/B/C.**

| Residue | N8 | S1 | S2 | S3 | S4 | Ia | Ib | F1a | F1b | UM | $L3_A$ | $L3_B$ | $L3_C$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| E/D 24 | E | K | S | S | S | S | A | K | A | W | S | S | T |
| T55 | T | T | T | T | K | P | P | S | H | L | I | I | H |
| R54 | K | H | D | D | D | V | L | R | K | L | P | P | T |
| D58 | D | E | Q | Q | Q | S | E | S | D | E | I | V | I |
| Y59 | Y | L | L | L | L | Q | Y | Y | Y | L | Y | Y | Y |
| N60 | K | G | E | E | E | G | G | G | G | D | E | E | R |
| I61 | I | H | H | H | H | L | L | I | I | L | K | K | K |
| Q62 | L | E | E | E | E | G | K | D | R | Q | D | D | D |
| Identity /8 | 5 | 1 | 1 | 1 | 0 | 0 | 1 | 3 | 3 | 1 | 1 | 1 | 1 |
| Similarity /8 | 6 | 5 | 6 | 6 | 5 | 3 | 3 | 6 | 5 | 6 | 5 | 5 | 4 |

No UBL proteins possess all 8 of the examined residues; however several possess up to 5 of the selected residues, in particular NEDD8 and FAT10 UBL domain 1 and 2. On the basis of sequence and structural similarity it is perhaps surprising that NEDD8 and the UBL of FAT10 do not interact with Znf_A20 given the conservation of properties across the majority of the 8 selected residues. Both NEDD8 and the 2nd UBL domain of FAT10 conserve the key Asp58 residue while in the 1st UBL of FAT10 this is replaced with a Ser which like Asp58 can form a hydrogen bond. However Ser has a single hydrogen bond donor compared to the two of Asp, also the extra carbon atom of the Asp chain may position the available

hydrogen bonds better than the Ser. However with the exception of the 1st UBL of FAT10 none of the UBL proteins possess the Arg54 side chain shown here to be key to the Znf_A20 interaction, in both NEDD8 and the 2nd UBL of FAT10 this residue is switched to lysine which maintains the positive charge. However it is the unique property of the arginine guanidinium group and its ability to form both polar and hydrophobic stacking interactions which is essential to the interaction with Znf_A20 domains which the lysine replacement lacks, and which may be key to the loss of interaction with UBL proteins.

Despite the high sequence and structural similarity between Ub and UBL proteins none maintains the two key residues namely Arg54 and Asp58 which are central to the interaction and potentially confer the Znf_A20 specificity for Ub.

## 4.4 Conclusions

The proposed interaction of the Znf_A20 with Ub is central to the function of ZNF216 as a proteasomal shuttling protein, warranting a more detailed study of the molecular detail of the interaction.

From a combination of biochemical and biophysical techniques the formation of a high affinity complex between Ub and the Znf_A20 domain of ZNF216 has been confirmed. The measured $K_d$ of ~10 μM is relatively strong for UBD mono-Ub interactions, marginally stronger than the reported value for the Znf_A20 of Rabex-5[17,20,21] with only the UBP of isopeptidase T showing a stronger interaction[17]. The interaction utilises the Asp58 centred patch although this residue is not sufficient for the interaction alone. The interaction shows no clear specificity for poly-Ub chain linkages in membrane-based binding assays or by ESI-MS although the structural investigation suggests the potential for interactions with the isopeptide linkage in Lys48-poly-Ub.

The NMR solution structure of the Znf_A20 non-covalent complex with Ub has been reported, the structure is consistent with the crystal structure reported for the Ub complex with Rabex-5. The key interactions involve hydrogen bonds between Ub-Asp58 and Znf_A20-Ser31 and Tyr21, and between Ub-Arg54 with Znf_A20-Phe20 and Cys30, resulting in the burial of around 771 ± 65 Å$^2$ of solvent exposed surface area.

Interestingly in the Rabex-5 study the Tyr at ZNF216 position Phe20 is proposed to form a hydrogen bond to Glu51, the authors go onto to propose that the

loss of this hydrogen bond would reduce the affinity for Ub, however in ZNF216 the replacement of Tyr for Phe and the loss of this proposed hydrogen bond does not appear to affect the affinity for Ub which appears higher affinity in ZNF216 with ITC measurements recording ~10 μM compared to 20 μM recorded for the Znf_A20 of Rabex-5[20,21]. The restoration of the Tyr in the F22Y mutation also causes no measurable CSP for Ub-Glu51 suggesting that the proposed hydrogen bond seen in the crystal does not hold in solution, and may be an artefact of the crystal environment.

The structure also provides insight into the apparent lack of specificity for poly-Ub with only Lys48 in close proximity to the Znf_A20 binding patch. The specificity for Ub over UBL proteins can also be explained on the basis of the presented structure due to the lack of a combination of Arg54 and/or Asp58 equivalents in UBL proteins.

It is worth noting that in this study the Znf_A20 and Ub originate from two different species (rat and yeast), while there is sequence variation with species for both Ub and Znf_A20 those residues making intermolecular contacts and in the immediate vicinity of binding interface are the same in both species. In the case of the Znf_A20 the ordered region of the Znf_A20 is identical across species. In the case of Ub the only mutated residue of significance is position 24 which in yeast Ub is an Asp, in other species including rat this a Glu, however this not a significant change, does not alter the structure of Ub, and is not in contact with the interface between the two proteins. The species variation is therefore not expected to affect the structure in this case.

The structure presented here combined with the previously solved structure of the Rabex-5 Znf_A20 complex with Ub and the sequence identity between Znf_A20 domains, provides evidence that Ub binding may be a common characteristic of Znf_A20 domains and provides further insight into the molecular detail of this important interaction.

# Chapter 5

# 5 Examination of the Full Length ZNF216 Protein

## 5.1 Introduction

The ZNF216 protein is classed as an atrogene with a potential role as a proteasomal shuttling factor which facilitates the degradation of muscle proteins by the proteasome during muscle atrophy. In order to better understand the molecular details of ZNF216 function the structure of the Znf_A20 domain of ZNF216 has been solved (Section 3.3) and its interaction with its key binding partner Ub has been characterised (Chapter 4).

As is common with NMR based studies, investigation of the Znf_A20 has been carried out on the isolated Znf_A20 domain. However the behaviour of an isolated domain can be subtly different in the context of the full length protein, inter-domain interactions can regulate domain function or alter the fold when compared to the domain in isolation. One such example of this is RAD23 which possess both UBA and UBL domains which interact in the full length protein, an interaction which must first dissociate before the UBA domain can bind to Ub[164]. Studying the properties and interactions of any isolated domain are ideally complemented by studies of the domain in the context of its full length protein.

### 5.1.1 Znf_A20 and Znf_AN1 Domain Containing Proteins

The ZNF216 protein possess two identified domains, an N-terminal Znf_A20 domain, which has been studied above (Sections 3 and 4), and a C-terminal Znf_AN1 domain. The large interdomain region of >100 residues has no significant homology with known domains and is predicted to be highly disordered. The Znf_AN1 domain has previously been studied structurally by the RSGI but very little is known about its function. In ZNF216 there is evidence that the Znf_AN1 mediates an interaction with TRAF6 although this is yet to be confirmed[117]. Previous reports have proposed that the Znf_A20 and Znf_AN1 domains of ZNF216 can form an inter-domain interaction which facilitates the formation of homodimers and homotetramers, with homodimers being the main form in gel filtration studies [117].

The pairing of Znf_A20 and Znf_AN1 domains within proteins is common, with the majority of proteins containing Znf_A20 domains also containing AN1 domains, including AWP1 a human regulator of NF-κB, PEM-6 an embryonic partition protein from ascidia, and a large family of proteins linked to improved environmental stress tolerance in plants[120,121,163,165]. In studies involving SAP 8 the reported interdomain interaction between Znf_A20 and Znf_AN1 was also observed, further highlighting the need to investigate this proposed interaction. The function of Znf_A20 and Znf_AN1 containing proteins is currently unclear but there is a clear conservation of the pairing of these two domains across species[120].

### 5.1.2 Summary and Aims

The importance of studying the Znf_A20 domain in the context of the full length protein may be key to understanding the molecular details of ZNF216 function. This is highlighted by the proposed interaction between the Znf_A20 and Znf_AN1 domains of ZNF216 which if proven to be the case could serve to modulate Ub binding by the Znf_A20 domain in a fashion analogous to the RAD23-UBA-UBL interaction[117,164]. In order to investigate the potential for a Znf_A20 interaction with the Znf_AN1 and any potential effects this may have on the ability of the Znf_A20 to bind to Ub, the full length ZNF216 protein was investigated by a number of biophysical approaches and by NMR spectroscopy in order to gain insight into the structure and potential function of ZNF216.

## 5.2 The ZNF216 is Monomeric in Solution with an Extended Conformation

### 5.2.1 Sequence Alignments

The ZNF216 protein is 213 amino acids in length and has only two domains predicted by SMART[89], this was confirmed by a BLAST search which showed no regions with high sequence similarity to known domains. Residues 11-35 are predicted to form an A20-type zinc finger (Znf_A20) domain consistent with previous reports[113,117], while residues 154-191 are predicted to form an AN1-type zinc finger (Znf_AN1). The remainder of the protein is predicted to be intrinsically disordered. The ZNF216 protein sequence was aligned with available orthologs identified within the UniProt databases[85] and shows a high degree of sequence

similarity across the entire protein across the available species (Figure 5-1). Only short regions show high variability which is reduced when only mammalian sequences are considered. In the 5 available mammalian species only a limited number of residues show variability and the two predicted domains show no changes. Considering that only 61 of the 213 residues are predicted to fold into known domains, and that no other known motifs or binding sites have been identified, such high sequence conservation is perhaps surprising without obvious evolutionary pressure to remain invariant.

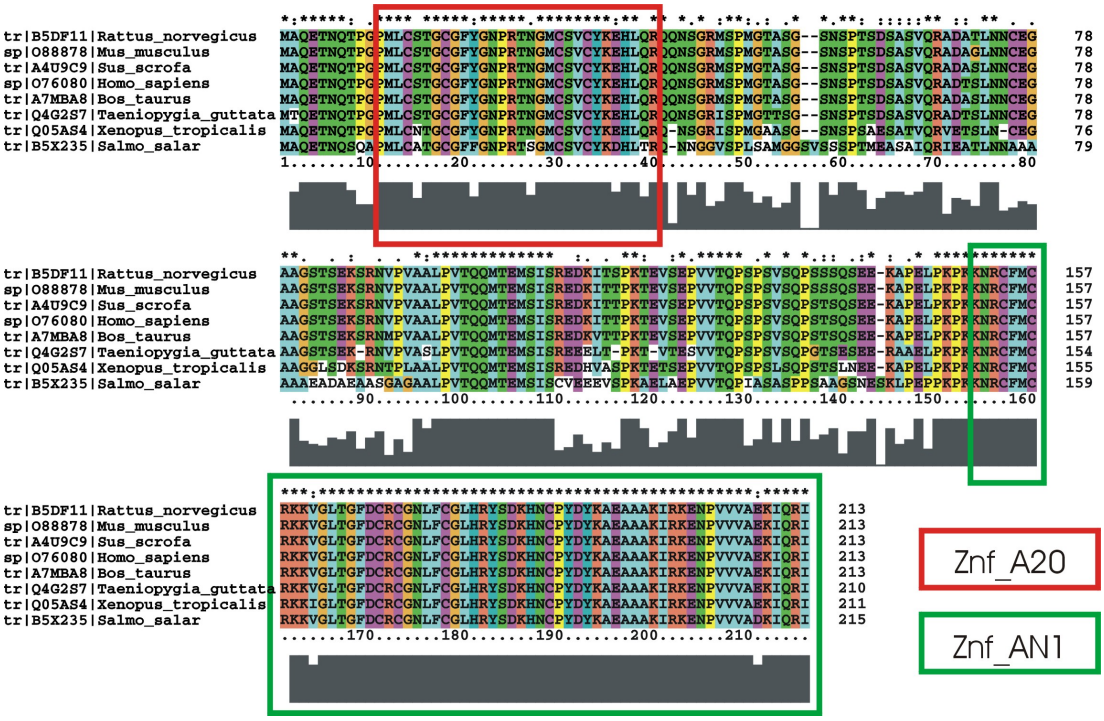



**Figure 5-1: Sequence Alignment of ZNF216 from multiple species, the two identified domains are highlighted.**

Despite the relatively high sequence conservation the large region separating the two domains presents no known domains. Analysis of the ZNF216 protein sequence using the programs DisProt[92], IUPred[93,166], DISpro[94], and DISOPRED2[95] predicts low disorder for the two structured domains, except for DisProt which only predicted the Znf_AN1. All predicted disorder in the linker region between the two domains consistent with previous reports[113,117]. However DISpro and DISOPRED2 predict low disorder in the residue range 90-109 and in these cases the value is marginally above the threshold for a disordered region, suggesting the potential for

structure in this region (Figure 5-2). Interestingly this residue range corresponds to a region of high cross-species sequence conservation in ZNF216 (Figure 5-1).

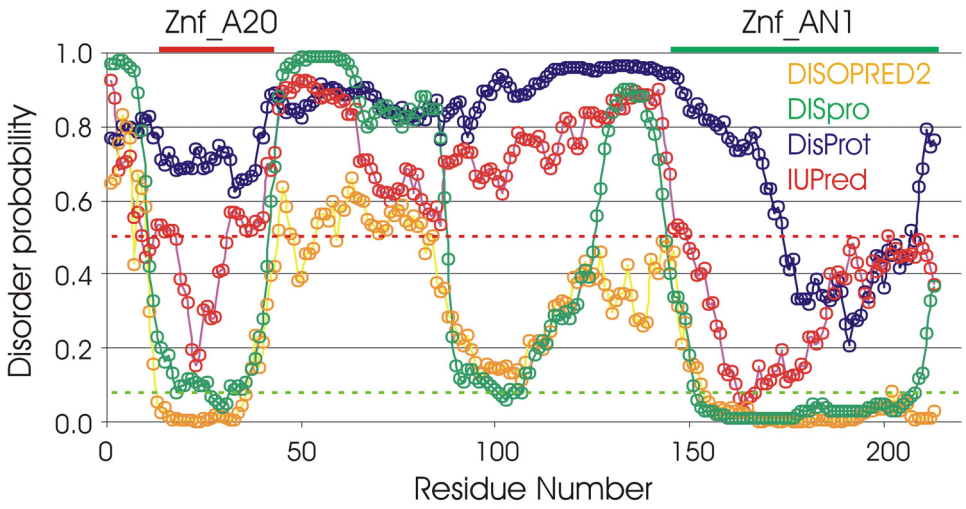



**Figure 5-2: Disorder prediction for the ZNF216 protein using DisProt[92](Blue), IUPred[93,166](Red), DISpro[94](Green) and DISOPRED2[95](Yellow); Disorder threshold values are indicated for DisProt and IUPred (Red) and DISpro and DISOPRED2 (Green ) the position of the Znf_A20 and Znf_AN1 domains are highlighted.**

The ZNF216 sequence was also analysed using PROSCAN[90] and ELM[91] to search for potential domains or functional motifs. Neither identified homology with known folds but did predict potential phosphorylation sites for PKC at positions 44, 84, 114 and 181, as well as potential casein kinase II phosphorylation sites at positions 82, 107, 137 and 164, although there are no reports of phosphorylation of ZNF216 in the literature.

The pairing of Znf_A20 domains and Znf_AN1 domains is common but interestingly it is not only the pairing which is common but the domain architecture is also strikingly similar with the Znf_A20 located within the first 50 residues of the protein followed by a long disordered linker with the Znf_AN1 typically at the extreme C-terminus (Figure 5-3). The two domains show good conservation across the examined proteins. However the linker separating these domains is variable between different proteins in length and sequence, but appears to have similar sequence compositions, with the inter-domain regions of all those examined, predicted to be highly disordered.

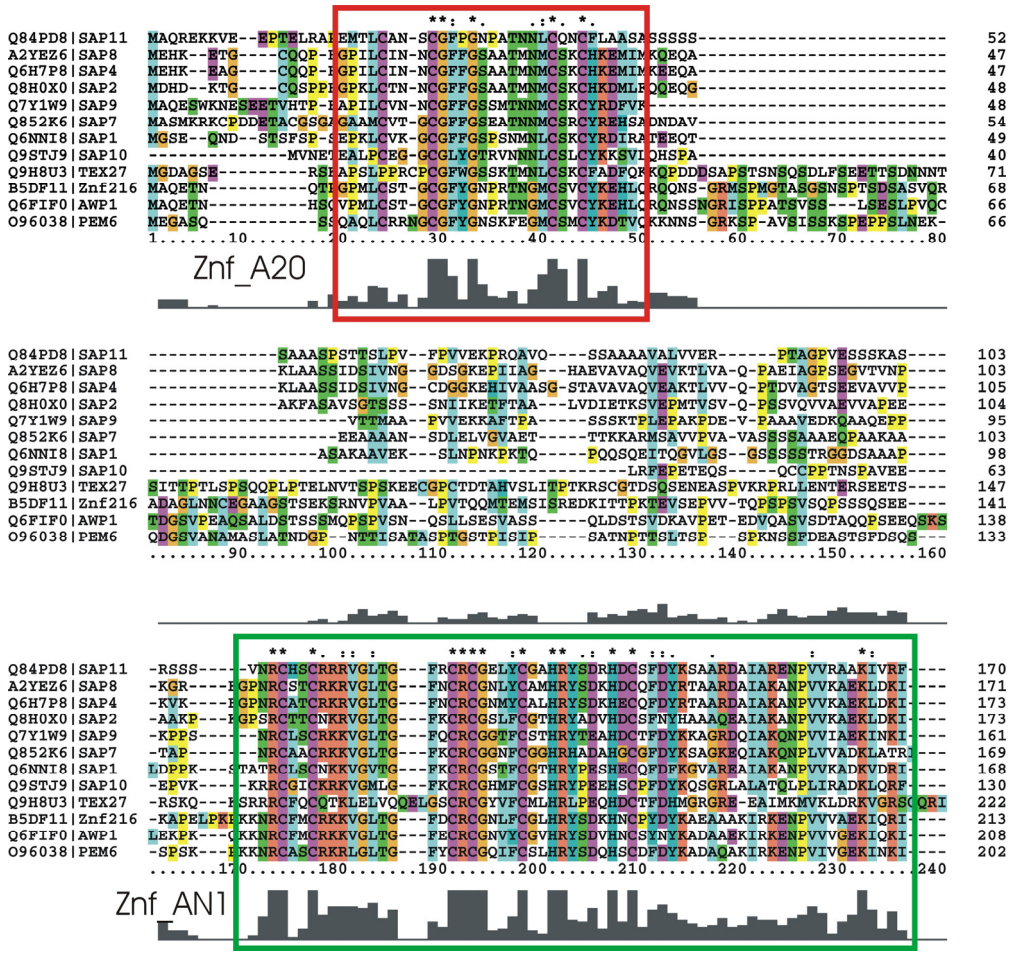



**Figure 5-3: Sequence alignment of Znf_A20 and Znf_AN1 containing proteins the two domains are highlighted.**

### 5.2.2 ZNF216 Protein is Monomeric in Solution

In order to examine the oligomerisation state of ZNF216, dilution ITC, ESI-MS, and AUC were recorded. ITC results gave some evidence for endothermic heat pulses as an indication of dissociation but results were inconsistent, giving low quality data in all cases with no consistency in the magnitude of the response, as a result ITC was deemed inconclusive as it was unable to distinguish between specific oligomeric dissociation, non-specific oligomer dissociation, or machine error.

Native ESI-MS recorded a native mass of 23421.05 ± 0.4 (Figure 5-4); the mass of the ZNF216 apoprotein is 23232.1, the expected holoprotein contains three zinc ions each with a mass of 65.38 (one bound to Znf_A20, two to the Znf_AN1). The recorded mass of 23421.05 is therefore consistent with the expected holoprotein in its monomeric state. No evidence was seen for dimeric or tetrameric species of ZNF216. Interestingly the charge-state distribution of ZNF216 is clearly split into

two distributions (Figure 5-4) with a potential third between them. This suggests the presence of two or three species during ionisation, one with a higher average charge-state distribution and a larger charge-state range than the other, suggesting a more unfolded state. The average charge-states of the two clear species are estimated to be 11 ± 0.4 and 22.6 ± 0.3 corresponding to a predicted solvent exposed surface area of 11903 ± 500 and 34536 ± 350 $Å^2$ respectively by the method of Kaltashov *et al*[46]. This could reflect an open and closed equilibrium in solution, possibly due to the proposed Znf_A20 inter-domain interaction with Znf_AN1. However as discussed earlier the proposed use of ESI-MS to estimate solvent exposed surface area may be influenced by a number of factors and should not be used without other evidence.

In order to investigate the solution oligomerisation state and shape of the ZNF216 protein SEC and AUC were performed. SEC was performed using a HiLoad™ Sephacryl™ S-100 HR gel filtration column pre calibrated with molecular weight standards as described previously (Section 3.2.2.1). SEC was recorded at 300 and 600 μM and in both cases gave a clean peak with an apparent molecular weight of ~70 kDa (Figure 5-4) representing a mass between a dimer and a tetramer. This is somewhat consistent with previous reports using SEC where dimeric ZNF216 was identified as the dominant species with some tetramer[117]. However in the original publication by Huang *et al* three distinct peaks were reported for ZNF216, while in this study we observe only a single peak and as discussed previously, SEC should not be used to measure mass rather more accurately to measure Stoke's radius. The recorded Stoke's radius for ZNF216 using SEC is 34.5 ± 1.5 Å which is consistent with a mass of around 70 kDa. However this could be the result of a highly extended protein of considerably smaller mass. The SEC gave identical elution volumes at 300 and 600 μM (Figure 5-4) suggesting the observed Stoke's radius is not the average of two species in rapid equilibrium. The gel filtration peak appears to have a small degree of asymmetry (Figure 5-4), this may indicate a small amount of interaction between the protein and column resin, an equilibrium between low populated species of smaller Stoke's radius (with a 20 Å minimum ) or small amounts of degradation. It is worth noting that although the ZNF216 is outside of the optimum separation range for the Sephacryl™ S-100 column, the ZNF216 elutes within the separation range of the column and within the range of the protein standards which show the anticipated correlation between Stoke's radius and elution volume suggesting that measurements within this range remain valid.

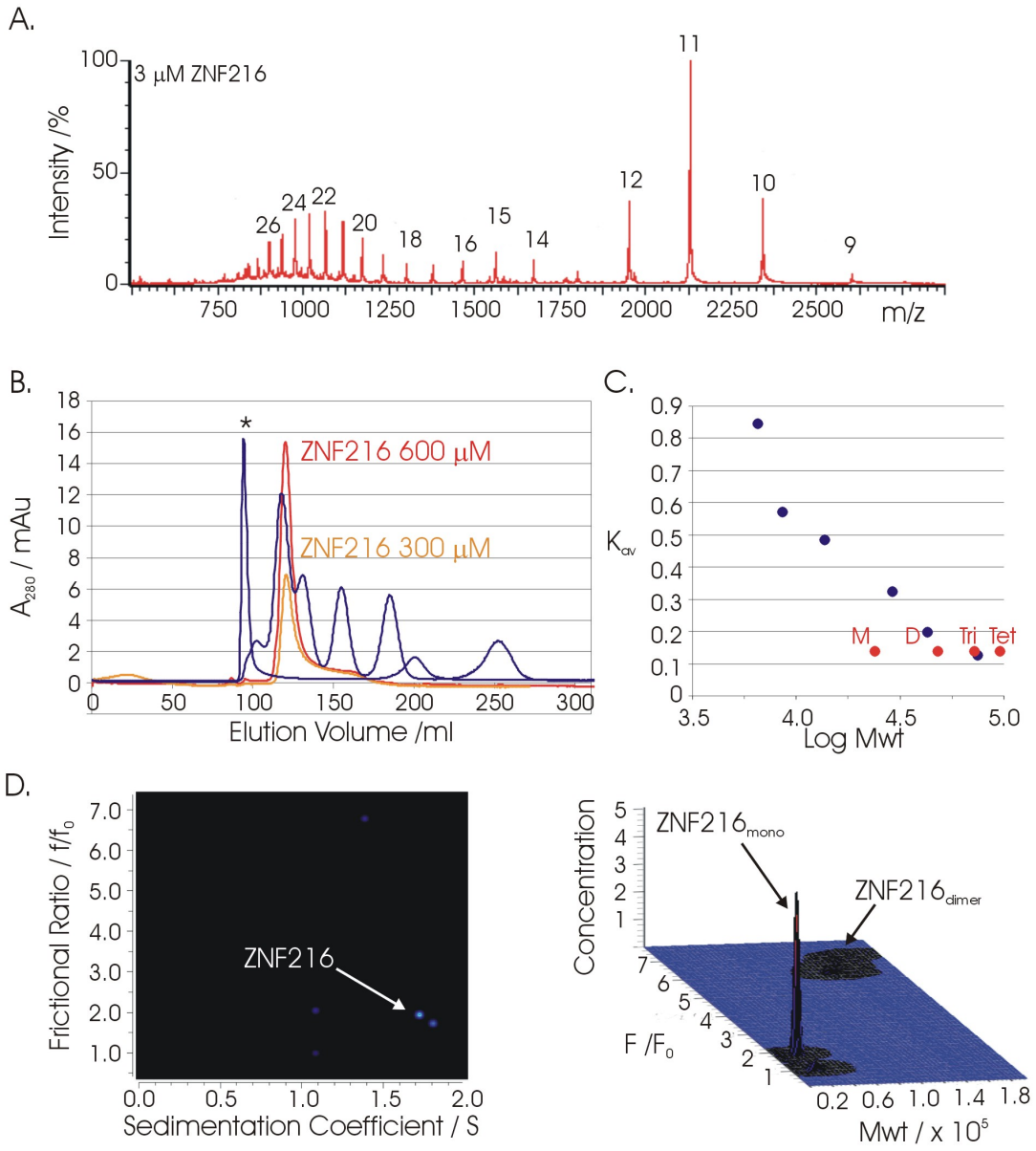



**Figure 5-4: (A) ESI-MS spectrum of ZNF216. (B) SEC elution profile of calibration proteins (Blue) and full length ZNF216 at injection concentrations of 300 (Orange) and 600 µM (Red) (* denotes blue dextran). (C) SEC $K_{av}$ vs LogMWt for ZNF216 monomer (M), dimer (D), Trimer (Tri), and Tetramer (Tet). (D) AUC analysis for the ZNF216.**

In order to resolve the solution molecular weight of ZNF216, AUC was performed by Dr Dave Scott (School of Biosciences) in the concentration range 0.1 mg/ml to 1 mg/ml (4-40 µM) using a Beckman XL-I analytical ultracentrifuge using interference optics recorded at 660 nm. AUC recorded a solution mass of around 24 kDa, consistent with monomeric ZNF216, however unlike the ESI-MS AUC gave a

single distribution around 24 kDa. The recorded frictional ratio of 2.0, and the observed sedimentation coefficient of 1.75 ± 0.1 S suggests an extended conformation. This is best illustrated by a comparison of the observed sedimentation coefficient to that estimated for globular proteins. The ZNF216 protein has a sedimentation coefficient well below that of globular proteins similar to that calculated for the open conformation of RAD23, suggesting a highly extended conformation. The calculated Stoke's radius for ZNF216 from the AUC for monomeric ZNF216 with a sedimentation coefficient of 1.75 ± 0.1 S equals 36 ± 1 Å consistent with that of 34.5 ± 1.5 Å from SEC. This supports the notion that the large apparent mass in SEC is again the result of a highly extended monomer rather that higher order oligomers.

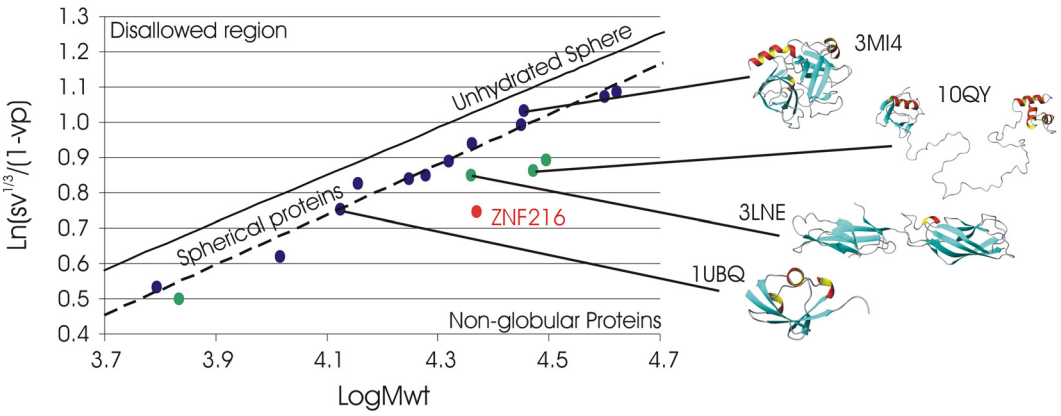



**Figure 5-5: Comparison of the observed sedimentation coefficient as a function of molecular weight for ZNF216 (Red) with that estimated for a set of globular proteins using HYDROPRO[139] (Blue). Predicted sedimentation coefficient for a set of non globular and highly extended proteins is also included (Green). A selection of representative proteins is illustrated.**

The AUC did detect some dimeric species although this was consistently very low even at the highest concentration used of 40 µM where it remained below 5% of the observed signal, this could be interpreted as non-specific dimerisation or a weak dimer with a $K_{dim}$ >1.4 mM.

SEC was also performed the on the Znf_AN1 domain of ZNF216 giving a predicted mass and Stoke's radius of ~10 kDa and 15.4 ± 1.5 Å. Although the recorded mass is high compared to the expected mass of 7.5 kDa, the estimated Stoke's radius for the published Znf_AN1 domain structure (RPDB ID: 1WFL with flexible termini removed) predicted using HYDROPRO[139] is 16 Å suggesting that

the Znf_AN1 is a well folded monomer and in good agreement with the published structure. However it is worth noting that the published structure is covers a different sequence range, with extra non Znf_AN1 residues included at the N-terminal end of the domain which are lacking in the construct used here. On the other hand the construct used here is extended C-terminally compared to the published structure. The published structure (RPDB ID: 1WFL) shows the N-terminal portion (which is missing in this construct) to be disordered and if this region is included, the estimated Stoke's radius increased up to ~22 Å. The considerably smaller measured Stoke's radius for the construct used here, which is of similar length to the published structure, may suggest that the extra C-terminal residues included here are well folded compared to the N-terminal region used in the published structure. This would be consistent with sequence analysis of Znf_AN1 which shows reasonably high sequence homology right up to the C-terminus (Figure 5-3).

### 5.2.3 The ZNF216 Znf_AN1 Domain Requires Zinc for Correct Folding.

In order to investigate the secondary structure of ZNF216 and the dependence of that structure on the presence of zinc, CD studies were performed on ZNF216 and the Znf_AN1 in the presence and absence of 5 mm EDTA. On first inspection ZNF216 gives a CD spectrum characteristic of a largely unfolded protein with strong negative ellipticity at 202 nm and little or no features indicating significant secondary structure with the negative ellipticity at 222 nm seen for Znf_A20 seemingly decreased in the ZNF216. An examination of the Znf_AN1 gives a clear explanation for this; the CD spectrum of the Znf_AN1 gives a strong negative ellipticity at 204 nm and an unusual positive ellipticity at 224 nm which cancels out the negative ellipticity at 222 nm observed for the Znf_A20 resulting in an overall weak negative absorbance at 222 nm. The overlapping characteristics of the CD spectra of the Znf_A20 and Znf_AN1 creates difficulty in assigning CD spectral characteristics to the remaining portions of the ZNF216, however there is a clear shift in the negative ellipticity at lower wavelength to 202 nm compared to 206 and 204 nm in the Znf_A20 and Znf_AN1 respectively. This small shift suggests a significant quantity of random coil character. This is consistent with the highly extended structure for ZNF216 proposed by SEC and AUC, and with secondary

structure predictions, which predict the linker region between the two domains to be highly disordered.

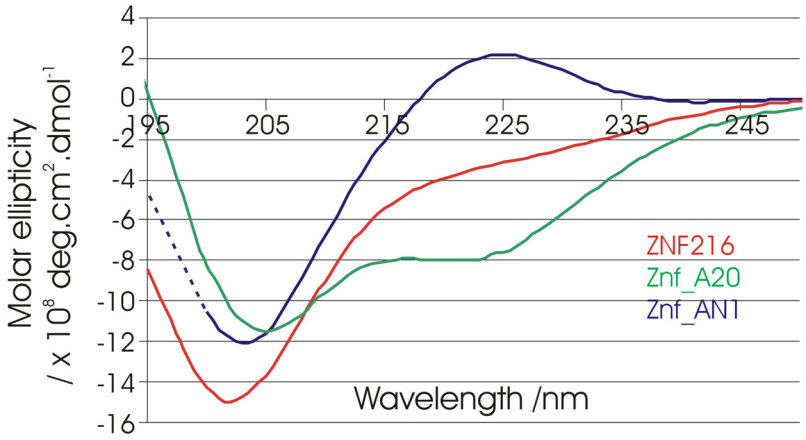



**Figure 5-6: CD spectra of ZNF216 (Red), Znf_A20 (Green) and Znf_AN1 (Blue)**

Treatment of the Znf_AN1 domain with 5 mM EDTA results in large changes in the appearance of the CD spectrum over the course of 2-3 hours, with the complete loss of the unusual positive band at 224 nm. A reduction in the intensity of the negative band at 206 nm coupled with a shift to 203 nm is also observed, consistent with unfolding of the domain in response to EDTA. Analysis of the ellipticity change as a function of time gave an EDTA induced unfolding rate which is similarly slow to that for Znf_A20 recording an approximate rate of $4.5 \pm 0.5 \times 10^{-4}$ $s^{-1}$($3.1 \pm 0.5 \times 10^{-4}$ for Znf_A20) (Figure 5-7). The unfolding transition has an isodichroic point at 212 nm suggesting a two-state transition. However there is potentially a small shift in the isodichroic point with time (Figure 5-7) suggesting the transition may not be two-state. This may not be surprising given the two adjacent zinc binding sites within the Znf_AN1.

The treatment of full length ZNF216 with 5 mM EDTA gave very little change in ellipticity due to the overlapping opposite ellipticity changes for the two domains. Examination of the small visible ellipticity change as a function of time revealed no clear isodichroic point and two clear transitions with the ellipticity at 222 nm initially increasing for the first 15 to 20 minutes, followed by an ellipticity decrease over a period of 2-3 hours (Figure 5-7). The calculated rate constants are unreliable largely due to the poor quality of the data, however it is clear from the change in ellipticity as a function of time that the Znf_AN1 transition is the slower of

the two transitions and that neither transition is drastically altered compared to the isolated domain.

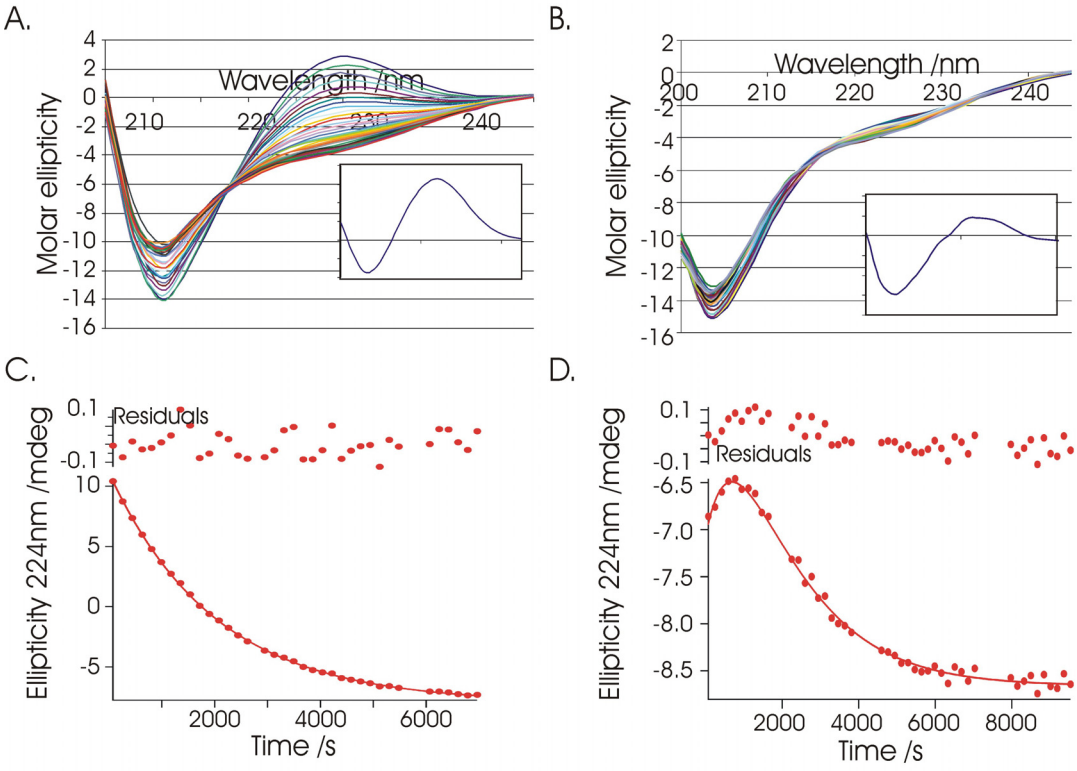



**Figure 5-7: EDTA induced unfolding monitored by CD for; (A) the Znf_AN1. (B) ZNF216. The derived difference spectra are included. (C-D) Ellipticity at 222 nm as a function of time with appropriate fits for; (C) Znf_AN1 (D) ZNF216.**

### 5.2.4 Summary

From a combination of bioinformatics and biophysical analysis the results strongly suggest that the ZNF216 is monomeric in solution at low concentration (~40 μM), and possess a highly extended conformation supported by SEC, AUC and CD analysis. The highly extended nature of the ZNF216 may suggest that the interaction between the Znf_A20 and Znf_AN1 of ZNF216 proposed by Huang *et al* [117]seems unlikely given the lack of solution phase data providing evidence for significant populations of a 'closed' conformation.

From analysis of SEC and CD data the Znf_AN1 domain forms a well-folded domain consistent with the published structure (RPDB ID: 1WFL). Like the Znf_A20, the Znf_AN1 requires zinc to maintain its correct fold and treatment of

either domain with EDTA results in a slow unfolding transition due to loss of the key structural zinc.

## 5.3 The Domains of ZNF216 are Independent and Separated by a Flexible Linker

### 5.3.1 The Znf_AN1 and Znf_A20 do not Interact

In order to investigate the proposed interaction between the Znf_A20 and Znf_AN1 domains of ZNF216, $^{1}$H-$^{15}$N-HSQC spectra were recorded of the full length ZNF216 protein and compared to that for the Znf_AN1 and Znf_A20 domains. For the $^{1}$H-$^{15}$N-HSQC spectra of the full-length ZNF216 a total of 196 of the 198 non-proline residue cross peaks were observed consistent with a single species. A number of resonances show good chemical shift dispersion consistent with the presence of folded domains (Figure 5-8). The $^{1}$H-$^{15}$N-HSQC is of good quality with little or no signs of the line broadening characteristic of molecules with high molecular weights. The observed signal intensity is consistent with the protein concentration suggesting that no significant signal intensity is lost to species which are 'invisible' due to extensive line broadening.

The $^{1}$H-$^{15}$N-HSQC spectra could be readily overlaid with those of the isolated Znf_A20 domain with no significant line broadening or CSP (CSP <0.025) (Figure 5-8). Thr62 was the only exception, which is the C-terminal residue in the Znf_A20 construct and would be expected to show considerable CSP. Given the sensitivity of the amide proton chemical shift to subtle changes in environment and its routine use in detecting even very weak protein-protein interactions[167] this strongly suggests that the Znf_A20 does not interact with the Znf_AN1 domain or any sequence within the linker region between the two domains. This is contrary to the work by Huang *et al* who measured an interaction between the Znf_AN1 and Znf_A20 of ZNF216 by both yeast 2 hybrid experiments and by GST-pull down experiments[117].

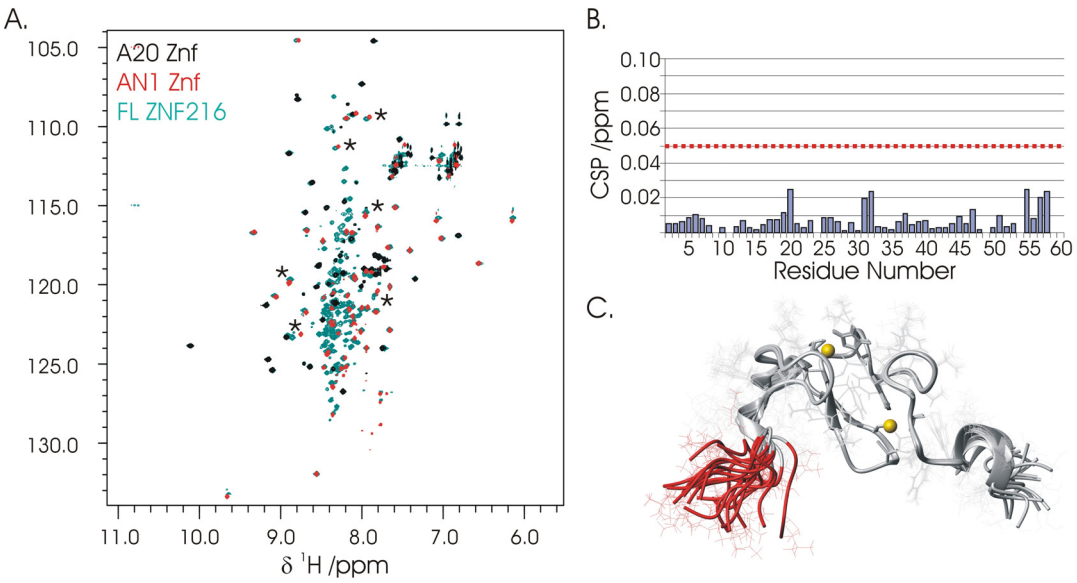



**Figure 5-8: (A) $^{1}$H-$^{15}$N-HSQC of ZNF216 (Blue) compared to that of the Znf_A20 (Black) and the Znf_AN1 (Red). A selection of perturbed Znf_AN1 residues are marked with (*). (B) Residue specific CSP for the Znf_A20 domain in the full-length protein. (C) The structure of the Znf_AN1 (RPDB ID: 1WFL) residues missing in the construct used here are highlighted in red.**

The Znf_AN1 on the other hand is less clear cut; the $^{1}$H-$^{15}$N-HSQC of the Znf_AN1 gives a clean spectrum, a total of 59 of the expected 64 non proline residues are visible with good peak intensity and good chemical shift dispersion suggesting a well folded domain in solution (Figure 5-8). A comparison of the full length ZNF216 spectrum with that of the Znf_AN1 shows a good correlation suggesting a conserved overall fold although there are a number of residues showing CSP between the isolated domain and the full length protein, these changes are typically small ~ 0.05 ppm but affect around 13 backbone resonances (Figure 5-8). This could provide evidence for the proposed interaction between the Znf_AN1 and the Znf_A20 but the lack of corresponding CSP for the Znf_A20 suggests that this is highly unlikely. However this does not rule out the possibility of an interaction between the Znf_AN1 and regions within the interdomain linker. An alternative explanation for the observed CSP comes from the choice of construct used in this study. The construct used here represents residues Lys150 up to the C terminus at Ile213, the published structure on the other hand runs from residues Ser135 up to Ala194. Examination of the published structure suggests that although the start of the Znf_AN1 construct used here lies within the domain boundaries and that residues

outside of this region are poorly defined, they may make contacts with portions of the Znf_AN1 domain (Figure 5-8) giving rise to CSP. It is also worth noting that the Znf_AN1 has a high degree of surface charge containing a total of 21 charged residues (Glu, Asp, Lys and Arg) in addition to the two histidine residues used for zinc ligation. This may result in high sensitivity to subtle buffer changes, in particular for the two histidine residues which sit at the centre of the domain and have $pK_a$ close to the buffer pH. Although all proteins were recorded in identical buffers subtle changes are possible arising from the different preparations for each of the proteins and this could contribute to the observed CSP.

The compact nature of the Znf_AN1 domain (Figure 5-8) gives rise to a large number of inter-residue NOEs for the majority of the Znf_AN1 backbone amide protons. This makes unambiguous separation of sequential NOE from medium- and long-range NOE difficult and precluded the unambiguous assignment of the Znf_AN1 domain without the use of through-bond sequential connectivities using a $^{13}$C-labelled sample. This was deemed outside of the remit of this investigation and an assignment was not performed. Without a full assignment of the Znf_AN1 the origin of the observed CSPs may remain unclear although an interaction with the Znf_A20 is highly unlikely.

### 5.3.2 The Interdomain Linker is Intrinsically Disordered

The remaining peaks of the $^{1}$H-$^{15}$N-HSQC of ZNF216 which do not originate from either the Znf_A20 or the Znf_AN1 show poor chemical shift dispersion with amide proton chemical shifts restricted to a narrow chemical shift range between 8-8.4 ppm (Figure 5-9) which is consistent with literature random coil values [63,141,168]. An examination of all peaks not corresponding to either the Znf_A20 or the Znf_AN1 in the ZNF216 $^{15}$N-TROSY-NOESY, show no inter-residue NOE correlations and only weak intra residue NOEs (Figure 5-9).

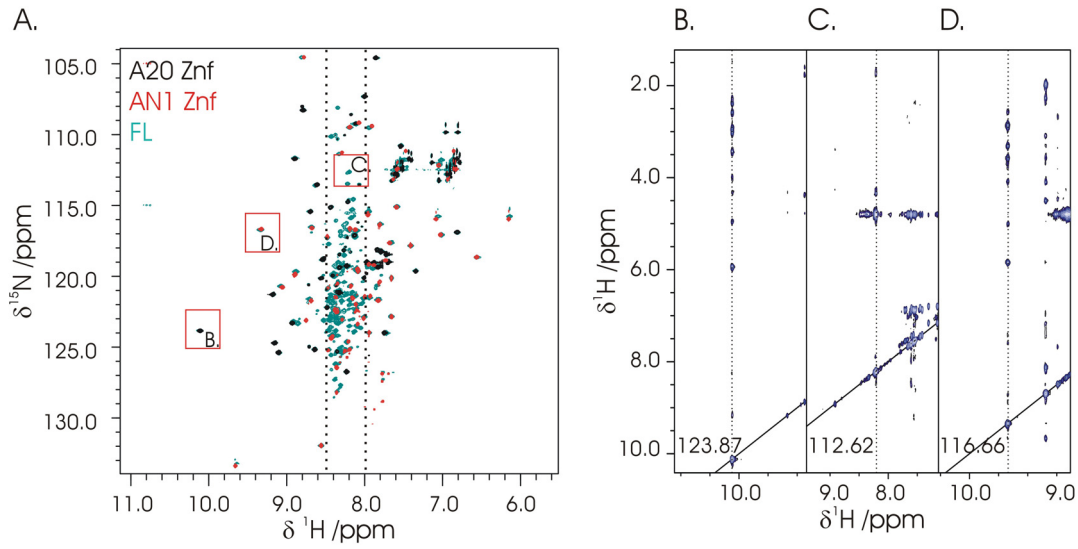



**Figure 5-9: (A) $^{1}$H-$^{15}$N-HSQC of ZNF216 (Blue) compared to that of the Znf_A20 (Black) and the Znf_AN1 (Red). Vertical lines represent the expected range for random coil peptides. (B-D) Portions of $^{15}$N-TROSY-NOESY showing representative residues from; (B) Znf_A20. (C) The interdomain linker. (D) The Znf_AN1.**

The lack of chemical shift dispersion suggests a lack of conformational restriction of the protein backbone and a lack of inter-residue contacts within the interdomain region[169]; this is supported by the complete absence of inter-residue NOE correlations for this region providing strong evidence that this region is intrinsically disordered.

A stable interaction between the Znf_AN1 and residues of the interdomain linker would most likely result in an increase in the chemical shift dispersion of the interdomain linker due to reduced conformational averaging of the peptide backbone and inter-residue contacts. This is particularly likely given the surface of the Znf_AN1 which is highly charged with a total of 21 charged side chains at the protein surface which would be expected to cause large contact mediated chemical shift changes[160]. However no such chemical shift dispersion is observed suggesting that no stable interaction occurs; this does not rule out such an interaction but does make it highly unlikely. A transient interaction which is low in population may result in only small chemical shift changes which maintain the disordered appearance but a stable interaction seems unlikely.

### 5.3.3 The Znf_A20 Interaction with Ub is Unaffected in the Full Length Protein

In order to compare the ubiquitin binding ability of the isolated Znf_A20 domain to the full length protein ITC and ESI-MS were performed to determine the affinity and stoichiometry of the interaction.

ITC titrations were performed whereby Ub at high concentration (~300 μM) was titrated into a calorimetric cell containing Znf_A20 at around 20 μM. The resulting exothermic pulses were used to calculate the thermodynamic parameters of binding (Section 2.3.2.1). The observed exothermic response was fit using the Origin™ software, fitting the data to a single site model to extract stoichiometry, $K_d$, ΔH, ΔS, and ΔG. The measured $K_d$ and thermodynamic parameters for the full length ZNF216 are summarised in Table 5-1. The appearance of the curve and the determined thermodynamic parameters are consistent with no change in the Ub binding ability of the Znf_A20 in the isolated domain compared to the full length protein (Figure 5-10).

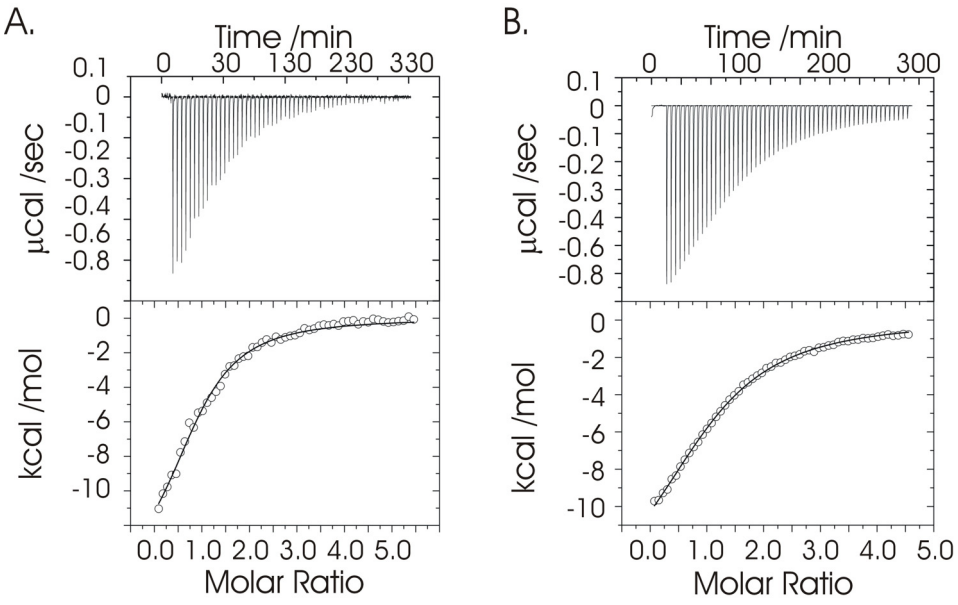



**Figure 5-10: ITC titrations of Ub with; (A) full length ZNF216. (B) Znf_A20.**

**Table 5-1: Thermodynamic parameters derived from ITC titrations of Ub into full length ZNF216 and the isolated Znf_A20 domain.**

| Protein in the Syringe | Mono-Ub | Mono-Ub |
|---|---|---|
| Protein in the Cell | ZNF216 | Znf_A20 |
| Stoichiometry | 0.96 (0.1) | 1.03 (0.1) |
| $K_a$ (±) /M | 110000 (10000) | 84800 (6400) |
| $K_d$ (±) /μM | 10 (1.3) | 12 (1.7) |
| ΔH (±) / kJ/mol | -67 (3) | -63.9 (1.7) |
| ΔS (±) / J/mol/K | -131 (20) | -120 (6) |
| ΔG (±) / kJ/mol | -28.7 (0.8) | -28.1 (0.6) |

ESI-MS was performed on samples containing mono-Ub and ZNF216 at a range of concentrations with a total protein concentration maintained below 6 μM. Mass peaks were observed corresponding to free mono-Ub, free ZNF216, and a 1:1 complex between ZNF216 and mono-Ub. No evidence for higher order complexes or homo-oligomers was observed (Figure 5-11).

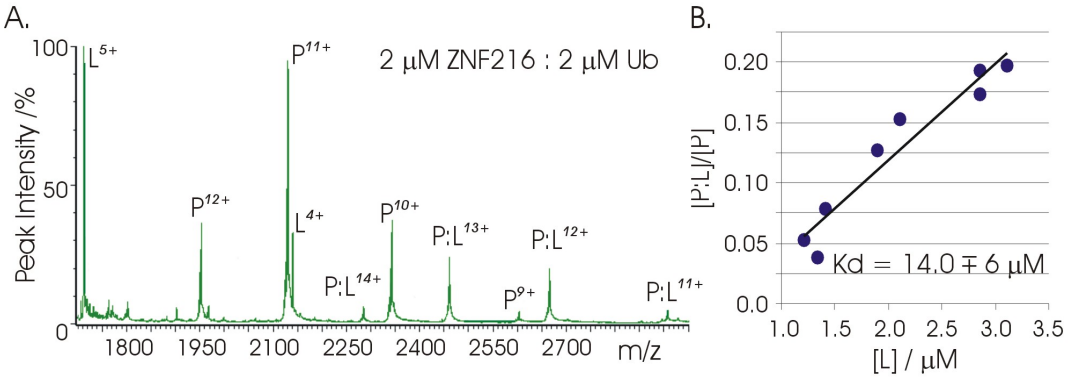



**Figure 5-11: (A) ESI-MS spectrum of 2 μM ZNF216 with 2 μM mono-Ub. (B) Plot of [P:L]/[P] against [L].**

Plots of [P:L]/[P] against [L] were used to determine the $K_d$ as described for the isolated domain, where P represents the concentration of free protein (in this case ZNF216), L represents the concentration of free ligand (in this case Ub) and [P:L] represents the concentration of the one to one complex of the two. For the titration of ZNF216 with mono-Ub a $K_d$ of 14 ± 6 μM was measured (Figure 5-11), consistent with that seen by ITC. The ESI-MS data in this case is lower quality than that for the Znf_A20 resulting in an increased uncertainty. This is the result of small quantities

of salt which co-purified with the ZNF216 resulting in small quantities of salt adducts and the broad charge-state distribution of ZNF216 resulting in reduced signal to noise.

### 5.3.4 Potential Insight into the Function of the Znf_AN1 domain.

The function of the Znf_AN1 is currently unknown, however observations made during the purification of the full length ZNF216 protein and the Znf_AN1 may provide an insight into the function of this domain. The initial attempt to purify the Znf_AN1 utilised SEC using a HiLoad™ Sephacryl™ S-100 HR gel filtration column to purify the protein after cleavage from the GST-tag. The cleaved Znf_AN1 eluted at a similar volume to blue dextran which has a molecular weight of 2,000 kDa. This suggests an aggregate or a very large molecular weight for the Znf_AN1, which is outside of the resolvable range for the Sephacryl™ S-100 gel filtration column. Despite having a high measured absorbance during SEC, little or no protein could be detected for that peak (or any other) by SDS-PAGE. Analysis of this high molecular weight peak by UV-spectroscopy using a NanoDrop ND-100 spectrophotometer revealed a large UV peak with an absorbance maximum at 260 nm. It is worth noting that the buffer used for the Znf_AN1 SEC contained DTT, oxidised DTT gives an absorbance maximum at 260 nm however the blank used for the UV spectroscopy was identical to that used for SEC. In order to account for changes in DTT with time the UV spectrum was recorded immediately after the run and the blank was tested using SEC fractions from the same run as the tested sample. The high absorbance at 260 nm could suggest the presence of DNA co-purifying with the Znf_AN1. In an attempt to separate the Znf_AN1 from potential DNA impurities, purification was attempted using ion exchange. Cleaved Znf_AN1 was passed through a cation exchange column followed by an anion exchange column. The cation exchange column effectively purified the Znf_AN1 domain eluting as a clean peak with good intensity. A second large peak eluted from the anion exchange column under high salt conditions ~1 M NaCl which gave a large UV absorbance at 260 nm. Interestingly this anion exchange column bound peak is also seen for ion-exchange involving the full length ZNF216. SEC recorded on the ion exchange purified Znf_A20 gave a Stoke's radius of 15.4 ± 1.5 Å consistent with monomeric Znf_AN1.

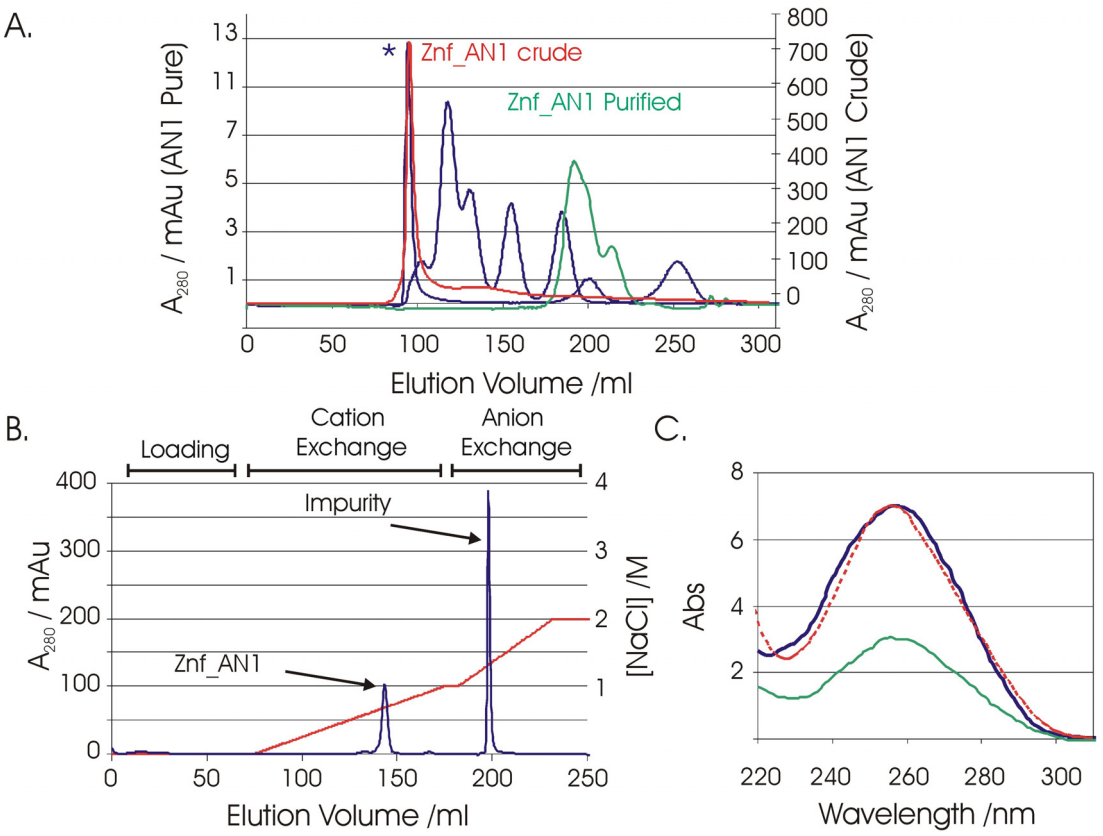



**Figure 5-12: (A) SEC of unpurified (Red) and purified (Green) Znf_AN1 with calibration proteins (Blue). (B) Ion exchange of the Znf_AN1 (Blue), the elution salt concentration is indicated (Red). (C) The UV spectrum of the large aggregate peak from SEC (Blue) and the anion exchange peak (Green), a standard DNA UV spectrum is included (Red).**

The Znf_AN1 elution from the cation exchange column is consistent with the high theoretical isoelectric point of 9.5. The peak eluting from the anion exchange however is consistent with a cationic species, and the high salt required for its elution suggests a highly polyanionic species. The high molecular weight, polyanionic nature and large absorbance at 260 nm, makes for a compelling argument for the presence of DNA co-purifying with the Znf_AN1 domain. During purification the Znf_AN1 domain is subject to a number of cleaning steps including pure water, 2M NaCl, and 0.1 % triton X, a combination of these three wash steps would be expected to remove the majority of non-specific bound impurities. The purification of DNA with the Znf_AN1 therefore provides strong evidence for a direct interaction between the Znf_AN1 domain and DNA.

Examination of the published structure of Znf_AN1 (RPDB ID: 1WFL) shows a considerable number of solvent exposed Arg and Lys side chains (15 in total) forming a large positively charged patch on the surface of the Znf_AN1 which is flanked by several Tyr and Phe side chains (Figure 5-13). The surface properties of the Znf_AN1 are therefore consistent with the proposed ability to bind to DNA, with positively charged residues for potential interaction with the negative phosphate backbone and aromatic residues for stacking interactions with DNA bases; this is a topic which may be worth examination in the future.

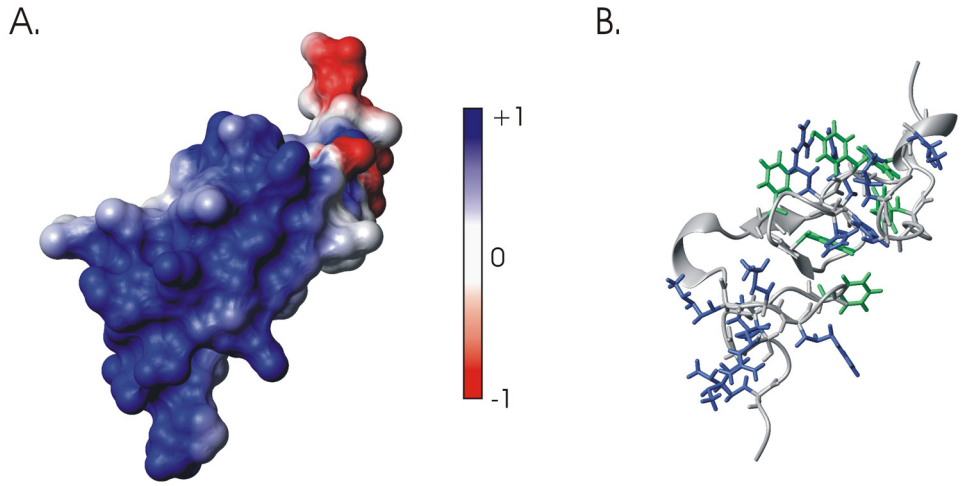



**Figure 5-13: Structure of the Znf_AN1 domain of ZNF216 (RPDB ID: 1WFL) (A) Electrostatic potential calculated using MOLMOL is plotted onto the surface. (B) The position of positively charged (Red) and aromatic (Green) side chains are illustrated onto a ribbon diagram of the Znf_AN1.**

## 5.4 Conclusions

A study by Huang *et al*[117] in 2004 suggested that the Znf_A20 and Znf_AN1 domains of ZNF216 interact leading to the formation of homodimer and tetramers. Using a combination of SEC, AUC, and NMR spectroscopy it has been demonstrated that this is not the case, and that the ZNF216 protein exists as a monomer in solution (at low concentration) with a highly extended conformation. The two domains do not appear to interact and instead function independently separated by a large intrinsically disordered linker. Given the low concentration of an individual protein inside a cell one may expect the monomer to be the *in vivo* form, however this would be influenced by a number of factors including molecular crowding, potential post-translational modifications, subcellular localisation and interactions with natural

substrates. Therefore it can only be concluded that the monomeric state is the *in vitro* state.

The type of domain architecture proposed here is seen for a number of Ub binding scaffold proteins including p62 and Rad23[164,170,171] and may be described as having a 'hook and line' architecture where the flexible linker facilitates multiple interactions by increasing the diffusional radius or capture radius of the two domains[172].

The Znf_A20 interaction with the Znf_AN1 has also been proposed for the related protein SAP8[126], although there is a high degree of sequence similarity between the Znf_A20 and Znf_AN1 domains of ZNF216 and other proteins the proposed interaction of the Znf_A20 and Znf_AN1 can not be ruled out for other proteins. One explanation for the result observed by Huang *et al* [117]could be the choice of construct; in their investigation they simply made two constructs each lacking one of the two domains but in both cases the inter-domain linker remained, the observed interaction could therefore be mediated through either domain interacting with the inter-domain linker through homotypic interactions between linker regions or mediated through binding partners capable of interacting with either of the two domains and/or the interdomain linker.

The Ub binding ability of the full length ZNF216 protein is comparable to that of the Znf_A20 in isolation supporting the notion that the Znf_A20 functions in isolation from the rest of the protein and that there is no measurable interaction with the Znf_AN1 domain.

Observations presented here also provide evidence that the Znf_AN1 has the ability to bind to DNA *in vitro*. This has not been reported for other Znf_AN1 domains however there is a reasonably high degree of sequence conservation across Znf_AN1 domains with many of the positively charged and aromatic residues being highly conserved (Figure 5-3). However this interaction would need to be confirmed by a more extensive study. Znf_A20 domains and Znf_AN1 domains are often found within the same protein and are often the only two functional regions. Should the Znf_AN1 interaction with DNA be confirmed the prospect of pairing a DNA binding domain with a ubiquitin binding domain is a curious one; ubiquitin has several known links to DNA including involvement in DNA repair and histone regulation[2], however how this fits with the already established role for ZNF216 as a shuttle protein would be an interesting question[113].

# Chapter 6

# 6 Characterisation of the UBA domain of p62: Effects of Paget's disease mutations

## 6.1 Introduction

### 6.1.1 The p62/SQSTM1 scaffold protein

The intricate signalling networks that determine cellular processes are often achieved with high specificity while utilising a relatively small number of proteins, several of which may be highly promiscuous[173]. High specificity is often mediated by the function of multi-domain adaptor/scaffold proteins, which possess no catalytic activity of their own, and instead function to bring reacting/functional partners together in multi-protein complexes in order to improve fidelity and regulate signalling pathways.

The SQSTM1/p62 protein is one such scaffold protein; the p62 protein is relatively small, it is only 440 amino acids in length with a MWt of 47.6 kDa. The p62 protein possesses three folded domains and despite its small size, it interacts with a large variety of different binding partners (Figure 6-1). The three folded domains include an N-terminal Phox and Bem1p (PB1) domain spanning residues 1 to 100. The PB1 domain facilitates homo-oligomerisation and interacts with other PB1 domain containing proteins including the atypical protein kinase Cs (aPKCs), NBR1, MEK5, MEKK3 and Akt[174-177]. Residues 128 to 163 contains an a typical or ZZ type zinc finger which interacts with the RIP1 which functions as an adaptor protein to link p62 to the TNF-α receptor family[178]. The third of the three folded domains spans residues 387-436 and contains a Ub associated (UBA) domain which forms non-covalent complexes with mono and poly-Ub and ubiquitinated proteins[179,180]. The region separating the ZZ and UBA domain forms a large disordered region containing two PEST sequences which are thought to reduce the half life of the protein. This region also hosts a number of small recognition sequences for signalling proteins including TRAF6, LIMD1, LC3, Keap1 and p38-MAPK[181-185](Table 6-1 and Figure 6-1).

**Table 6-1: The known functional regions of p62 and the associated binding partners.**

| Region | Name | Confirmed binding partners |
|---|---|---|
| 1-100 | PB1 domain | NBR1[176], aPKCs[177], MEKK3[175], MEK5[177], Caspase-8[173,186], Proteasome[179] |
| 125-165 | ZZ domain | RIP1[178], IDR1[187] |
| 170-220 | LBD | p38 MAPK[184], LIMD1[182,188], Ajuba[182], ALFY[189] |
| 225-221 | TBD | TRAF6[190] |
| 266-294 | PEST | NA – regulates stability[185] |
| 321-342 | LIR | LC3[191] |
| 348-355 | KIR | Keap1[183] |
| 345-377 | PEST | NA – regulates stability[185] |
| 387-436 | UBA domain | Ubiquitin[179,180] MURF2[176], CYLD[31] |
| LBD: LIM Binding Domain, TBD:TRAF6 Binding Domain, LIR: LC3 Interacting Region, KIR: Keap Interacting Region | | |

Through its many binding partners p62 has been implicated in a wide variety of important cellular processes including bone, muscle and T-cell differentiation[192], the shuttling of ubiquitinated proteins to the proteasome[193], regulation of TRAF6 polyubiquitination and NF-κB activation[181,190], selective macroautophagy (autophagy)[173,194], regulation of energy homeostasis[195], receptor trafficking[196], apoptosis[186], adaptation of muscle to mechanical load[176], and regulation of ion chanels[197]. The p62 protein is found localised to cytosolic speckles or sequestesomes formed of PB1 mediated p62 oligomers, p62-aPKC complexes and ubiquitinated proteins which are often misfolded[173,181,191]. These cytosolic speckles function as signalling hubs where p62 interacts with key functional partners including TRAF6[190], caspase-8[186], and Keap1.

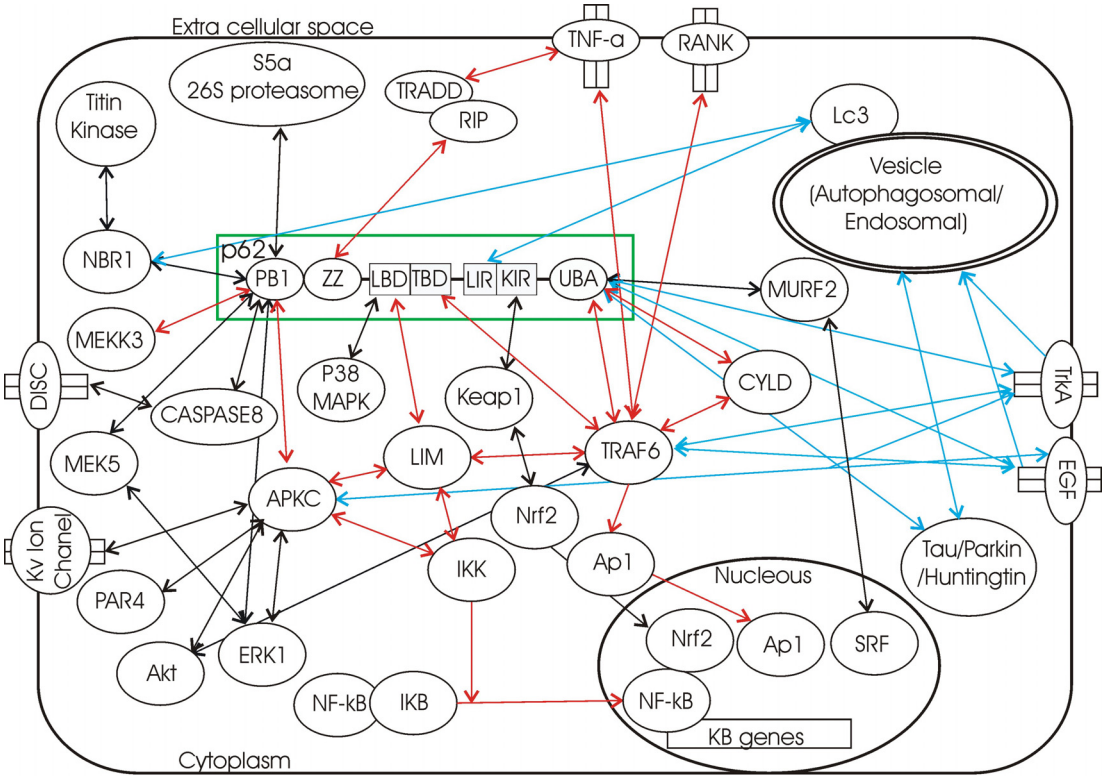



**Figure 6-1: Confirmed direct and indirect interaction partners of p62 to form a signalling network, p62 is indicated by a green box, the regions involved in protein-protein interactions are labelled PB1 = PB1 domain, ZZ = ZZ zinc finger, LBD= LIM binding domain, TBD = TRAF6 binding domain, LIR = LC3 interacting region, KIR = Keap1 interacting region and UBA = UBA domain. Red arrows indicate interactions linked to the regulation of NF-κB activity, Blue lines represent those associated with autophagy and/or receptor trafficking.**

Recently the p62 protein has been highlighted as playing a key role in autophagy. Autophagy is a bulk protein degradation pathway responsible for the turnover of long lived proteins, the disposal of excess or damaged organelles, the removal of invading bacteria, and the degradation of aggregated or misfolded proteins. Bulk cytoplasm containing material for degradation is engulfed by isolation membranes to form autophagosomes, which subsequently fuse with the lysosome resulting in degradation of its contents by lysosomal hydrolases[198]. The classical view of autophagy considered it to be a non-selective bulk degradative pathway, the primary function of which was to recover amino acids and generate energy during times of starvation. The recent discovery of autophagy receptors or 'shuttling-factors' has demonstrated that a process of selective autophagy exists, whereby

targets are selected for degradation by the autophagy-lysosome system[199,200]. p62 interacts with LC3, a key component of the autophagy machinery. The identification of the p62 interaction with LC3 led to the identification of p62 as one of the first receptors or shuttling factors for selective autophagy of cytosolic components[173,191,194,199]. The role of p62 in selective autophagy has a clear link to several diseases. p62 facilitates the removal of bacteria, damaged organelle, and misfolded protein aggregates or inclusion bodies. Importantly, p62 has been identified as a component of inclusion bodies in a number of human diseases, including the neurodegenerative diseases Alzheimer's, Parkinson's, and amylotrophic lateral sclerosis and in the liver diseases alcoholic hepatitis, hepatic steatosis, and hepatocellular carcinoma[173,193,194,201].

The dominant downstream effect of p62 in response to activation of interleukin 1 (IL1), receptor activator of NF-κB (RANK) and the TNF-α receptor family has been shown to be activation of NF-κB[192,202]. Activation of NF-κB drives the expression of inflammatory and survival genes. In osteoclasts this stimulates osteoclastogenesis and disruption of this axis leads to bone abnormalities in mice[192]. The precise mechanism of p62 involvement is not fully understood but there is a clear link between several p62 associated factors and the regulation of NF-κB (Figure 6-1). NF-κB is retained in the cytoplasm by the inhibitor of NF-κB (IκB) under the control of the IκB Kinase (IKK) complex. When activated, IKK phosphorylates IκB, leading to its Lys48-linked polyubiquitination and degradation by the UPS. This frees NF-κB to translocate to the nucleus where it is able to drive gene expression. The p62-TRAF6 interaction is thought to be of particular interest to the regulation of NF-κB; TRAF6 has been implicated in the activation of IKK in a p62 dependent manner by two potential mechanisms. The first is by modulating phosphorylation of the IKK β subunit which is required for its activation. The second is by tagging of the IKK β subunit with Lys63 linked poly-Ub which has also been shown to be important for activation. TRAF6 activation is dependent on auto-ubiquitination which has been shown to require both the PB1 and UBA domains of p62[190,192].

Another set of important binding partners of p62 are LIM domain containing proteins. The LIM domain containing proteins Ajuba and LIMD1 have been shown by biochemical studies to interact with p62 (170-220) in response to RANK and TNF

family receptors leading to NF-κB activation[182]. LIMD1 has also been shown to modulate formation of the AP1 complex as part of RANK regulated osteoclast development[182,188].

### 6.1.2 Paget's Disease of Bone and p62

Paget's disease of bone (PDB) is the second most common metabolic bone disorder; it affects approximately 1.5 million people in the US and around 3 % of people over 55 in the UK. PDB has a clear geographical distribution, being common in countries with Anglo-Saxon derived populations[202-204]. PDB is characterised by increased bone turnover resulting in painful deformities and focal lesions[204]. Despite first being described over 130 years ago many aspects of the pathogenesis of PDB remain undefined. At a cellular level PDB is characterised by an increase in the quantity of bone-resorbing osteoclasts which are larger than normal and contain unusually large numbers of nuclei. However the molecular mechanisms which underpin PDB remain unclear[204].

Both environmental and genetic factors have been implicated in the pathogenesis of PDB, several genetic loci have been linked to PDB with a number of genes identified including the RANK gene[203,205], the OPG gene[205,205], and in the SQSTM1/p62 gene[203,204]. Mutations have been identified in both OPG and in SQSTM1 from patients with PDB. While OPG causes juvenile onset PDB, mutations in SQSTM1 appear commonly in late-onset PDB, with mutations identified in 25-30 % of cases of familial PDB and in 10 % of sporadic PDB[203-206]. Interestingly all three isolated genes are associated with RANK-L mediated NF-κB activity indicating a clear link between PDB and RANK-L mediated NF-κB signalling[203,204].

Until recently all of the known PDB mutations found within p62 clustered within its C-terminal UBA domain including the most common PDB mutant P392L which represents around 75 % of p62 related cases[207]. Although recently mutations have been identified which are in the long flexible region between the UBA and ZZ domains, the high concentration of mutations within the p62-UBA suggests a role for the p62-UBA in the cellular events which contribute to PDB. p62 interacts with many components of the NF-κB activation pathway, many of which are in a Ub dependent manner. The key role of the UBA domain is therefore not surprising. Despite several years of investigation and a number of theories the precise

mechanism by which PDB mutants affect the p62-UBA domain, and the role that any effects may play in the pathogenesis of PDB remains unclear.

Several previous investigations into the p62-UBA confirmed that the p62-UBA does indeed form a classical UBA three helix bundle with at least two structures currently available in the RPDB. Mapping of a selection of the known PDB mutants onto the available structures of the p62-UBA shows little correlation in terms of location, with PDB mutants spread over the entire domain (Figure 6-2). However recent investigations have cast doubt on the monomeric models of the p62-UBA suggesting that instead the p62-UBA forms a dimer in solution[61], as a result any previously recorded data on the p62-UBA and its mutants may need to be re-examined in the light of a correct structural model.

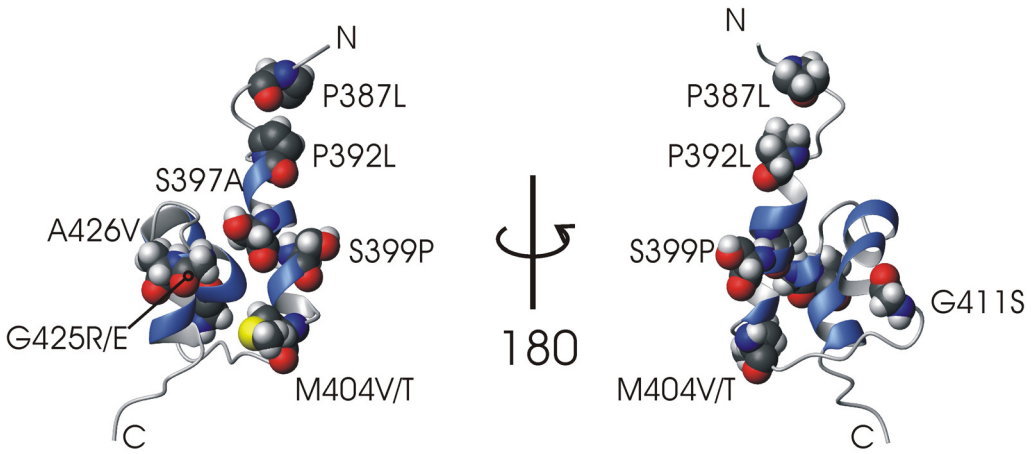



**Figure 6-2: A selection of PDB mutants of the p62-UBA mapped onto the monomeric structure of the p62-UBA (RPDB ID: 1Q02).**

### 6.1.3 Summary and Aims

The p62 protein is a multi-functional protein with many binding partners involved in a large number of cellular processes of profound interest. The C-terminal UBA domain of p62 binds non-covalently to mono-Ub and is required for many of the cellular functions of p62, most noticeably in TRAF-6 poly-ubiquitination and NF-κB activation. A subset of sufferers of PDB, a debilitating bone disorder characterised by excessive bone turnover and deformities, harbour mutations within the p62-UBA domain. As a direct result, several studies by a number of groups have been undertaken in order to characterise the molecular detail of the function of the

p62-UBA and the effect of PDB mutants on the function and properties of the p62-UBA.

Recent investigations have provided evidence that the p62-UBA exists as a dimer in solution. In order to better understand the molecular effects of PDB mutations, the p62-UBA dimer has been characterised using structural and biophysical approaches. This will provide a basis for further investigations of the effects of PDB mutants on the p62-UBA. This investigation has examined several PDB mutant including S399P, G411S and G425R in order to examine their affects on the properties of the p62-UBA.

## 6.2 The p62-UBA is Dimeric in Solution

The p62-UBA domain has been subject to continuous study within the Searle research group over the past 8 years[61,170,180,202,208,209]. The NMR spectrum of the p62-UBA gives clean spectra with 47 of the possible 49 backbone amide protons clearly measurable with good chemical shift dispersion suggesting a single well folded species. The number of observed resonances led to the assumption in previous structural studies that the p62-UBA was monomeric in solution[36,180,209]. This was supported by previous concentration dependence studies of the ratio of the spin-spin to spin-lattice relaxation rates ($R_2/R_1$), in the range 0.2 – 1.8 mM at 298K, which suggested no concentration dependence below 1.4 mM, above which some evidence for aggregation was detected[36]. Some concentration dependence was observed below 0.3 mM at 310K which was attributed to poor signal to noise at low concentrations[36].

However, during purification size exclusion chromatography (SEC) was performed on the p62-UBA which recorded a elution volume of around 186 ml. This is consistent with a considerably larger molecule with an apparent mass of around 13 kDa (Figure 6-3). The apparent MWt recorded by SEC can be affected by a number of factors and is better described in terms of Stoke's radius. SEC was repeated with a set of low MWt proteins (GE HealthCare) which were used to generate a calibration curve for the calculation of apparent MWt and/or Stoke's radius (Appendix C). From SEC a Stoke's radius for the p62-UBA of 16.6 ± 1 Å was calculated.

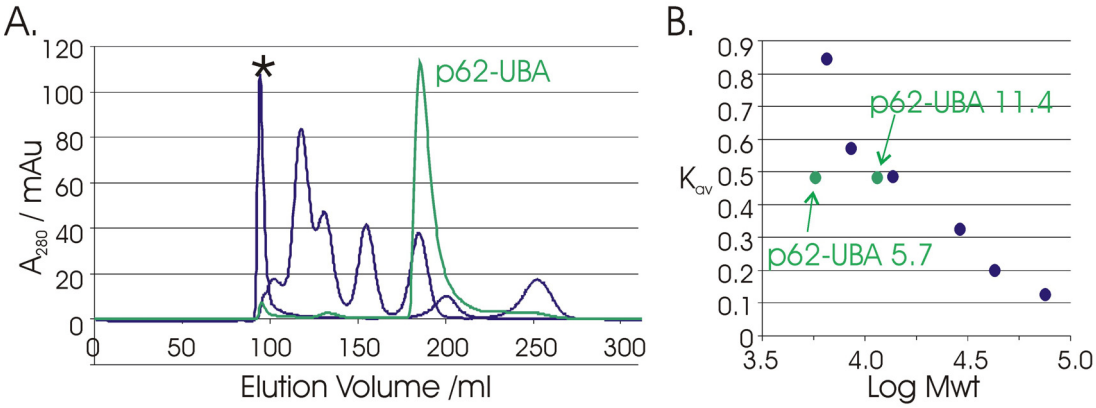



**Figure 6-3: (A) SEC trace for p62-UBA (Green) Salt (Red) and Low MWt protein (Blue) (A.)(* denotes blue dextran) (B) The standard curve used to calculate the apparent native mass of p62-UBA. The $K_{av}$ against expected mass for monomeric and dimeric p62-UBA are shown in red and labelled.**

The larger than expected Stoke's radius may suggest either higher order oligomerisation or a highly extended conformation in solution. However neither of the previously determined structures for the p62-UBA appear highly extended and neither possess large flexible termini which could result in a larger than expected Stoke's radius. The program HYDROPRO can be used to estimate Stoke's radius from a given protein structure, the two previously determined structures for the p62-UBA (RPDB ID: 1Q02 and 2JY7) gave estimated Stoke's radii of 13.1 and 11.6 Å representing a non-spherical but relatively compact structure[180,208]. This discrepancy suggests that the structures previously presented for the p62 may be incorrect and may be explained in a number of ways. The first being the oligomerisation state, where the increase in radius is explained by a higher order structure such as a dimer. The second being dynamics; if the structure was in rapid equilibrium with a structure containing significant quantities of disorder the SEC may well record a structure with an increased apparent radius. However measurements of the dynamics of the p62-UBA recorded previously gave no evidence for regions of high dynamics with the majority of the domain appearing to be relative rigid[210].

### 6.2.1 Low Concentration NMR of the p62-UBA

Previous NMR studies of the p62-UBA interaction with Ub had revealed evidence for the presence for a low populated Ub-binding species which is in equilibrium with the major species observed under the conditions of the NMR[180].

Previous NMR studies were restricted to relatively high concentrations in the range 0.6 - 1 mM due to the sensitivity of NMR instrumentation. In order to examine the possibility of concentration dependent changes in the NMR spectrum at concentrations below those previously obtainable, $^{1}H$-$^{15}N$-HSQC were recorded by Dr M. J. Pandya, Prof M. P. Williamson and Dr C J. Craven at the University of Sheffield using a Bruker DRX600 with a cryoprobe. The increased sensitivity this gave allowed measurements at concentrations down to 10 μM recorded at 298 and 310 K[61]. At the low concentration of 10 μM, double the original number of peaks was observed in the $^{1}H$-$^{15}N$-HSQC suggesting a concentration dependent change in the visible species which are in slow chemical exchange on the NMR time scale (Figure 6-4)[61,145]. This is consistent with a change in oligomerisation state suggesting that the species for the p62-UBA observed in previous studies was indeed a higher order oligomer. The second species was consistent with the previously identified low populated Ub-binding species and this facilitated the assignment of the low concentration species by comparison to the Ub bound form (Discussed in Chapter 7). The chemical shift dispersion of the low concentration species is comparable to that seen for the high concentration species giving little or no indication of unfolding in the low concentration species. There are however relatively large chemical shift perturbations observed upon dissociation of the dimer to monomeric units suggesting considerable conformational changes (Discussed in section 6.3)

The apparent molecular weight observed by SEC is consistent with the formation of a dimeric species, from this it can be assumed that the observed species in the NMR spectrum at 10 μM are dimeric p62-UBA, representing the major species at high concentration, in slow equilibrium with a monomeric species which is only evident at low concentration.

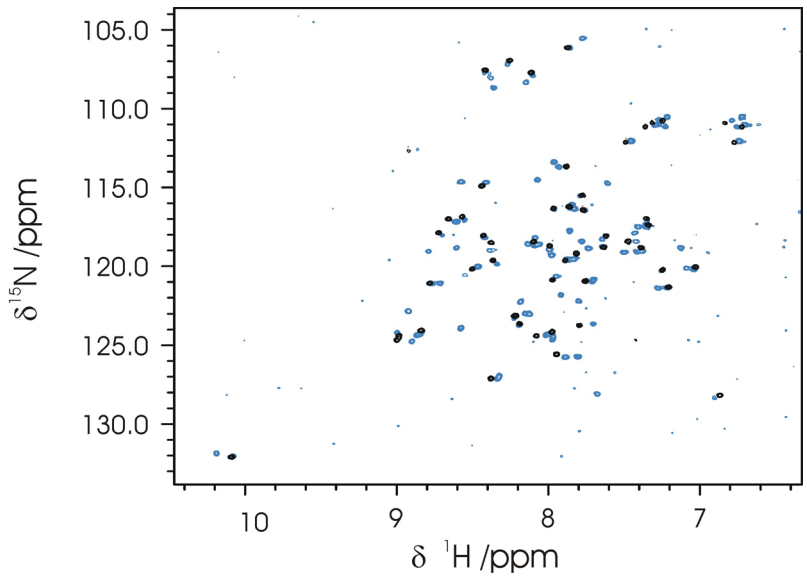



**Figure 6-4: $^{1}$H-$^{15}$N-HSQC of the p62-UBA recorded at total protein concentration of 1 mM at 298 K (Black) or 10 µM at 310 K (Blue).**

The relative intensity of the $^{1}$H-$^{15}$N-HSQC peaks of the two species was used to determine the relative concentrations of the two species. No considerable difference in line shapes for the monomer and dimer species was observed thus the intensity was assumed to be proportional to the monomer concentrations in each species. From the apparent concentration of species an apparent $K_d$ for the monomer-dimer equilibrium ($K_{dim}$) of 6 ± 3 and 14 ± 3 µM were measured at 298 and 310 K respectively. This represents a relatively high affinity dimer. the lowest concentration recorded in previous studies was 200 µM, where the dimer would be expected to be in vast excess, the lack of observed changes in the value of $R_2/R_1$ in previous studies is therefore unsurprising[36].

The observation of a second species at low concentration can only be explained by changes in oligomerisation and SEC suggests that the higher order species is dimeric. However the proposed dimeric species gives a single set of resonances in NMR spectra at high concentration, the implication of this is that the proposed dimer must be highly symmetric resulting in chemical shift equivalence between corresponding resonances in the two subunits of the dimer molecule.

### 6.2.2 Biophysical Characterisation of the p62-UBA Dimer

#### 6.2.2.1 ESI-MS and ITC Confirm the Stoichiometry and Strength.

In order to confirm the oligomerisation state of the p62-UBA, and to determine the thermodynamic parameters of the interaction, a range of biophysical approaches were utilised. In order to characterise the stoichiometry of the p62-UBA, ESI-MS was recorded at a range of total protein concentrations from 0.25 – 4 μM. Protein ESI results in multiple species visible in the ESI-MS spectrum corresponding to multiple charge-states, with the recorded signals representing values of mass to charge ratio (m/z). As a result values for m/z for dimer and monomeric species can coincide, where the mass and charge are both doubled giving the same value of m/z. However the average charge-state distribution is considered to be proportional to the solvent exposed surface area during ionisation and as such the charge-state distribution of the monomer and dimer species would be expected to differ. What's more where a dimeric species held an odd numbered charge-state, peaks would appear in the ESI-MS spectrum coinciding to apparent (impossible) half charge-states for the monomeric species. Two characteristics in the ESI-MS spectrum can be used to indicate the presence of a higher order structures. The first is a concentration-dependent shift in the observed average charge-state for the apparent monomer species due to overlapping monomer and dimer peaks. The second is the presence of peaks in between the charge-states assigned to the monomeric species. The ESI-MS spectrum of the p62-UBA in 25 mM ammonium acetate at pH7 gives an apparent molecular weight of 5738.9 ± 1 consistent with the expected apoprotein mass of 5739.4. However the apparent average charge-state shows a clear concentration-dependence, shifting from 4.2 at 0.25 μM to 3.8 at 4 μM (Figure 6-5). There is also the presence of mass-ion peaks which are between the monomeric peaks. These peaks are consistent with a dimeric species with a measured mass 11478 ± 1 (Figure 6-5). No mass peaks consistent with trimer or tetramer could be seen even at relatively high concentrations (up to 4 μM).

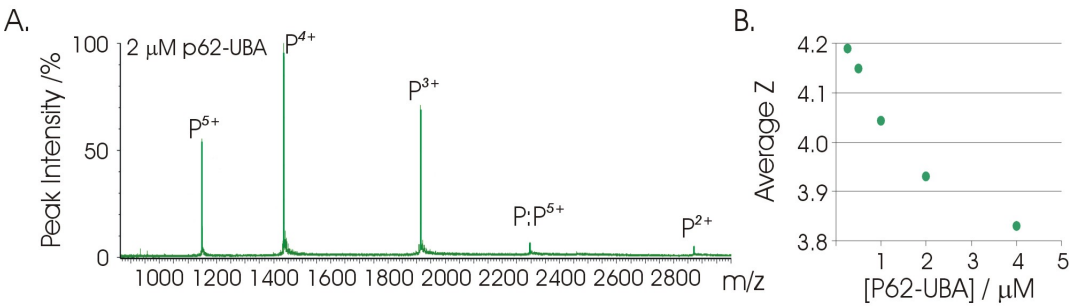



**Figure 6-5: (A) ESI-MS spectrum of 2 µM p62-UBA (A) monomer (P) and dimer (P:P) peaks are labelled. (B) Plot of the concentration dependence of the observed average charge-state for the p62-UBA.**

Although the overlap of mass ion peaks corresponding to monomer and dimer species precludes a calculation of their relative concentrations from the peak intensities in the mass spectrum, the presence of dimer peaks and the lack of peaks corresponding to higher order species provides strong evidence for the stoichiometry of the p62-UBA.

In order to determine the thermodynamic parameters for the p62-UBA dimmer, ITC dilution experiments were recorded whereby concentrated p62-UBA (~300 µM) was diluted by sequential injections into buffer in order to observe the calorimetric response of dimer dissociation induced by rapid dilution. A series of endothermic heat pulses of decreasing intensity were observed resulting from gradually decreasing dissociation to monomer with increasing protein concentration within the calorimetric cell, until such point where saturation occurs and dissociation is no longer observed (Figure 6-6). Once the concentration of p62-UBA within the calorimetric cell reached above 20 µM little or no endothermic heat pulse was observed. The resulting dissociation isotherm of the integrated heat response was fitted to a dimer dissociation model to give a $K_{dim}$ = 7 ± 1 µM and a ΔH of 6.8 ± 3 kJ/mol at 298 K. From this a $\Delta G_{298}$ of 29.4 ± 1 kJ/mol and a ΔS of -75 ± 5 J/mol/K was determined (Table 6-2). Repeating the ITC dilution experiments with the related monomeric, RAD23-UBA domain (Unpublished data by Miss J. Adlington), gives a mild exothermic heat response which is indistinguishable from buffer alone (Figure 6-6).

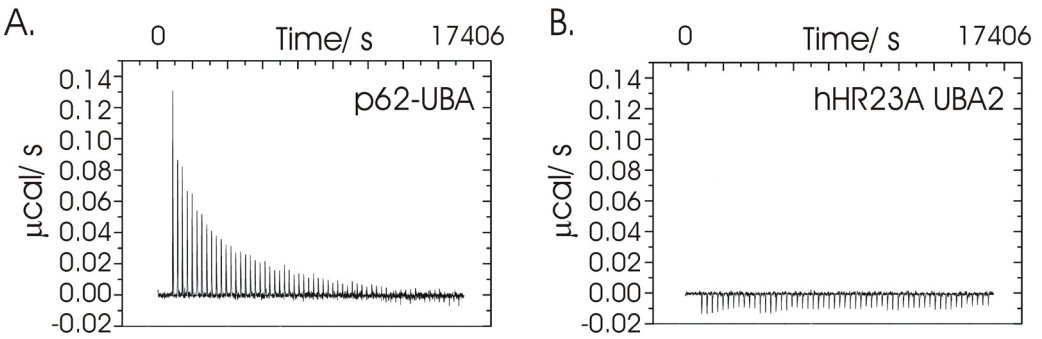



**Figure 6-6: ITC dilution thermal output for; (A) the p62-UBA and (B) RAD23-UBA.**

**Table 6-2: Thermodynamic parameters for the p62-UBA dimer derived from ITC.**

| | $K_{dim}$ (±) / µM | ΔH (±) / kJ/mol | ΔS (±)/ J/mol/K | $\Delta G_{298}$ (±) / kJ/mol |
|---|---|---|---|---|
| p62-UBA | 7 (1) | 6.8 (3) | -75 (5) | 29.4 (1) |

The ITC dissociation data fits well to a dimer dissociation model consistent with mass spectrometry and SEC observations. The determined $K_{dim}$ of 7 ± 1 µM is in excellent agreement with that determined from the signal intensity in NMR dilution experiments which gave an estimated $K_{dim}$ of 6 ± 3 µM. From the ITC analysis the ΔH contribution to the $\Delta G_{298}$ of 29.4 ± 1 kJ/mol is relatively small compared to a large entropic contribution. This is consistent with the observed interface identified by NMR which shows a largely hydrophobic interface suggesting that the burial of hydrophobic surface area may be the thermodynamic driving force behind the dimer association. However ITC reports only on global thermodynamics and other factors may contribute to the entropic driving force other than the classical hydrophobic effect.

#### 6.2.2.2 Far-UV Circular Dichroism Studies

In order to investigate relative stabilities of the monomer and dimeric species of the p62-UBA equilibrium unfolding experiments were recorded in the concentration range 1.5- 200 µM monitored using far-UV CD. A total of 10 concentrations were recorded at non-regular intervals, with reduced concentration intervals recorded at concentrations close to the $K_{dim}$.

The far-UV-CD spectrum of the p62-UBA is consistent with a protein containing a high percentage of α-helix with characteristic negative ellipticity bands at 222 and 208 nm and a positive band at 192 nm[41]. This is wholly consistent with previously reported structures for the p62-UBA which reported a classical UBA three helix bundle. This is also in agreement with previous CD studies of the p62-UBA[36,41,170,180,208,211]. Several of the recorded CD spectra showed significant signal overloading at medium to low wavelengths due to high protein concentration, several others showed very low signal intensity resulting in very poor signal to noise. In these cases the CD spectra were considered as inaccurate in terms of intensity and characteristics, however of the ten measured concentrations a total of six gave suitable CD spectra representing the concentration range 57-1.5 μM. The helical characteristics of the CD spectra remain at all concentrations suggesting that the monomeric species is not unfolded which is consistent with NMR observations. The intensity of the two helical bands varies approximately linearly with concentration suggesting that concentration errors were minimal. However there is some deviation from linearity potentially as a result of the monomer to dimer transition. This is supported by a comparison of the ratio of the ellipticities at 222 and 208 nm ($\theta_{222}/\theta_{208}$) which changes with concentration shifting from 1.1 at 41 μM to 0.85 at 1.5 μM. An ellipticity ratio of $\theta_{222}/\theta_{208}$ of 1.1 has previously been used as an indicator of coiled-coil character or $3_{10}$ helices[67,168,212,213]. This has been attributed to the increase in helical pitch in these structures. Whether this is the case in the p62-UBA can not be determined on this basis alone, however the gradual shift in the $\theta_{222}/\theta_{208}$ ratio is a clear indicator of a structural transition and suggests some changes in the helical conformation between the monomer and dimer.

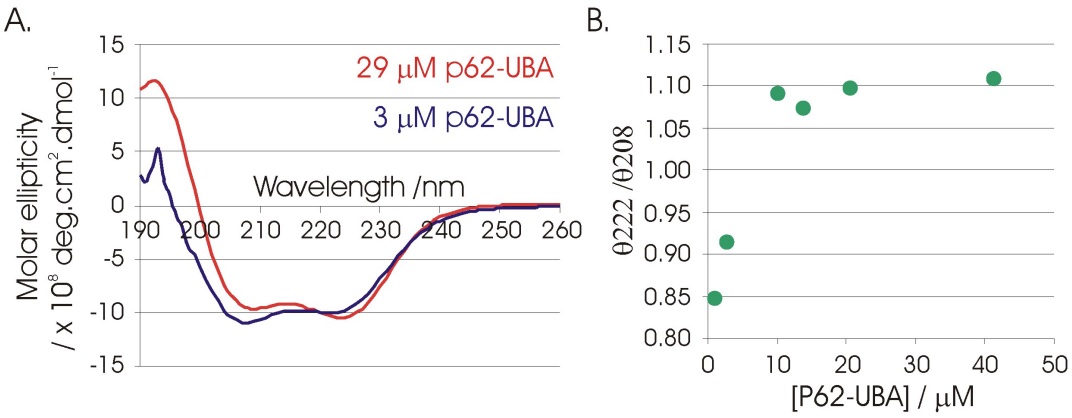



**Figure 6-7: (A) CD spectrum of p62-UBA at 29 (Red) and 3 (Blue) µm. (B) Plot of $\theta_{222}$ nm/$\theta_{208}$ nm as a function of concentration.**

Equilibrium unfolding was recorded in the temperature range 288.1 to 368.1 K, and the unfolding transition was monitored using the far-UV CD band at 222 and 224 nm. The temperature was raised in 0.5 K steps with 30 seconds equilibration time at each temperature. For the highest concentration both unfolding and folding profiles were recorded to confirm the reversibility of the transition, however in the interest of time, for all concentrations thereafter only unfolding transitions were measured. Apparent mid-points of the folding transition were determined by direct fitting to the folding equations outlined by Greenfield *et al*[42] in combination with first derivative plots of the folding transition[42].

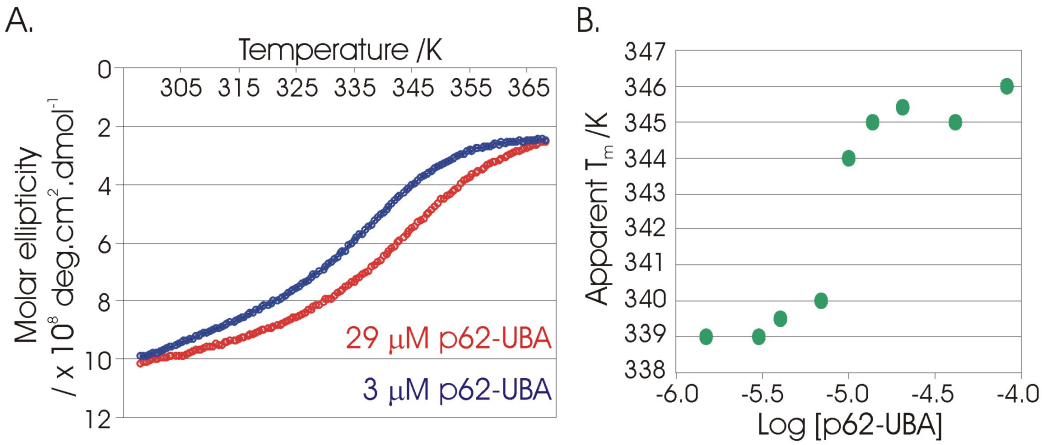



**Figure 6-8: (A) Thermal unfolding profile for the p62-UBA 29 µM (Red) and 3 µM (Red). (B) Plot of the concentration dependence of the apparent $T_m$ for the p62-UBA.**

The apparent mid-point of the thermal unfolding transition appears relatively stable at high concentrations (> 20 μM), giving a $T_m$ of around 347 ± 2 K. However at concentrations below 20 μM a clear and sharp reduction in the apparent mid-point occurs, settling at a $T_m$ of 339 ± 2 K at 1.5 μM, representing an increase in stability of ~ 8 K as a result of dimer formation. Fits to the thermal unfolding profiles give an estimated ΔH of around 230 ± 20 kJ/mol at high concentration while at low concentration an estimated ΔH of 180 ± 30 kJ/mol was determined. The concentration at which the transition occurs is consistent with a $K_{dim}$ of around 10 ±8 μM consistent with the above NMR and ITC data.

### 6.2.3 Summary

From a combination of NMR dilution experiments, SEC, ESI-MS, ITC, and equilibrium unfolding data it has been demonstrated that despite previous structural studies of the p62-UBA being solved as a monomer unit, the p62-UBA is in fact a symmetrical dimer in solution with a $K_{dim}$ of around 6 μM. No evidence is seen for higher order species by any of the techniques employed here. The chemical shift dispersion in the NMR spectrum of the monomer suggest a well folded domain, this is supported by far-UV CD data which suggests a high percentage of α-helix in both the dimer and the monomer. However there is evidence for considerable conformational changes within the monomer unit upon dissociation implicated by both NMR and far-UV CD.

## 6.3 NMR Investigations of the p62-UBA Dimer

In light of the clear biophysical evidence that the p62-UBA forms a high affinity dimer in solution, a re-evaluation of the data used to derive the initial NMR structures was required. Analysis of the NMR spectrum recorded by Dr M. J. Pandya et al showed large CSP for the majority of residues with very large CSP (> 0.5) for residues Leu413 and Thr430 through to Tyr433 with residue Ile431 having the largest CSP of 2.37 ppm. Using the previously solved monomeric structures as a guide, the large chemical shift perturbations seen for the dissociation of the dimer to monomer are seen to cluster to a large contiguous binding surface with Trp412 at its centre. The CSP data along with intermolecular NOE's and aided by the requirement for C2 symmetry were used by Dr Jed Long to generate a solution structure of the p62-UBA dimer which is consistent with the observed CSPs (Figure 6-9)[61]. The

solved structure forms a symmetrical anti-parallel dimer with regard to the relative alignment along helix 2, the side chain of Trp412 sits at the centre of the interface and the C-termini of helix 3 of each monomer are in close proximity (Figure 6-9)[61]. The dimer interface buries around 590 $Å^2$ of solvent-accessible surface area for each monomer unit. The interface is essentially comprised of a hydrophobic core formed from the side chains of Trp412, Leu413, Leu416, and Thr430 with additional electrostatic and polar contacts around the periphery including a potential salt bridge between Glu409 and Lys420, and a potential hydrogen bond from the phenol-OH of Tyr433 to the backbone carbonyl of Asp429[61]. The size and composition of the interface is consistent with other homodimeric interfaces of a similar size and affinity[214].

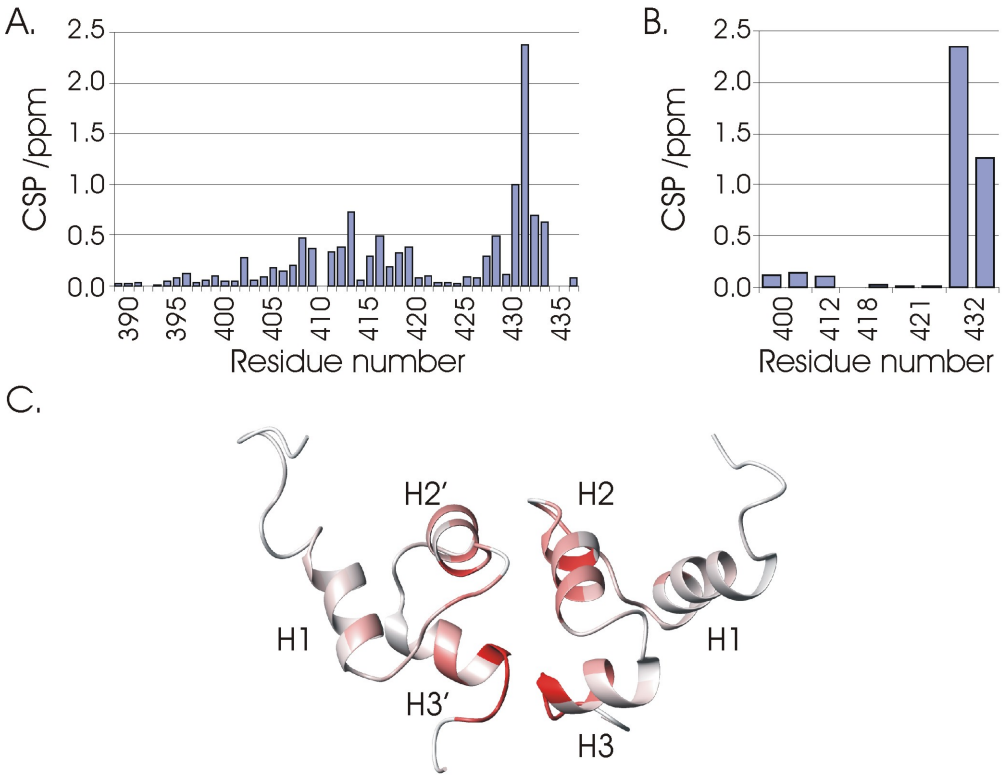



**Figure 6-9: (A) Residue specific CSP for the dimer dissociation of p62-UBA for; (A) backbone NH and (B) side chain $NH_2$ or NH. (C) Large CSPs are plotted onto the dimeric p62-UBA structure (RPDB ID: 2KNV).**

The observed CSPs are considerable and larger than those seen for a typical protein-protein interaction, suggesting the potential for significant conformational changes upon dissociation. In order to examine the origins of the large CSPs observed upon dissociation, Dr Pandya *et al* extended the examination of the dimer

induced CSPs to predict the dihedral angles of the p62-UBA monomer and dimer from their chemical shifts using the program TALOS[84]. The predicted dihedrals are consistent with maintenance of the overall secondary structure of the UBA domain in both structures, however helix three appears to extend to the C-terminus in the monomeric form. In the solved solution structure of the dimer the C-terminal end of helix 3 forms part of the dimer interface forming an intermolecular hydrogen bond between Asp429 in helix 3 and the side chain of Tyr433 (Figure 6-10). This appears to prevent helix 3 from extending beyond Thr430, however in the monomeric state the loss of the intermolecular hydrogen bond may result in extension of the helix 3 beyond Thr430.

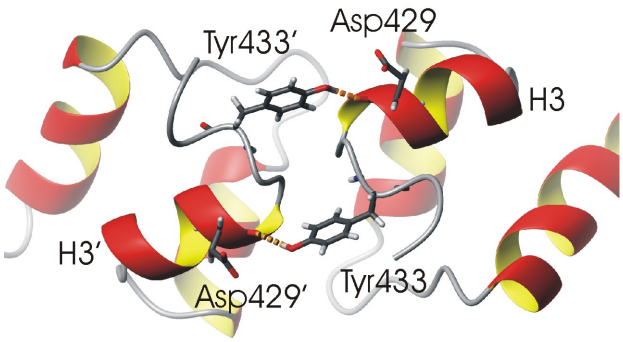



**Figure 6-10: Ribbon diagram of the p62-UBA dimer showing the potential intermolecular hydrogen bond between Asp429 and the side chain of Tyr433.**

In previous NMR studies of the p62-UBA the large chemical shift perturbations, explained here by dimerisation, were attributed to a large conformational change involving the re-packing of the three helix bundle into an alternative fold (Figure 6-11).

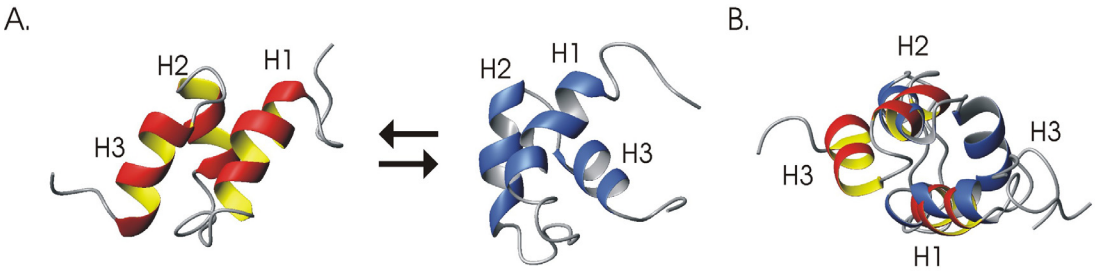



**Figure 6-11: (A) Conformational switch suggested by the previously solved structures of the p62-UBA. (B) Overlay of the two structures using helix 1 and helix 2.**

The determination of this fold was achieved using Ub bound p62-UBA using NOEs, chemical shift derived dihedral restraints, and RDCs[180]. The large CSPs are seen here to result from dimer dissociation, therefore any large conformational rearrangement contributing to the magnitude of the CSP must be coupled to dissociation. The conformation of the p62-UBA monomer remains unclear, however given that the previously determined structure for the Ub bound form of the p62-UBA was determined independently of knowledge of the free form, one would expect that the interpretation of the data used to generate the structure is unaffected by the realisation that the free p62-UBA forms a dimer in solution. Given the magnitude of the CSPs caused by dimerisation, the monomeric form of the p62-UBA will be assumed to adopt the overall conformation previously determined by Long *et al* for the Ub binding conformation of the p62-UBA[180]. For the remainder of this thesis data will be interpreted on the basis of the dimeric p62-UBA structure which is in equilibrium with the conformationally switched monomeric p62-UBA structure (RPDB ID: 2KNV and 2JY8).

### 6.3.1 Solution PRE Measurements

The use of paramagnetic agents to influence nuclear spin relaxation via the strong local magnetic fields produced by the magnetic moments of the unpaired electron, has been widely used to probe the structure and dynamics of biomolecules[68,215-217]. Water soluble paramagnetic molecules which are not covalently attached to the biomolecule of interest are assumed to form non-specific interaction with the surface of biomolecules, these interactions are transient and governed primarily by diffusion[215]. The observed paramagnetic relaxation enhancements are largely limited to surface residues giving an indication of the degree of solvent exposure for a given nucleus. This provides potentially useful information for structural and dynamics studies of the protein backbone [215,217,218].

In order to identify residues with significant exposure to solvent, $^{1}H$-$^{15}N$-HSQC were recorded on a 1 mM sample of p62-UBA in the presence of the soluble nitroxide-radical containing compound HyTEMPO, at various concentrations between 15 and 90 mM. PRE was determined as the ratio of the $^{1}H$-$^{15}N$-HSQC peak intensities in the presence and absence of HyTEMPO.

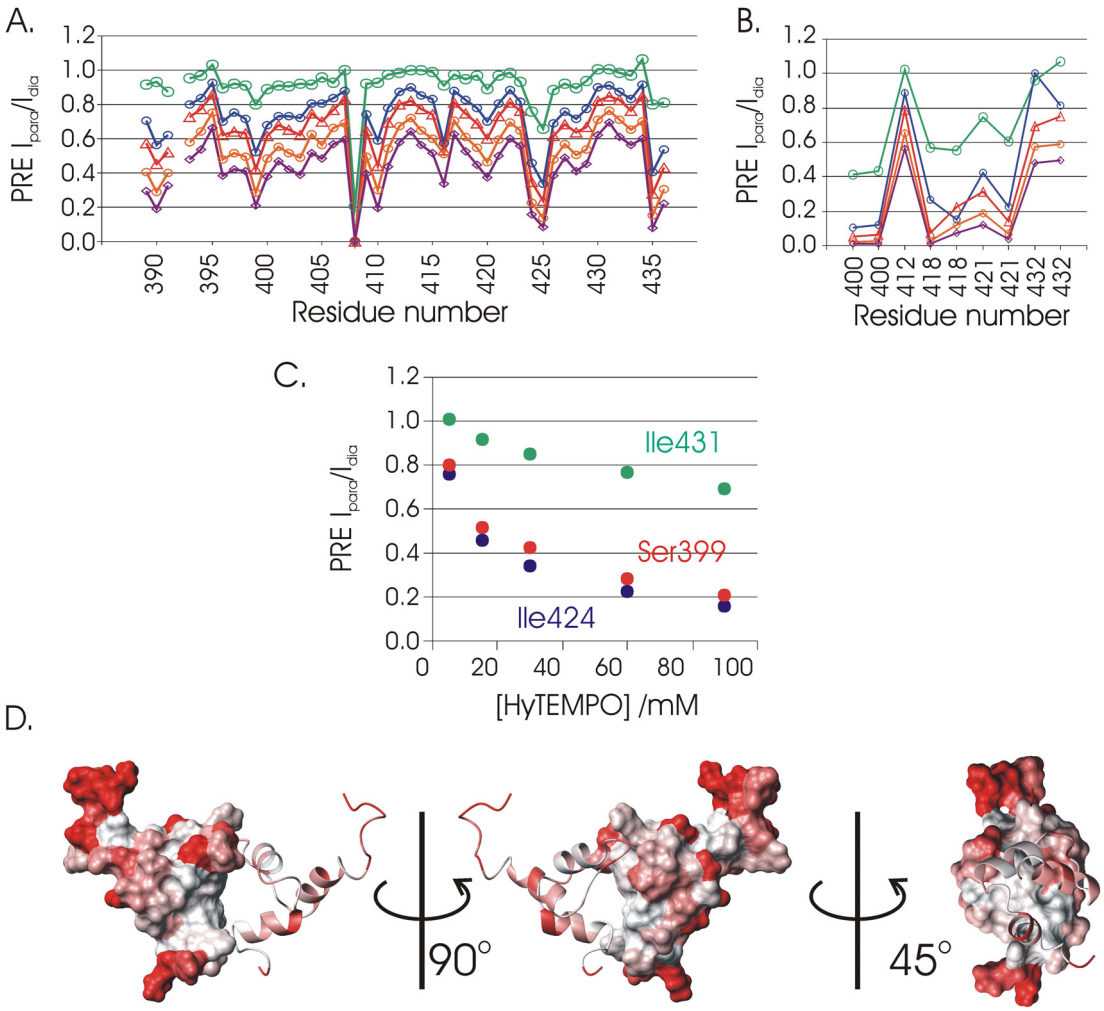



**Figure 6-12: (A-B) Residue specific PRE for the p62-UBA in the presence of HyTEMPO at 5 mM (Green), 15 mM (Blue), 30 mM (Red), 60 mM (Orange) and 90 mM (Purple). Shown for (A) Backbone and (B) Side chain resonances. (C) Observed PRE as a function of HyTEMPO concentration for the backbone NH of residue Ser399 (Red), Ile431 (Green) and Ile424 (Blue). (D) Large PRE plotted onto the surface of the p62-UBA dimer (RPDB ID: 2KNV).**

The majority of the domain shows only very small PRE effects at 5 mM HyTEMPO with the notable exception of Asp408 which is reduced to an intensity of around 20 %, several other residues also show moderate effects (Intensities <80%) including Ser399, Leu416, Ile424, Gly425, Lys435, and His436 (Figure 6-12) suggesting a degree of exposure to the solvent environment for these backbone NH. Noticeably all of the measured side chain (Asn, Gln and Trp) NH or $NH_2$ show strong PRE effects (signal intensities <30%) with the clear exceptions of Trp412 and Gln432 which show little or no PRE. (Intensity >98 %). All resonances show a non-

linear increase in the observed PRE as the concentration of HyTEMPO is increased from 15 mM to 90 mM. At 90 mM HyTEMPO all resonances show PRE with significant PRE seen for Glu389, Ala390, Asp391, Ser399, Asp408, Gly410, Leu416, Lys420, Ile424, Gly425, Lys435, and His436 (Figure 6-12).

The relatively small PRE seen for the majority of residues at low HyTEMPO concentrations is consistent with the observed structure, where the majority of backbone NH are involved in intramolecular hydrogen bonds or buried at the dimer interface. The exceptions to this are the backbone NH of Asp408 in loop one, which is not proposed to be involved in an intramolecular hydrogen bond, and those residues of the C-terminus which are considerably more solvent exposed. The particularly large PRE for Asp408 may potentially indicate an increased lifetime of the HyTEMPO at this site due to some weak interaction, however it may be expected that this would result in large PRE for surrounding residues which does not appear to be the case[215].

The large PRE observed at high HyTEMPO concentrations results from a decrease in the average distance between HyTEMPO molecules and protein molecules in solution, and an increased likelihood of non-specific interactions with the protein surface. Mapping the large PRE effects onto the surface of the p62-UBA dimer shows clear PRE effects across the entire surface, this notably includes clear PRE effects along helix two which is involved in the dimer interface (Figure 6-12). The PRE effects for helix two are reduced compared to the solvent exposed surface regions suggesting that the dimer appears to provide a degree of protection from HyTEMPO in solution. However the degree of protection appears markedly reduced compared to the regions of loop one and the C-terminal portion of helix three which form the non-helix two portion of the dimer interface and which show a high degree of protection (Figure 6-12).

In previous studies PRE effects have been used to probe conformational dynamics in systems which adopt conformations which alter in their degree of solvent exposure[215,216]. The reduced effects seen for the helix two portion of the dimer interface may suggest that this portion of the dimer interface may adopt more open conformations than that suggested by the structure. However it is worth noting that the C terminal portion of the dimer interface which shows particularly low PRE values is rich in charged side chains as well as the C-terminal carboxylate. Previous reports have suggested that the neutral nature of HyTEMPO prevents it from

penetrating the hydration shell surrounding charged side chains preventing transient interactions with the protein surface in their vicinity[215]. It is therefore possible that the lower PRE values observed for the helix three portion of the dimer interface compared to the helix two portions may be contributed to by the hydration shell of charged residues around the C-terminal region.

The large PRE values for the side chain $NH_2$ of asparagine and glutamine residues is coupled to significant CSP values suggesting a more specific interaction in these cases, the large PRE values are consistent with their high degree of solvent exposure. The potential interaction with asparagine and glutamine side chains could lead to a correlation with large PRE values in the proximity of these side chains, leading to misleading results. From simple mapping of the PRE values onto the p62-UBA dimer structure no clear correlation with the proximity to asparagine or glutamine is observed. The only observed $NH_2$ which does not show significant PRE is Gln432 which is located at the C-terminus of helix three which shows significant protection as discussed above.

### 6.3.2 Summary

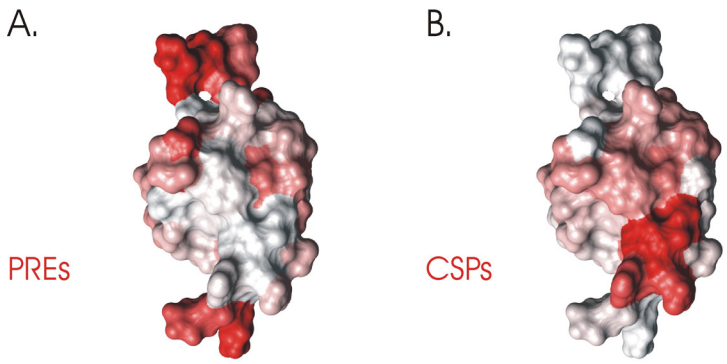



**Figure 6-13: (A) PRE values for the p62-UBA dimer and (B) dimer induced CSP. Plotted onto the structure of the p62-UBA dimer in Red (RPDB ID: 2KNV). Only one monomer unit is displayed to reveal the proposed dimer interface.**

The solution PRE results are wholly consistent with the published structure and complement the dissociation induced CSPs. Both suggest a dimer interface which has the side chain of Trp412 at its centre, extends along helix two, and involves the C-terminal end of helix 3. The dimer interface is stable on the NMR

time scale indicated by the lack of solvent exposure or more accurately the PRE effects from the soluble paramagnetic agent HyTEMPO.

## 6.4 Amide Temperature Dependence of the p62-UBA

### 6.4.1 Theory Behind the Amide Temperature Dependence

In order to gain insights into the structure and dynamics of the p62-UBA dimer and to complement the structural and PRE studies, measurements of the amide temperature coefficients for the backbone NH of the p62-UBA were recorded.

The temperature dependence of the chemical shift of the backbone amide proton has been shown to give a good indication of hydrogen bond donors[219]. The relationship between temperature and the amide proton chemical shift has been shown in several studies to be linear in proteins not displaying conformational averaging up to ~15 K below the $T_m$[63,219,220]. This linear change in chemical shift is the result of linear expansion of the protein molecule with increasing temperature, the expansion leads to an increase in the average hydrogen bond length, leading to a decrease in the deshielding effects caused by the hydrogen bond acceptor[219]. Intermolecular hydrogen bonds to solvent are generally weak, so the temperature induced expansion is relatively pronounced resulting in large negative temperature coefficients. Stronger intramolecular hydrogen bonds, such as those found in regular secondary structure, are generally stronger and therefore resist expansion, giving rise to a smaller temperature dependence[219]. However the chemical shift depends on the inverse third power of the distance between the amide proton and the acceptor, therefore a similar amplitude expansion gives rise to a chemical shift temperature dependence which is larger for shorter hydrogen bonds than for longer hydrogen bonds. Temperature coefficients therefore give an indication of both the presence of hydrogen bonds but also the length and strength of that bond[219].

Studies have shown that in general amide proton temperature coefficients more positive than -4.6 parts per billion (ppb)/K indicate hydrogen bonds with an average value of -3.2 ± 2.0 ppb /K, while amide protons involved in transient solvent hydrogen bonds average -7.1 ± 2.5 ppb/K. These values are not without exception and extremes for both classes have been reported between the ranges -17.5 and +6.7 ppb/K [63,219].

When interpreting amide temperature coefficients a third group must also be considered; the chemical shift of protons which do not form hydrogen bonds but lie close to aromatic rings are strongly influenced by the proximity to that ring and by rotamer conformation, and in some cases the ring π-electron cloud may even act as a hydrogen bond acceptor. This regularly gives rise to gradients that are more positive than -4.6 ppb/K with a typical range of -2 – 0 ppb/K. However values as high as 6.7 ppb/K have been reported[219].

### 6.4.2 Curvature in Amide Temperature Dependence

Several studies have demonstrated that not all amide proton chemical shifts show linear relationships. This deviation from linearity has been suggested to be caused by small conformational changes and could potentially be used to identify and characterise low energy excited states occupied in the native state ensemble [62,219,221,222]. If one considers a single state 1 with a temperature coefficient defined by Equation 6-1A, and a higher energy, alternative state 2, with a temperature coefficient defined as Equation 6-1B.

**Equation 6-1:** **A.** $\delta_1 = \delta^o_1 + g_1T$ **B.** $\delta_2 = \delta^o_2 + g_2T$

Where $\delta_1$ and $\delta_2$ represent the chemical shift at a given temperature T for state 1 and 2, g denotes the gradient of their chemical shift temperature dependence, and $\delta^o_1$ and $\delta^o_2$ denotes their chemical shift at absolute zero. These two states must be separated by a free-energy difference ΔG defined by Equation 6-2.

**Equation 6-2:** $\Delta G = \Delta H - T\Delta S$

Where ΔS and ΔH refer to the entropy and enthalpy difference between the two states. This free energy difference determines the populations of the two states according to Equation 6-3.

**Equation 6-3:** $p_2/p_1 = \exp(-\Delta G/RT)$

Where $p_2$ and $p_1$ are the populations of state 2 and 1 at temperature T and R equals the gas constant. The observed chemical shift can therefore be described by Equation 6-4.

**Equation 6-4:** $\delta_{obs} = \delta_1 p_1 + \delta_2 p_2 = \{(\delta_1 = \delta^o_1 + g_1T) + [(\delta_2 = \delta^o_2 + g_2T) \exp(-\Delta G/RT)]\}/ 1+ \exp(-\Delta G/RT)$

Using this model to simulate the temperature dependence of a two-state system several groups have demonstrated that the free energy difference between the two states has the largest influence on curvature, and that at a free energy difference above 20 kJ/mol the population of the excited state becomes too small to produce curvature. These groups also propose that the excited states reflect only local changes affecting around 5 residues centred around the amide showing curvature[62,221,222].

Andersen and co workers have demonstrated that taking the chemical shift deviation (CSD) into account when interpreting the temperature coefficients of amide protons greatly improves the accuracy. The rationale behind this is a simple one, the chemical shift of the amide proton is largely affected by its hydrogen bonding (as discussed above). The deviation of the chemical shift values from random coil values has been widely used as an indicator of secondary structure by itself[63,141]. It is therefore logical that a combination of these two pieces of information may improve the accuracy of hydrogen bond predictions[63,221]. This has also been employed to explain the sign and magnitude of the curvature of nonlinear temperature coefficients, assuming the excited state is a less folded state the magnitude and direction of the curvature should correlate with the CSD. This has been shown to be true in various cases, though the magnitude has been shown to be dependent more on the ΔG between the two states[63,221].

### 6.4.3 Amide Temperature Dependence of the p62-UBA

Temperature dependence studies were performed on a 1 mM sample of $^{15}$N-p62-UBA. Given the determined $K_{dim}$ of ~10 μM the p62-UBA would be presumed to exist almost exclusively in its dimeric form at this concentration (Section 6.2.2). From CD studies the p62-UBA dimer has a $T_m$ of around 347 K and remains dimeric up to the unfolding transition (Section 6.2.2). $^{1}$H-$^{15}$N-HSQC were recorded at temperatures in the range 278-323 K at 2 or 3 K intervals and referenced to DSS which has a temperature independent chemical shift[51,221]. The proton chemical shift was plotted against temperature and fitted to a straight line. The gradient for each residue against residue number is plotted in Figure 6-14, the average value for hydrogen bonded (Green) and non hydrogen bonded (Blue) protons have been included, the red line represents a cut off where hydrogen bonded (Less negative) and non hydrogen bonded (More negative) protons are more likely[219]. Temperature coefficients were measured for all 47 non proline residues of the p62-UBA domain;

all but 5 of those residues gave the expected negative temperature coefficient. Those protons found in regular secondary structure, which would be expected to be involved in hydrogen bonding, gave consistently smaller temperature coefficients with few exceptions (Figure 6-14).

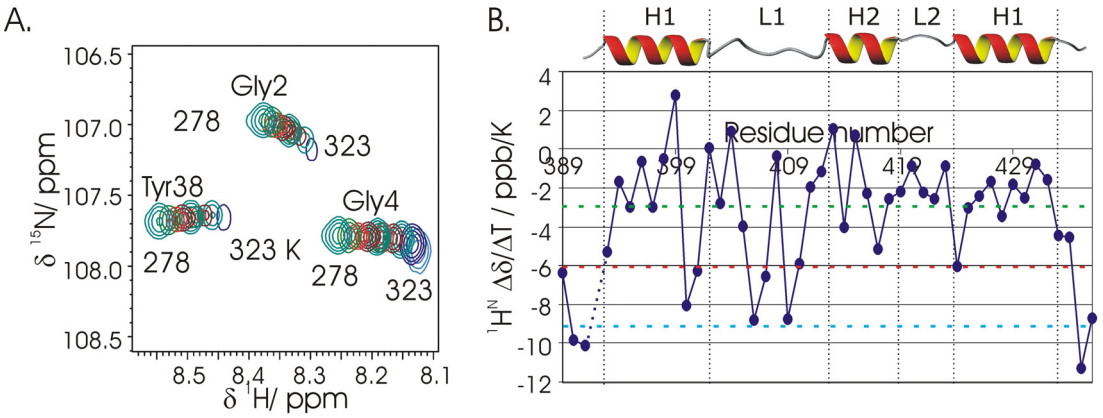



**Figure 6-14: (A) Section of $^{1}H$-$^{15}N$-HSQC of 15N-p62-UBA at a range of temperatures between 278 and 323 K. (B) Plot of amide temperature coefficient against residue number for the p62-UBA. The cut off for hydrogen bonded coefficients (Red line), the average hydrogen bonded amide temperature coefficient (Green line), and the average non hydrogen bonded amide temperature coefficient (Blue line) and the secondary structure boundaries are illustrated (H = Helix L = Loop).**

Temperature coefficients for each backbone amide were plotted against their chemical shift deviation using the approach of Andersen *et al*[63]. Their study showed that plots of chemical shift deviation against temperature coefficients could be used to predict the presence of hydrogen bonding with improved accuracy compared to temperature coefficients alone. Such a plot is displayed in Figure 6-15. Protons in the bottom left of such plots are predicted to be hydrogen bonded while those in the top right as non hydrogen bonded. Andersen *et al* propose that the use of two "cut off" lines could greatly improve the interpretation, and that a cut off line of $\Delta\delta/\Delta T = -2.11 - (2.4CSD)$ (Blue line) yields a false positive rate of <11% while a gradient of $\Delta\delta/\Delta T = -3.37 - (1.8CSD)$ (Red line) gives ~20% false negatives[63].

The NMR structural ensemble for the p62-UBA dimer (RPDB ID: 2KNY) was used to identify hydrogen bonded amides in the p62-UBA. The measured backbone amides were split into three categories; the structural ensemble contains a total of 10 dimeric structures giving 20 monomeric units, backbone amides showing

a consistent hydrogen bond in over 90 % of the structural ensemble were deemed as hydrogen bonded, those forming a hydrogen bond in 90-50 % of the structural ensemble were deemed as potentially involved in hydrogen bonding, while any amide involved in hydrogen bonds in <50% of the ensemble were deemed as not hydrogen bonded (or solvent hydrogen bonded).

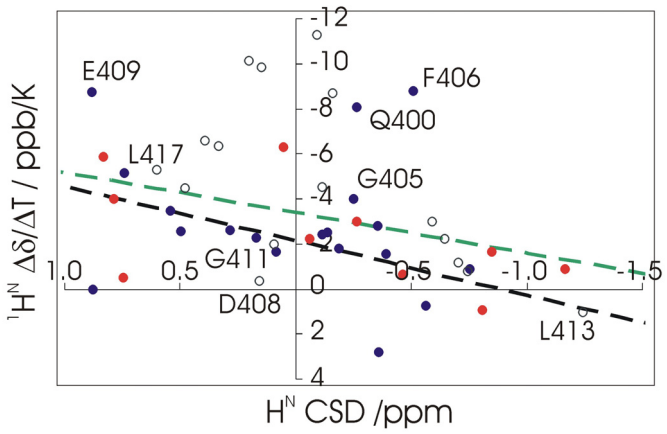



**Figure 6-15: Plot of backbone amide temperature coefficient against CSD, the lowest false positive threshold (Black dashed line) and false negative threshold (Green dashed line) has been included; residues which form hydrogen bonds in 100% of the structural ensemble (Blue), form hydrogen bonds in 50-90 % of the ensemble (Red) or < 50% of the ensemble (Open circles) are indicated (RPDB ID: 2KNV). Residues in unexpected regions are labelled.**

The majority of backbone amides categorised as hydrogen bonded were found in either the hydrogen bonded or potentially hydrogen bonded region of the 'Andersen plot'. Similarly the majority of non-hydrogen bonded amides also fell into the appropriate region. However there were a number of clear exceptions (Figure 6-15).

The backbone amide protons for Gln400, Ser403, Gly405, Phe406, Glu409, and Leu417 are all predicted to be hydrogen bonded based on the p62-UBA dimer structural ensemble (RPDB ID: 2KNV), yet fall above the threshold line for hydrogen bonded amides[63]. Interestingly three of these residues form a continuous section, Ser403-Phe406 at the beginning of loop one, previous dynamics studies of the p62-UBA have demonstrated increased dynamics from residue Gln400 to Leu413 encompassing the C-terminus of helix 1 and the entire loop one[36,210]. The large amide temperature coefficients seen for residues Ser403-Phe406 may indicate that the

hydrogen bonding observed between these residues and residues of helix one may be an artefact of the structural ensemble which is not describing the full structural dynamics of these residues.

A similar rationale may apply to Glu409, in the structural ensemble Glu409 consistently forms an intra-residue hydrogen bond to its own side chain carboxylate. Such intra-residue hydrogen bonds are common for Glutamate side chains in crystal structures[223], however such intra-residue hydrogen bonds are highly dependent on the backbone torsion angles with only a very limited range of preferred combinations[223]. Glu409 has backbone torsion angles which lie just outside of this range. Combine this with the proposed dynamics for Glu409 and surrounding residues and it is possible that again the measured hydrogen bond may be considered as a structural artefact. Of those residues predicted to form an intramolecular hydrogen bond occupying a region of the 'Andersen' plot consistent with solvent hydrogen bonding, Leu417 is the only seemingly unexplainable example. Leu417 sits within helix 2 and is consistent with formation of a stable hydrogen bond and shows very little PRE consistent with no solvent interaction. Leu417 possesses a relatively large CSD and amide temperature coefficient and is close to the threshold value for hydrogen bonded amides. Given the large CSD, the larger that expected temperature coefficient for hydrogen bonded amides may be deemed as acceptable in this case.

Of the residues predicted not to be involved in intramolecular hydrogen bonds only three exceptions are noted these are the backbone amides of Asp408, Gly411, and Leu413 which do not form hydrogen bonds in the structural ensemble for the p62-UBA dimer but occupy regions of the 'Andersen' plot consistent with hydrogen bond formation. Leu413 gives an unusual positive gradient suggesting that this is not determined by hydrogen bond formation. This will be discussed in more detail below. Both Asp408 and Gly411 show relatively small amide temperature coefficient and small CSD suggesting long intramolecular hydrogen bonds[219], however a small amide temperature coefficient could equally be explained by the absence of a hydrogen bond. In the case of Gly411 this may well be the explanation, the backbone NH of Gly411 sits at the N-terminus of helix two and is buried by the side chains of Asp408 and Glu409 resulting in poor solvent exposure which is consistent with low observed PRE values for Gly411. In the structural ensemble Asp408 is clearly solvent exposed which is supported by considerable PRE values, yet it possesses a particularly small amide temperature coefficient consistent with a

long hydrogen bond or a lack of hydrogen bonding. In the structural ensemble, although the backbone NH of Asp408 is solvent exposed it is in close contact with its own carboxylate, this opens two possible sources for hydrogen bonding, the first being an intra-residue hydrogen bond to its adjacent side chain carboxylate. A second possibility is for a hydrogen bond to a surface bound water molecule, which may have a higher than typical resident time on the protein surface as a result of the hydration shell of the adjacent carboxylate. this may result in a more stable hydrogen bond than those typically seen for solvent hydrogen bonds resulting in a small amide proton temperature coefficient.

In general the sign and magnitude of the amide temperature gradients are as anticipated and consistent with the secondary structure as defined by the NMR structure. However five residues experience positive gradients of varying magnitude including Ser399, Leu402, Met404, Thr415, and Leu413 with the highest being Ser399 with a gradient of 2.78 ppb/K, unusually this includes Leu413 which is defined as non-hydrogen bonded. The explanation for positive gradients is not straight forward, it has been suggested that positive temperature gradients originate from structural perturbations such as change in distance from aromatic rings or an increase in aromatic side chain mobility[221]. However of these five residues only Leu413 lies within 5 Å of an aromatic ring and the closest ring to S399 is over 9 Å away. Another explanation for positive gradients is that, in the case of residues that have a very negative CSD and relatively weak hydrogen bonding, the thermal expansion of the protein results in a temperature dependence which is determined by the chemical shift returning to a more random coil state due to increased local motions, giving a positive gradient. Of the five residues, all but Leu402 have relatively large negative CSD and could be explained in this way[221]. Leu402 although records a positive gradient, on closer inspection it actually shows pronounced curvature and the observed positive gradient is a product of the curvature (See section 6.4.4).

### 6.4.4 Curved Amide Proton Temperature Coefficients on the p62-UBA

As discussed above, curvature in the amide temperature coefficient can be used to indicate the presence of low energy conformational states which are sampled at equilibrium. In order to investigate the possibility of low energy excited states, the

recorded amide temperature coefficients were examined in the temperature range corresponding to 15 K below its $T_m$[62,221,222]. Plots of $\delta^1H^N$ against temperature were fitted to a straight line in IGOR Pro™ 5.05A and the residuals examined for curvature. When using temperatures of 278–323 K many residues show a non-linear temperature-dependence in their chemical shifts, with only 9 out of 47 non-proline residues showing linear profiles, these include Glu389, Leu413, Thr419, Lys420, Tyr422, Ala426, Leu428, Tyr433, and Ser434 (Figure 6-16).

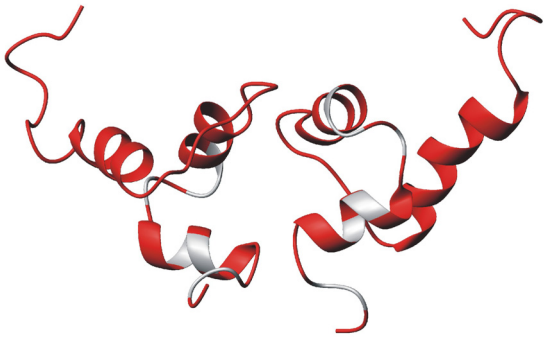


**Figure 6-16: 'High temperature' curvature of amide temperature coefficients mapped onto the surface of the p62-UBA dimer in red (RPDB ID 2KNV).**

Given the broad melting transition for the p62-UBA domain (as seen by CD Figure 6-8), the 15 K below the $T_m$ criteria suggested in previous studies was extended so that only temperatures in the range 278-310K (37 K below the $T_m$) were considered so that early melting events would not influence the results. Under these conditions the number of amide protons showing pronounced curvature was reduced to 11 including Leu398, Gln400, Leu402, Phe406, Ser407, Asp408, Thr414, Arg415, Leu417, Ala427, and Lys435 with a further 10 showing potential curvature including Ile395, Glu396, Ser403, Gly410, Leu416, Asn418, Asp423, Ile424, Asp429, and Lys434 (Figure 6-17). Of the 11 residues showing pronounced curvature 6 show concave curvature including Leu398, Leu402, Asp408, Thr414, Arg415, and Ala427, while 5 are convex including Gln400, Phe406, Ser407, Leu417, and Lys435.

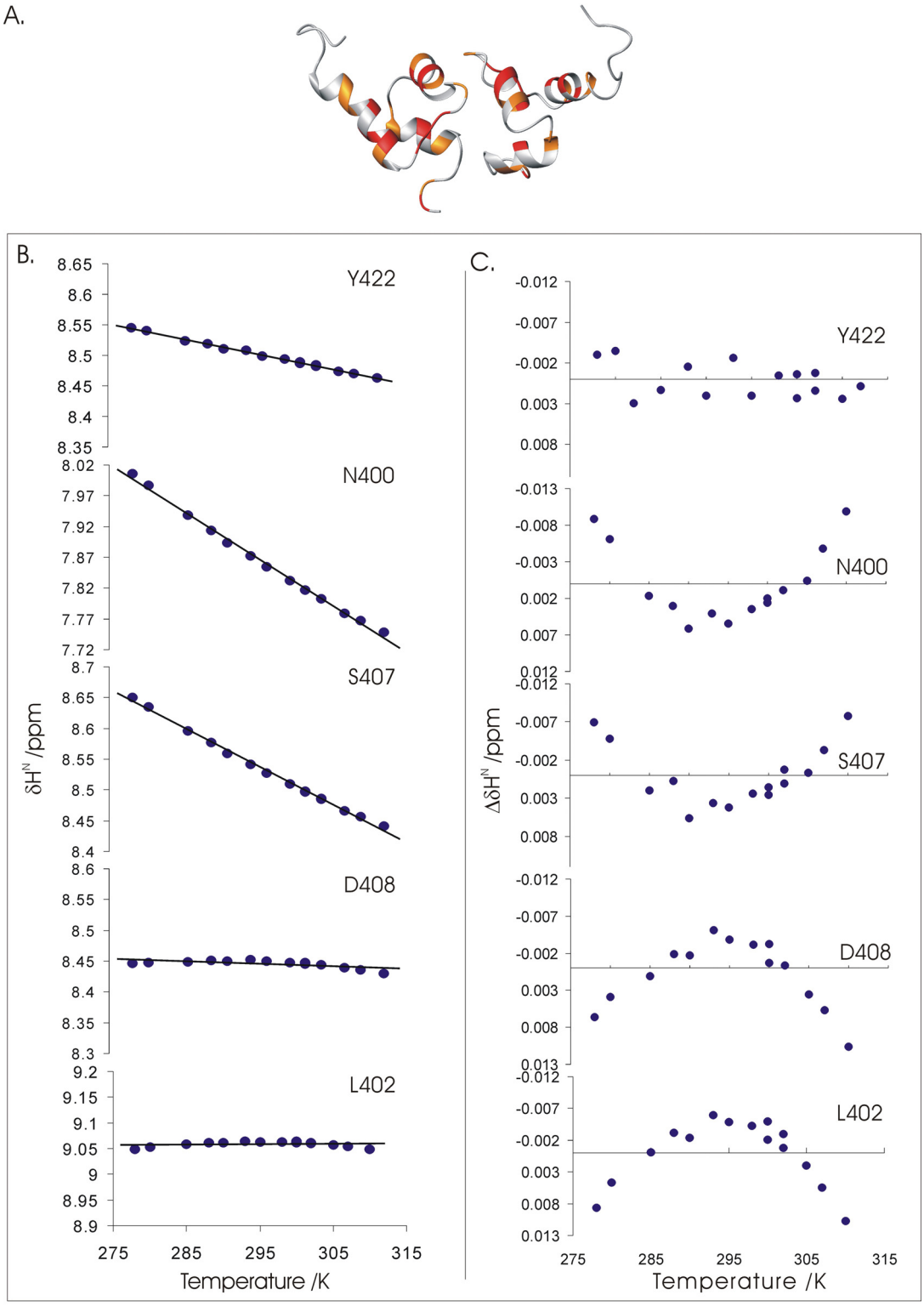



**Figure 6-17: (A) Residues showing pronounced curvature (Red) and potential curvature (Orange) in the temperature range 278 -310 K mapped onto the structure of the p62-UBA dimer (RPDB ID: 2KNV). (B) $\delta^1H^N$ against temperature for residues Tyr422, Gln400, Ser407, Asp408 and Leu402 fit to a linear relationship (C) Residuals to the linear fits in B.**

In order to examine the suggestion that the curvatures seen at high temperatures result from processes associated with the beginning of the unfolding transition, the amide temperature dependence study was repeated on the Paget's disease associated mutation G425R, which has been shown in previous studies to be stabilised compared to wt-p62-UBA with a $T_m$ which is increased by around 5 K. The gradients determined for the p62-UBA-G425R mutant are indistinguishable from wt-UBA with the exception of those residues immediately adjacent to Gly425 and the following turn of helix three which show changes in gradient with Ala426 and Asp429 showing a ~1 ppb/K difference (Figure 6-18). This may suggest a shortening of the hydrogen bonds associated with these NH by around 0.2 Å and may suggest that the G425R mutation leads to a stabilisation of helix 3 which may contribute to the overall stabilisation seen by CD[219].

The curvature seen at higher temperatures for the p62-UBA-wt is still present in the G425R mutant and affects the same set of residues, however in the case of G425R the curvature appears to be reduced and have shifted in temperature, increasing by around 8 K (Figure-6-18). While the origin of this curvature is not necessarily caused by the folding transition the correlation of the shift in the curvature is consistent with the increase in $T_m$ between wt p62-UBA and the G425R mutant, giving a strong indication that this high temperature curvature is associated with the folding transition. Interestingly the curvatures seen at low temperature for the wt p62-UBA affects the same residues in the G425R mutant and have magnitudes and signs which are indistinguishable in most cases (Figure 6-18).

The curvatures seen at high temperature affect the entire domain and may therefore be the product of a global folding linked transition, however the 11 low temperature curvatures, which are present in both the wild type protein and the G425R mutant, are more likely to be characteristic of low energy excited states[62,221]. The origins of non-linear temperature coefficients has been discussed in several studies in the literature and are thought to be the product of local excited conformations which are 4 -10 kJ/mol above the ground state[62,221,222,224]. It is thought that these excited states involve up to five contiguous residues centred on the residue showing curvature and that these excited conformations are the product of local unfolding, subtle conformational changes, changes in aromatic side chain position or motions, and different protonation states of histidines[62,221,222]. In the case of the p62-UBA the observed curvatures are spread out over the whole domain without clear

correlation. This could indicate that no single alternative state can explain the observed curvatures. Each curved amide was therefore examined individually for the potential origins of these effects.

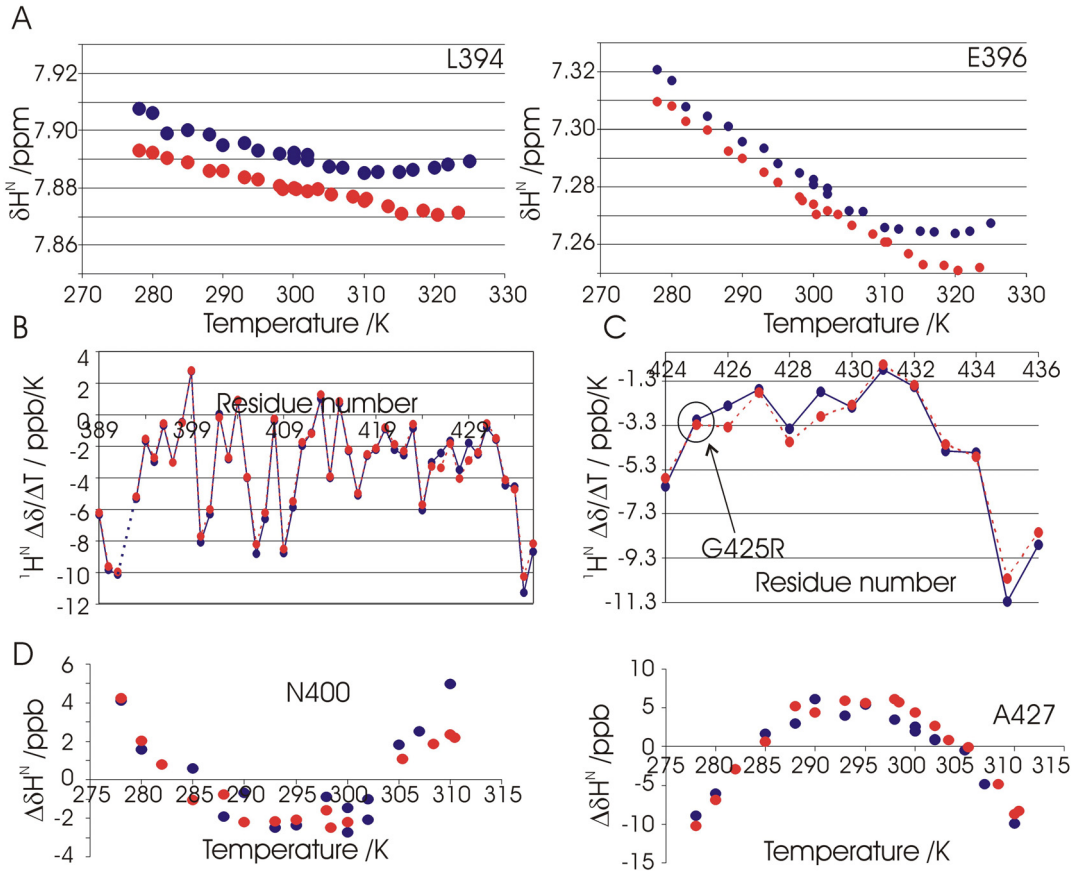



**Figure 6-18: Summary of the amide temperature data for G425R: (A) $\delta^1H^N$ plotted against temperature in the range 278-323K for L394 and E396 for p62-UBA-Wt (blue) and G425R (red). (B) Plots of $^1H^N\Delta\delta/\Delta T$ against residue number for p62-UBA-Wt (blue) and G425R (red). (C) An expansion of helix three (from B). (D) Residuals to the linear fit for the temperature coefficients for N400 and A427 from p62-UBA-Wt (blue) and G425R (red).**

If the excited state is a partially unfolded state then whether the curvature of an NH temperature coefficient is concave or convex should correlate with the CSD of the measured proton. If the excited state is less folded than the native state then the chemical shift should have a smaller CSD, if this is the case then concave curvatures should be seen for residues with a positive CSD and a convex curvature seen for residues with a negative CSD. This is the case for all but 3 of the 11 residues, with the exceptions including Ser407, Arg415, and Leu417 (Figure 6-19).

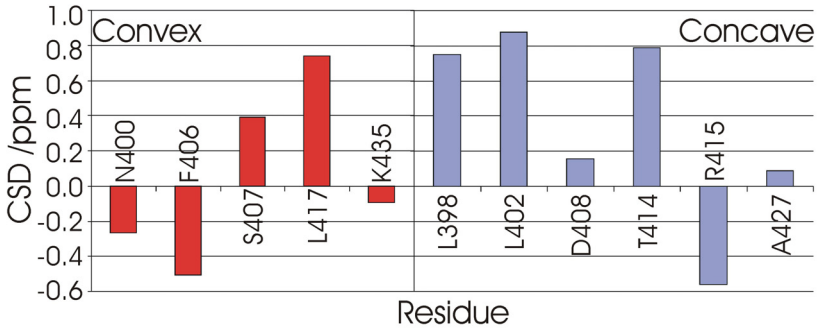



**Figure 6-19: CSD for convex (Red) and concave (Blue) curved temperature profiles.**

Arg415 has a concave curvature with a negative CSD, however Arg415 has a positive temperature coefficient in the native state. An excited state with a negative gradient like that seen for the majority of residues, or a smaller CSD would result in a concave curvature. Both a small CSP for Arg415 and a large negative temperature coefficient would suggest a less folded excited state and result in a concave temperature coefficient.

Only Ser407 and Leu417 have curvatures which are characteristic of a larger CSD and therefore more folded-like chemical shift in its excited state. The most obvious explanations for Ser407 come from the neighbouring aromatic ring of Phe406. If the mobility of loop 1 increases with temperature this may alter the effect of the π electron cloud of the phenylalanine ring on Ser407. Similarly Leu417 lies close to the ring of Tyr422 in loop 3, and the observed gradients therefore may be influenced by changes in mobility of these aromatic side chains or even transient $H^N$-π interactions [158].

The 11 residues showing pronounced curvature are spread out across the p62-UBA structure, however as discussed above, one potential cause of the measured excited states may be changes in aromatic side chain positions. All of the curved amides are located relatively close to aromatic side chains. During an investigation into the folding landscape of SUMO, Kumar *et al* noted that ring flipping can cause structural perturbations including local unfolding that could give rise to curved temperature coefficients. This may well contribute to the observed curved temperature dependence[221].

Residues Leu398, Leu417, and Ala427 all lie close together in space and form contacts across the ends of helices 2 and 3 to the middle of helix 1 (Figure

6-19). It is possible that these three residues undergo a local conformational change such as opening of the ends of the three helices possibly as an early step in unfolding or in the proposed conformational rearrangement which would require a separation of the three helices. However as discussed above this may equally be explained by the proximity of Tyr422 to these residues, which have both main chain and side chain atoms in close proximity to the Tyr ring, therefore rotamer conformations of Tyr422 may well be influencing structural changes for these residues(Figure 6-20).

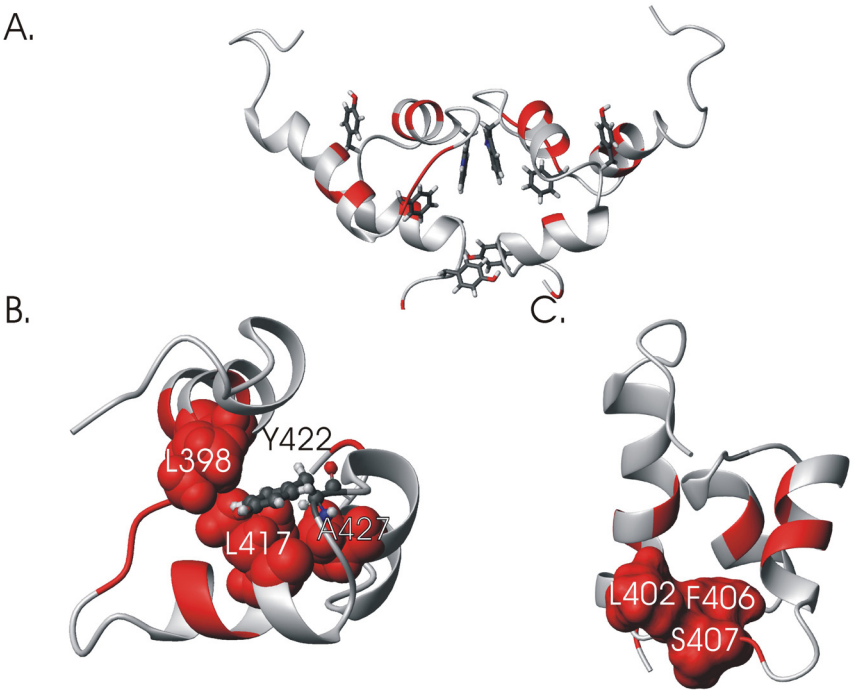



**Figure 6-20: (A) Position of amides with non-linear temperature coefficients (Red) compared to aromatic residues, in the structure of p62-UBA (RPDB ID: 2KNV). (B) Illustration of contacts between residues L398, L417 and A427 and their position relative to Ty422 (C) Contacts between Helix 1 and loop 1. In both cases only one monomer unit is included for clarity.**

Residues Leu402, Phe406 and Ser407 all lie at the interface between helix 1 and loop 1 and show potential contacts (Figure 6-20). As discussed previously loop 1 shows increased dynamics and any potential contacts between helix 1 and loop 1 may be transient and would most likely be interchanging between two or more conformations resulting in curvature. This may also contribute to curvature in Gln400 which lies on the same helical turn as Leu402 and Leu398 which is hydrogen bonded to Leu402. Alternatively these could be the result of unwinding of the

extreme C-terminus of helix 1 which would be consistent with curvatures towards an unfolded state.

The curved amide temperature dependence for Asp408 maybe consistent with the previous observation of an unusually low temperature gradient and high PRE. As the backbone amide is in close proximity to its side chain carboxylate, it is possible then that this is involved in an intra-residue hydrogen bond or stable solvent hydrogen bond, both of which may be transient causing Asp408 to exchange between alternative observed states causing curvature[223].

Finally the curvature associated with Lys435 may be the result of the adjacent His436 side chain, the buffer in which all NMR experiments was recorded is at pH 7, the $pK_a$ of the histidine side chain is ~ 6 – 6.5, at pH 7 it may be expected to be in equilibrium with a protonated form resulting in repulsion between the side chains of Lys435 and His436 resulting in occupation of an alternative state giving rise to curvature.

It is worth noting that a linear profile does not exclude the presence of excited state conformers and that a very small change in chemical shift and small difference in gradient between the native and excited state could lead to an undetectable curvature. Overall the precise nature of the excited states can not be unambiguously determined only suggestions made about their likely origin. However the excited states do not seem consistent with a cooperative alternative state linked to dimer dissociation or the proposed conformational switch[61,180].

## 6.5 The p62-UBA is Isolated by a Long Flexible Linker

To date all structural and biophysical studies of the p62-UBA domain have been performed on the isolated domain alone. However the behaviour of an isolated domain can be subtly different in the context of the full length protein or by simply adding the adjacent regions. In the case of p62, the UBA domain is one of three known domains including an N-terminal PB1 domain followed immediately by a small ZZ type zinc finger. The UBA domain is separated from these two adjacent domains by a linker of around 200 residues which hosts a variety of binding sites for proteins including LC3, TRAF6 and LIMD1. Bioinformatics searches predicts no folded domains within this linker, with a BLAST search showing no regions with high sequence similarity to known domains[181,182,191].

The 100 amino acid region immediately preceding the UBA domain, which sits at the extreme C-terminus, contains no known binding patches and contains a so called 'PEST' sequence. PEST sequences are characterised by an unusually high incidence of Pro, Glu, Asp, Ser, and Thr residues and previous structural studies have demonstrate them to be largely disordered regions which may function as control elements for protein stability[185,225-227].

In order to better characterise the p62-UBA domain the domain was extended N-terminally in a stepwise manner with the intention of gaining insight into the structure and potential function of those regions affecting the p62-UBA domain.

### 6.5.1 Regions Adjacent to the p62-UBA are Intrinsically Disordered

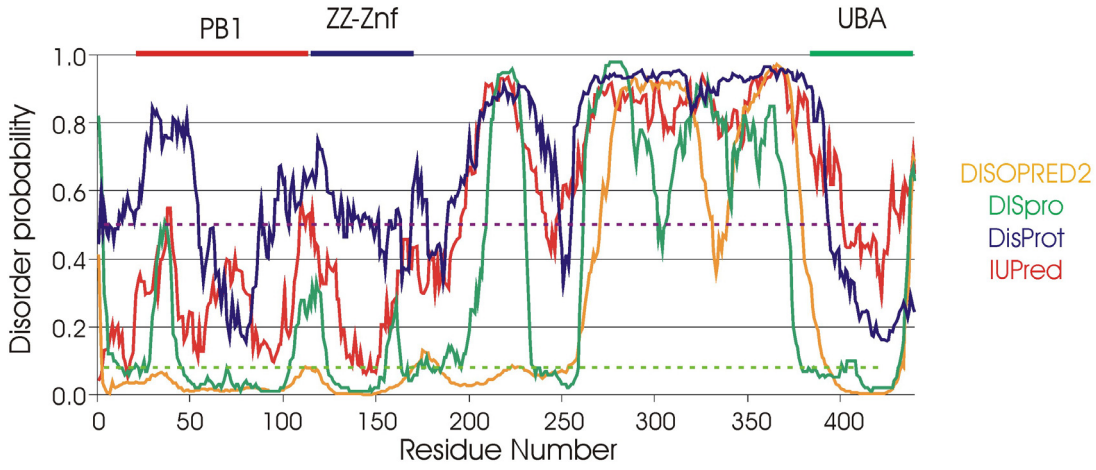



**Figure 6-21: Disorder prediction for the p62 protein using DisProt[92](Blue), IUPred[93,166](Red), DISpro[94](Yellow) and DISOPRED2[95](Green), Disorder threshold values are indicated for DisProt and IUPred (Purple dashed line) and DISpro and DISOPRED2 (Green dashed line) the position of the three folded domains are highlighted.**

The p62 protein sequence was analysed for disorder using the programs DisProt[92], IUPred[93,166], DISpro[94], and DISOPRED2[95]. All of the programs predicted low disorder for the three domains, with the exception of DisProt which only predicted some structure for the ZZ-zinc finger but successfully predicted the PB1 and UBA domains (Figure 6-21). All predictors indicate that the two PEST regions are highly disordered, however interestingly all of the predictors suggested a loss of

disorder between Glu230- Glu260 suggesting a propensity for folding within this region for which there is no structural data to date (Figure 6-21).

In order to examine the structure for the regions immediately N-terminal to the p62-UBA, p62 constructs corresponding to residues 341-440, 261-440 and 221-440 were analysed by far-UV CD spectroscopy and compared to the p62-UBA domain. As discussed above the CD spectrum of the p62-UBA gives a spectrum which is consistent with a three helix bundle with strong negative ellipticity bands at 222 nm and 208 nm. In the CD spectrum of the N-terminally extended construct 341-440, the strong band at 222 nm is maintained, however the ellipticity at 208 nm is increased significantly by the addition of the extra residues. This trend is seen again when the construct is extended further to include residues 261-440 again suggesting little or no contribution from the new regions to the secondary structure content of the CD spectrum with only extra ellipticity at 208 nm observed consistent with an increase in disordered protein content (Figure 6-22). Further extension of the N-terminal region to include residues 221-440 again showed little or no extra structured material but unexpectedly seemed to contribute only very little to the disordered region (Figure 6-22). On first inspection this may be unusual, however it is worth noting that although the increase in ellipticity of the 208 nm band is likely to be a product of an increase in disordered protein, disordered CD spectra do not have a minima at 208 nm but at around 195 nm[41]. By plotting the minima wavelength as a function of the length of the N-terminal extension, it can be seen that as the extra residues are included the intensity of the band originally at 208 nm does increase but also shifts from 207 nm in the p62-UBA to 202 nm in the 221-440 construct with an almost linear dependency (Figure 6-22).

In order to confirm this proposed disordered region, $^{1}$H-$^{15}$N-HSQC were recorded on the p62 constructs 341-440 and 261-440 and compared to the $^{1}$H-$^{15}$N-HSQC of the p62-UBA. Comparison of the $^{1}$H-$^{15}$N-HSQC of the 341-440 and the 261-440 constructs with that of the p62-UBA show low CSP for all but 5 residues including Glu389, Ala390, Asp391, Lys435, and His436 (Figure 6-23). The change in those 5 residues can be attributed to them lying close to the domain boundaries and therefore next to the residues of the extended construct causing changes in local environment. The low CSP can also be seen in the $^{1}$H-$^{15}$N-HSQC -NOESY spectrum of the 341-440 construct where the majority of cross peaks are maintained for the folded UBA domain. Residues Glu389 and Ala390 of the p62-UBA were assigned

using the expected inter-residue NOE's to identify the unique PEA sequence using the $^{1}H$-$^{15}N$-HSQC -NOESY spectrum. The high conservation of the chemical shifts of the p62-UBA in these longer constructs suggests a lack of intramolecular contacts between the UBA domain and the added PEST sequences.

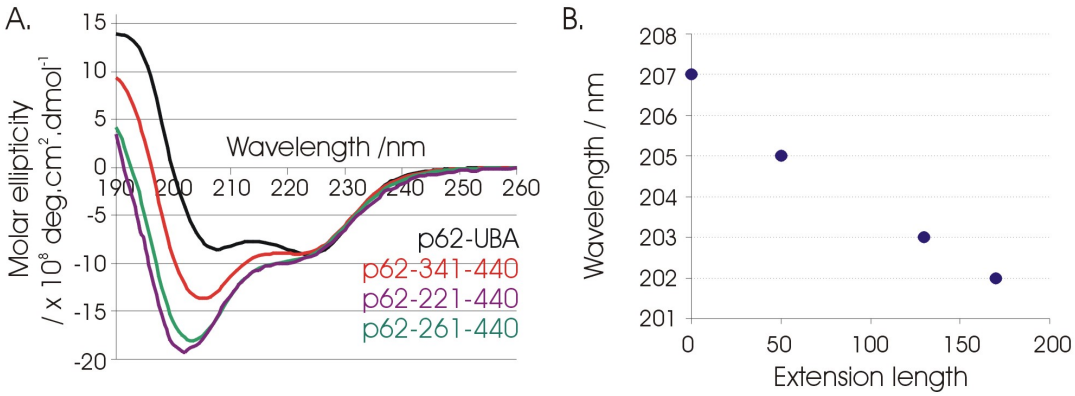



**Figure 6-22: (A) CD spectrum for the p62-UBA (Black), p62 residues 341-440 (Red), 261-440 (Blue) and 221-440 (Purple). (B) Plot of the wavelength of the '208 nm' minima as a function of the length of the extension N-terminal from the p62-UBA.**

Both the 341-440 and 261-440 give good quality spectra although there is evidence for degradation products in the 261-440 spectrum which increased over a period of days at room temperature. By comparing the $^{1}H$-$^{15}N$-HSQC of the p62-UBA, the 341-440 construct and the 261-440 constructs it is evident that those cross peaks not corresponding to residues of the p62 show poor CSD, with chemical shifts restricted to a narrow chemical shift range between 7.7-8.5 ppm consistent with literature values for random coil peptides (Figure 6-23)[63,141,168].

Non p62-UBA residues which show poor CSD also show only weak intra residue NOE in $^{1}H$-$^{15}N$-HSQC-NOESY spectrum, little or no inter residue NOE and intense NOEs to solvent. These features are consistent with an intrinsically disordered chain which undergoes unrestrained conformational dynamics and is fully exposed to solvent. Although a $^{1}H$-$^{15}N$-HSQC for the p62-221-440 construct was not recorded, the 1D proton spectrum of the amide proton region recorded in $H_2O$ was compared to that in $D_2O$ to identify amide protons and shows no evidence for improved CSD with its amide proton signals clustered into the 7.7-8.5 ppm region with no evidence for extra structured regions (Figure 6-23).

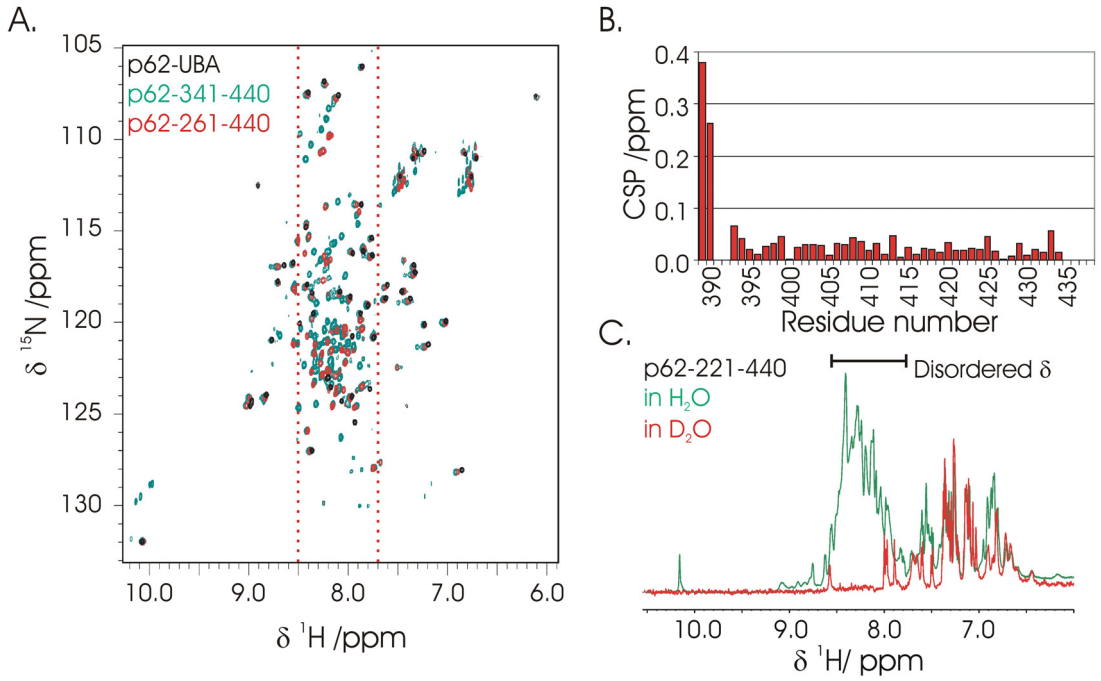



**Figure 6-23: (A) $^{1}$H-$^{15}$N-HSQC spectra of the p62-UBA (Black) and p62 residues 341-440 (Red) and 261-440 (Blue), the chemical shift range expected for intrinsically disordered residues is illustrated with red dashed lines. (B) CSP for the p62-UBA when extended to include residues 341-440. (C) 1D spectrum of the amide region of the 221-440 construct in $H_2O$ (Blue) and $D_2O$ (Red).**

### 6.5.2 Intrinsically Disordered Regions do Not Alter the Stability of the p62-UBA Dimer

NMR structural characterisation of the p62-UBA extended to include resides 341-440 shows no evidence for structural alterations to the p62-UBA, however the construct used for the p62-UBA includes residues 387-436, the 341-440 construct therefore extends at both termini and these extensions have the potential to contribute or interfere with the dimerisation interface.

ESI-MS spectra were recorded on the 341-440 construct to confirm the oligomerisation state and to confirm the absence of modifications or cofactors within the PEST sequences. The ESI-MS spectra revealed a native mass of 11336.6 ± 1.0 consistent with the expected monomer mass of 11337.6, as with the p62-UBA dimer species were clearly visible with no evidence for higher order structures.

In order to examine the potential for changes in the affinity of the p62-UBA dimer, ITC dilution experiments was recorded whereby concentrated p62-UBA (~300 μM) was diluted by sequential injections into buffer in order to observe the calorimetric response of dimer dissociation induced by rapid dilution. As before the resulting dissociation isotherm of the integrated heat response was fitted to a dimer dissociation model to give a $K_{dim}$ = 7.6 ± 1 μM and a ΔH of 13 ± 4 kJ/mol at 298 K. From this a $\Delta G_{298}$ of 29.4 ± 1.0 kJ/mol and a ΔS of -45 ± 8 J/mol/K was determined. The values are consistent with those for the p62-UBA domain suggesting that the extension to include residues 341-440 has little or no effect on the p62-UBA dimerisation (Figure 6-24 and Table 6-3).

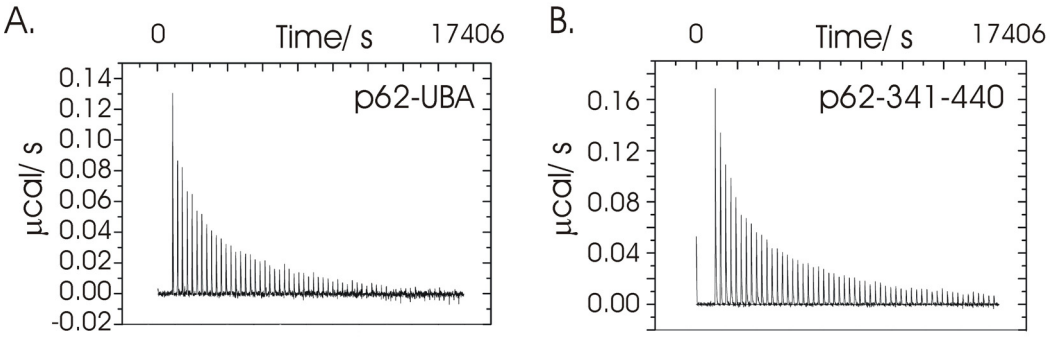



**Figure 6-24:ITC dissociation isotherms for the; (A) p62-UBA. (B) p62 341-440.**

**Table 6-3: Thermodynamic parameters for the p62-UBA and p62-341-440 dimer derived from ITC.**

| | $K_{dim}$ (±) /μM | ΔH (±) / kJ/mol | $\Delta G_{298}$ (±) kJ/mol | ΔS (±) / J/mol/K |
|---|---|---|---|---|
| Wt | 7.1 (1.0) | 6.8 (3) | 29.4 (1) | -75 (5) |
| 341-440 | 7.6 (1.0) | 13.4(4) | 29.4(1) | -45 (8) |

### 6.5.3 Summary

The combination of far-UV CD and NMR data strongly suggests that the regions immediately N-terminal to the p62-UBA are intrinsically disordered and do not alter the structure of the p62-UBA or the affinity of its dimerisation. The separation of the UBA domain from the remainder of the protein by a flexible linker is an architecture seen for many Ub-binding receptors, particularly those which act as proteasomal shuttling proteins[164,171]. The rationale behind this is still unclear, although one explanation might be that in order for a scaffold protein such as p62 to

be able to bind its multiple binding partners effectively to form functional multi-protein complexes, the large unstructured region may act to simply allow for steric bulk attached to the protein. Alternatively they may facilitate the so called fly-casting mechanism whereby large disordered regions function to increase the diffusional radius of the domain to increase its 'capture radius' in order to improve the kinetics of binding[172].

## 6.6 Effects of Paget's Disease of Bone Mutations in the p62-UBA

Much of the literature interest in the p62-UBA revolves around the clustering of mutations linked to Paget's disease of bone (PDB) within its domain boundaries[170,202,204,228,229]. The functional implications of these PDB mutations remain unclear, although disruption of the ability of the p62-UBA to interact with Ub has a clear role in most cases[202,204]. In order to gain insight into the molecular mechanism by which PDB mutations alter the function of p62 and give rise to a pre disposition to PDB, the structural and functional effects of PDB mutations in the p62-UBA have been investigated.

In principle there are a number of ways by which PDB mutations may disrupt the function of the p62-UBA domain, focusing on its interaction with Ub. The first would be through direct interference with the interaction between the p62-UBA domain and Ub, either by removal of side chains which are involved in key interactions or the insertion of side chains which prevent interaction on steric or electrostatic grounds. The second is a related mechanism whereby mutations alter the overall tertiary structure of the p62-UBA domain so that it no longer presents the same interaction surface as the wt-protein distorting the Ub binding patch. Mutations may affect the stability of the p62-UBA domain so that at cellular temperatures a significant portion of the protein exists in an unfolded state which can not perform the functions of the folded domain; such destabilisation may also decrease the proteins cellular half-life leading to low cytosolic concentrations[227]. Finally mutations may disrupt the function of the p62-UBA by altering the dimerisation affinity of the p62-UBA altering the populations of the monomer and dimer species at equilibrium.

The investigations presented here are focused on the Paget's disease mutations S399P, G411S, and G425R which have previously been identified in patients and have been proposed to interfere with the Ub binding properties of the p62-UBA with varying severity[207,208,230]. The most common mutation P392L[207], and the related mutation P387L, which have been investigated structurally previously within the group are also featured[36,208].

## 6.6.1 Biophysical Characterisation of PDB Mutants

### 6.6.1.1 CD studies of PDB mutations

In order to investigate the effect of PDB mutations on the stability of the p62-UBA domain, equilibrium unfolding experiments were performed by monitored by far-UV CD spectroscopy, on the mutants S399P, G411S, and G425R. As with the studies of the wt p62-UBA the ellipticity at 222 nm was recorded as a function of temperature in the range 283 – 368 K. The apparent $T_m$ for the transition was calculated at a range of concentrations in the range of 1 – 200 μM (Figure 6-25 and Table 6-2).

The observed CD spectrum for the examined PDB mutants gave the same classical helical CD spectrum with no significant differences between the observed spectrum and that seen for wt p62-UBA at high or at low concentrations. The S399P mutation however appears to have significantly lower ellipticity than the other mutants. This may be an artefact of the concentration measurements or an indication of a structural change. All PDB mutants of the p62-UBA give strong negative bands at 222 nm and 208 nm, as observed for the wt p62-UBA. The ratio of $\theta_{222}/\theta_{208}$ shows clear concentration dependence which is relatively invariant at high concentration then reduces from ~1.1 to around 0.9 at low μM concentrations consistent with a $K_{dim}$ in the low μM range (5-20 μM) (Figure 6-25). The only exception to this is the G425R mutant which shows only weak concentration dependence in the $\theta_{222}/\theta_{208}$ ratio with a ratio of less than 1 not measured at any point. Assuming that the observed change in $\theta_{222}/\theta_{208}$ ratio is associated with the dimer to monomer population shift (as discussed above), the reduction in concentration-dependence may indicate an increase in the dimer stability so that the population is no longer shifting significantly in the measured concentration range. Alternatively this may suggest that the tertiary structure of G425R is similar in terms of helical packing

and/or pitch in the monomer to that seen in the dimer which results in the relatively high $\theta_{222}/\theta_{208}$ ratio[67,168,212,213](Figure 6-25).

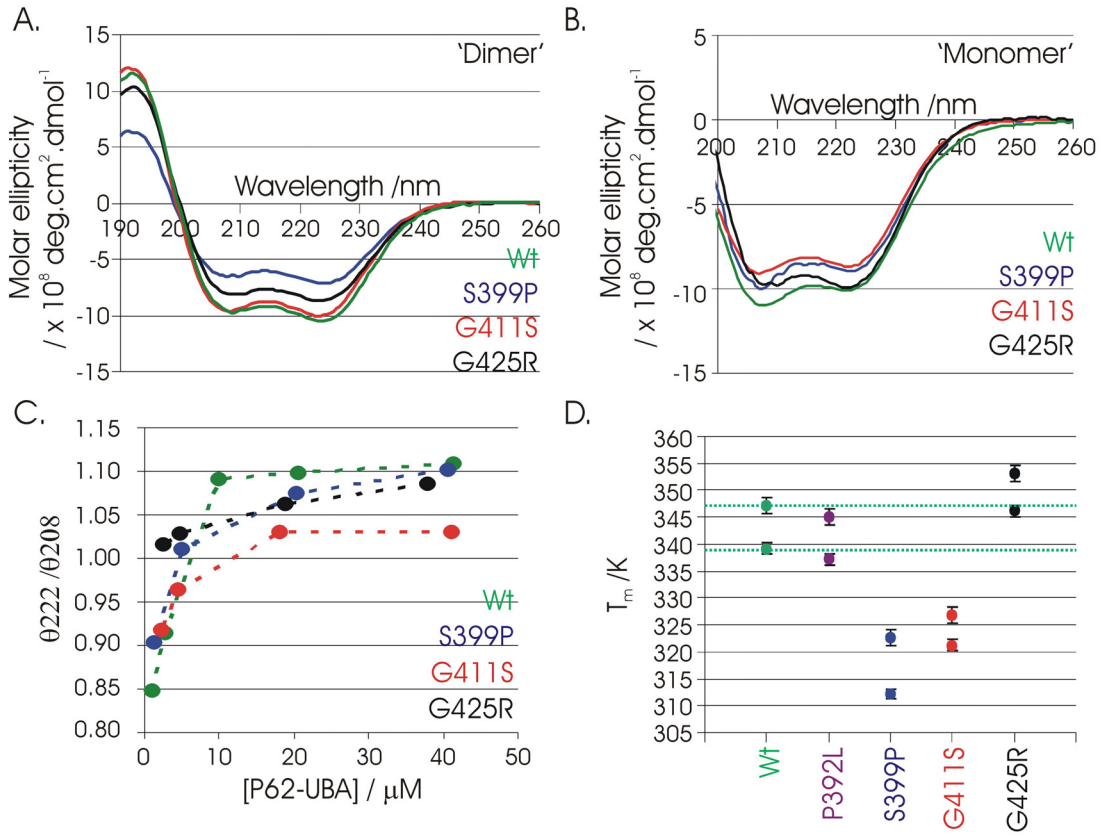



**Figure 6-25: Far-UV CD spectra for the p62-UBA wt (Green) and the PDB mutants S399P (Blue), G411S (Red) and G425R (Black) recorded at (A) high concentration (200-150 µM) and (B) low concentration (4-1 µM). (C) Plots of the concentration dependence of the $\theta_{222}/\theta_{208}$ ratio for p62-UBA mutants. (D) Graph of the $T_m$ of the measured PDB mutants, upper values represent highest measured value at high concentrations, lower value represents the lowest measured value at low concentration; values for the wt p62-UBA are indicated with a horizontal line.**

Equilibrium unfolding experiments were also performed on the P392L and P387L mutants; both of these mutations are close to the N-terminus of the p62-UBA domain (387-436) and were therefore studied in the p62 341-440 construct. The 50 extra residues resulted in increased signal overloading at low wavelengths reducing spectral quality and limiting the maximum recordable concentration to around 50 µM. The reported values are estimates based on the available data recorded between

~50 μM and 2 μM. It is worth noting that the $T_m$ of the p62-UBA and 341-440 construct are indistinguishable and they possess near identical $K_{dim}$.

The observed $T_m$ for the p62-UBA at high and low concentration varies both in the recorded values and in the difference between high and low concentrations for different PDB mutants. Both the P392L and P387L give little or no change in the recorded Tm with P392L showing a marginal reduction in both the $T_m$ of the high and low concentration values (Table 6-4). Both the S399P and G411S record significantly reduced $T_m$ by around 24 K and 21 K respectively at high concentrations, and by 27 K and 18 K at low concentration, suggesting that both the monomer and dimer states are destabilised as a result of these mutations (Table 6-4). This loss of stability may well contribute to the mechanism by which these two mutations disrupt the function of the p62-UBA.

**Table 6-4: Thermodynamic parameters for PDB mutants of the p62-UBA derived from equilibrium unfolding experiments measured by far-UV CD. The fraction folded (FF) at 298 K and 310 K.**

| | $T_m$ (±)/ K | $\Delta T_m$ / K | $\Delta G_{298K}$ (±)/ kJ/mol | $\Delta H$ (±)/ kJ/mol | $\Delta S$ (±)/ J/mol/K | $FF_{298K}$ / % | $FF_{309K}$ / % |
|---|---|---|---|---|---|---|---|
| | 'Dimer' | | | | | | |
| Wt | 347.2 (1.5) | - | 31.6 (1.8) | 229.8 (1.6) | 660 (6) | 99.9 | 98.5 |
| P392L | 345.0 (2) | -2.2 | 24.5 (2.5) | 188.9 (8.4) | 551 (9) | 99.2 | 97.5 |
| S399P | 322.8 (1.6) | -24.4 | 13.8 (1.3) | 188.5 (1.2) | 586 (5) | 95 | 82.0 |
| G411S | 326.8 (1.5) | -20.4 | 15.7 (1.2) | 185.0 (1.7) | 568 (4) | 99.5 | 93.8 |
| G425R | 353.2 (1.5) | 6 | 39.0 (2) | 255.9 (1.9) | 727 (7) | 99.9 | 99.6 |
| | 'Monomer' | | | | | | |
| Wt | 339.2 (1) | - | 24.1 (0.6) | 201.4 (2.9) | 595 (2) | 99.2 | 96.7 |
| P392L | 337.2 (1.5) | -2 | 15.3 (0.9) | 136.6 (3.3) | 406 (4) | 99.0 | 96.0 |
| S399P | 312.2 (1) | -27 | 5.7 (1.1) | 114.2 (7.6) | 363 (4) | 94.9 | 78.8 |
| G411S | 321.3 (1) | -17.9 | 12.5 (0.9) | 167 (1.4) | 519 (3) | 96.3 | 89.2 |
| G425R | 346.2 (2) | 7 | 33.1 (0.6) | 240.0 (8.3) | 694 (2) | 99.8 | 99.0 |

Although the reduction in stability is considerable and the resulting $T_m$ are relatively low (Table 6-4) in order for them to disrupt function based on stability alone the PDB mutations must be considered in terms of the relative populations of folded and unfolded protein at equilibrium. The typical cellular temperature has recently been measured to be around 309 K[231]. At 309 K the wt protein would be expected to have a fraction of its population in the folded state equal to around 99 % at high concentration (the dimer) and 97 % at low concentrations (the monomer). For the G411S mutant these numbers are reduced to around 94 % and 89 % which are reduced further in the S399P mutant to 82% and 78 % demonstrating a significant loss in the population of folded protein for these two PDB mutants (Table 6-4). However in both cases, the majority of protein remains in its fully folded state suggesting that while stability may well play a role it seems likely that other factors are involved. It is also worth noting that reports have suggested that unfolded regions of proteins may act as weak signals for degradation within the cell, and the increased population of unfolded p62-UBA in these mutants may well increase the rate of degradation within the cell impairing the function of p62 by reducing its cellular concentration[227].

The opposite is seen for the PDB mutant G425R which shows a significant increase in the $T_m$ compared to wt p62-UBA with an increased $T_m$ at high concentration of around 5.5 K and of around 7 K at low concentration. The ratio of $\theta_{222}/\theta_{208}$ suggested that the G425R mutant may not undergo a full population shift from dimer to monomer in the concentration range examined here, if this is the case then the increase in $T_m$ of 7 K for the monomer may be slightly higher than the actual value. However even if this is the case, the 5 K increase in $T_m$ for the dimer is significant. The G425R mutant would be expected to be almost completely folded with a fraction folded at 309 K of over 99 % for both the dimer and monomer, therefore instability does not contribute to the G425R disease mechanism. It should be noted however that an increase in stability does not necessarily maintain function; the increased stability may reduce dynamics involved in partner recognition of other functional aspects or prevent other conformational changes required for function.

#### 6.6.1.2 ITC Studies of PDB Mutations

In order to examine the potential for changes in the stability of the p62-UBA dimer, ITC dilution experiments were recorded. As before the resulting dissociation

isotherm of the integrated heat response was fitted to a dimer dissociation model to give $K_{dim}$ and ΔH, from which $\Delta G_{298}$ and ΔS were determined. The general appearance of the calorimetric heat response of all PDB mutations was similar (Figure 6-26) with all giving the expected endothermic heat response consistent with dimer dissociation. The determined value of $K_{dim}$ was in the low μM range in all cases consistent with the changes in $\theta_{222}/\theta_{208}$ ratios observed by far-UV CD, however significant variation in the determined $K_{dim}$ was observed across the PDB mutations of p62 (Table 6-5).

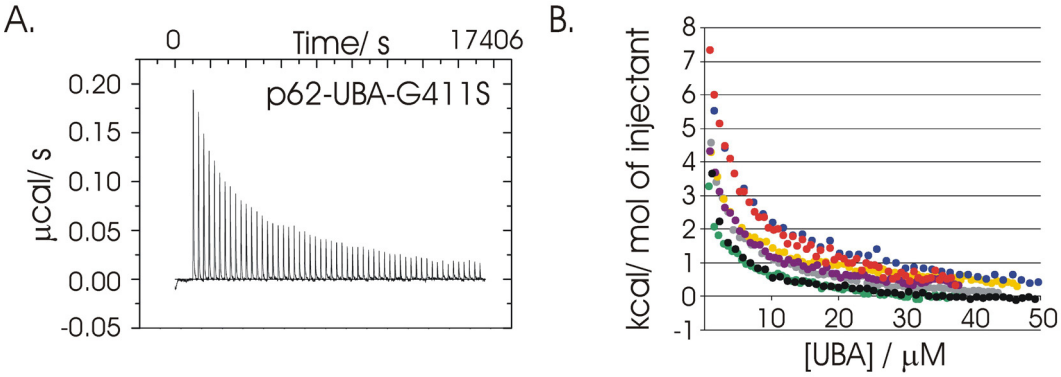



**Figure 6-26: (A) Typical ITC dissociation isotherm for the p62-UBA PDB mutant illustrated using G411S. (B) An overlay of calorimetric heat responses for wt p62-UBA (Green), p62-341-440 (Grey) and all studied PDB mutants including P387L (Yellow), P392L (Purple), S399P (Blue), G411S (Red) and G425R (Black).**

**Table 6-5: Thermodynamic parameters at 298 K for the dimer dissociation of the p62-UBA and associated PDB mutations determined by ITC.**

| | $K_{dim}$ (±) /μM | ΔH (±) / kJ/mol | $\Delta G_{298}$ (±) kJ/mol | ΔS (±) / J/mol/K |
|---|---|---|---|---|
| Wt | 7.1 (1.0) | 6.8 (3) | 29.4 (1) | -75.0 (5) |
| S399P | 17.5 (3.0) | 13 (1.5) | 27.2 (1) | -47.6 (5) |
| G411S | 10.6 (0.7) | 16.7 (2) | 27.4 (1) | -72.0 (12) |
| G425R | 3.0 (0.9) | 16 (4) | 31.5 (1) | -51.0 (6) |
| 341-440 | 7.6 (1.0) | 13.4 (4) | 29.4 (1) | -45.0 (8) |
| P392L | 9.6 (0.8) | 10.3 (2) | 28.6 (1) | -61.5 (4) |
| P387L | 9.8 (0.6) | 13.0 (1) | 28.6 (1) | -51.9 (4) |

The two PDB mutations which showed decreased thermal stability similarly show increased recorded values of $K_{dim}$, with G411S recording a marginally reduced affinity of 10 ± 0.7 μM and S399P giving a further increased $K_{dim}$ of 17 ± 3 μM. Overall there is a good correlation between the observed $T_m$ and the observed $K_{dim}$ with the two marginally destabilised PDB mutants P392L and P387L showing marginally reduced affinity with $K_{dim}$ of 9.8 ± 0.7 μM and 9.6 ± 0.8 μM respectively. On the other hand the G425R mutant which recorded an increased $T_m$ shows a corresponding increase in affinity with a $K_{dim}$ of 3 ± 0.9 μM as suggested by the low concentration dependence of the $\theta_{222}/\theta_{208}$ ratio.

The functional implications of either an increased or decreased dimer affinity are unclear, however it is proposed that dimer formation may serve to regulate Ub binding by the p62-UBA which appears to bind Ub only in its monomeric form (Discussed in Chapter 7). An increase in the dimer affinity such as that seen for the PDB mutant G425R could have knock on effects on function by modulating the populations of available monomer[61].

#### 6.6.1.3 ESI-MS and PDB Heterodimer Formation

The dimeric state of the PDB mutants S399P, G411S, and G425R was examined by native ESI-MS in the concentration range 8 to 0.5 μM. In all cases the highest order oligomer detected was a dimer as seen for the wt protein. As discussed previously, although the populations of monomer and dimer can not be accurately determined the population changes in monomer and dimer will be reflected in changes in the average charge-state of apparent monomer ions, which contain varying mixtures of monomer and dimer ions at different concentrations. The wt protein and G425R mutants show relatively large concentration dependence in average charge-state, although both S399P and G411S show concentration dependent variations in the average charge-state the magnitudes are lower than those seen for the wt protein and G425R. This may well be a reflection of the $K_{dim}$, which is higher in G411S and S399P than in wt and G425R. It would be expected therefore that the increase in dimer concentration for the wt protein and the G425R mutants would be larger, in the measured concentration range than for the G411S and S399P mutants, giving rise to reduced concentration dependence in the average charge-state.

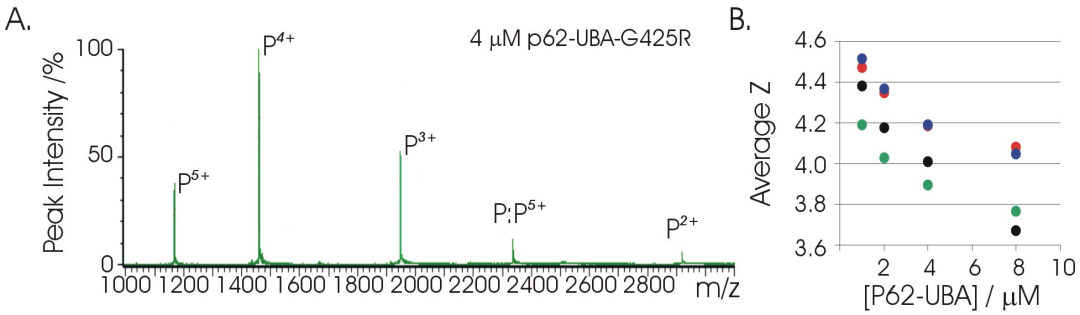



**Figure 6-27: (A) Example ESI-MS of a p62-UBA PDB mutant represented by 2 μM G425R. (B) Concentration dependence in the average charge-state of p62-UBA PDB mutants including wt (Green), S399P (Blue), G411S (Red) and G425R (Black).**

The majority of sufferers of PDB who carry mutations in the p62-UBA domain possess only a single copy of the PDB mutated gene, with the other allele being the wt gene. Therefore it is assumed that in most cases the p62-UBA will form a mix of homo- and hetero-dimers of wt p62-UBA and PDB mutant. In order to confirm the possibility of heterodimer formation of wt p62-UBA with PDB mutants *in vitro*, ESI-MS was performed on an approximately 1:1 mix of wt p62-UBA with the PDB mutants S399P, G411S, or G425R. In all cases the expected heterodimer was observed at total protein concentrations up to 4 μM.

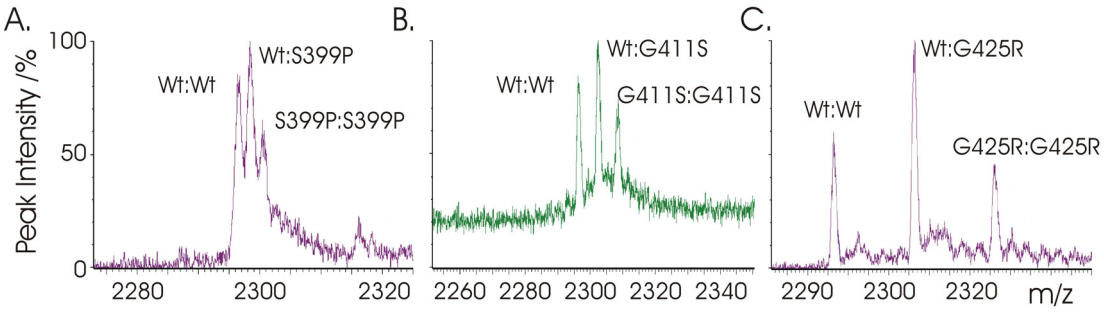



**Figure 6-28: Sections of ESI-MS showing the 5+ ion of homo- and hetero-dimers between wt p62-UBA with; (A) S399P. (B) G411S. (C)G425R.**

In all cases the heterodimer appeared at higher signal intensity that the two homo dimmers. This is consistent with the expected 1:2 ratio of homodimer to heterodimer although in many cases the data was of insufficient quality to accurately determine populations (Figure 6-28).

### 6.6.2 NMR Examination of PDB Mutants

In order to examine the possibilities of structural perturbations in the p62-UBA as a result of PDB mutations, the sensitivity of the backbone amide proton to its chemical environment was utilised. The chemical shift of the backbone amide proton and the attached $^{15}N$ is dependent on a combination of backbone dihedral angles, the side chain $\chi 1$ angle of the preceding residue, and the proximity of other side chains as a result of the tertiary structure. Structural perturbations caused by mutations would be expected to alter at least one of these properties leading to changes in the chemical shifts, this is often used to give indications of perturbations with per residue resolution.

$^{1}H$-$^{15}N$-HSQCs were recorded on 1 mM samples of $^{15}N$-p62-UBA wt and the PDB mutants S399P, G411S, and G425R. Assignments of all mutants was achieved using $^{1}H$-$^{15}N$-HSQC-NOESY, $^{1}H$-$^{15}N$-HSQC-TOCSY, $^{1}H$-$^{1}H$-NOESY and $^{1}H$-$^{1}H$-TOCSY using standard methodology (Discussed in Chapter **3**), with the previously assigned p62-UBA assignment as a guide[180], in the case of S399P and G425R data acquisition and chemical shift assignment was performed by Dr Jed Long and Dr Clare Evans.

Both the G411S and G425R mutants gave large CSPs localised to the mutated residue and the following residue due to direct contact with the mutation site (Figure 6-29). G411S shows some extra CSP in the helical turn following the mutation site and in the loop preceding the mutation site which would be in close contact with the Ser411 side chain (Figure 6-29). Overall neither G411S nor G425R appear to result in large structural perturbations to the p62-UBA as no large chemical shift perturbations are seen other than in the immediate vicinity of the mutation site.

However in the case of the PDB mutant S399P, significant CSPs (>0.05 ppm) are seen across the entire p62-UBA domain suggesting a large conformational change. Although significant perturbations are seen across the whole domain a set of contiguous large CSPs (>0.2 ppm) are seen extending from Leu394 to Ser403 encompassing the entire of helix 1 which harbours the S399P mutation (Figure 6-29).

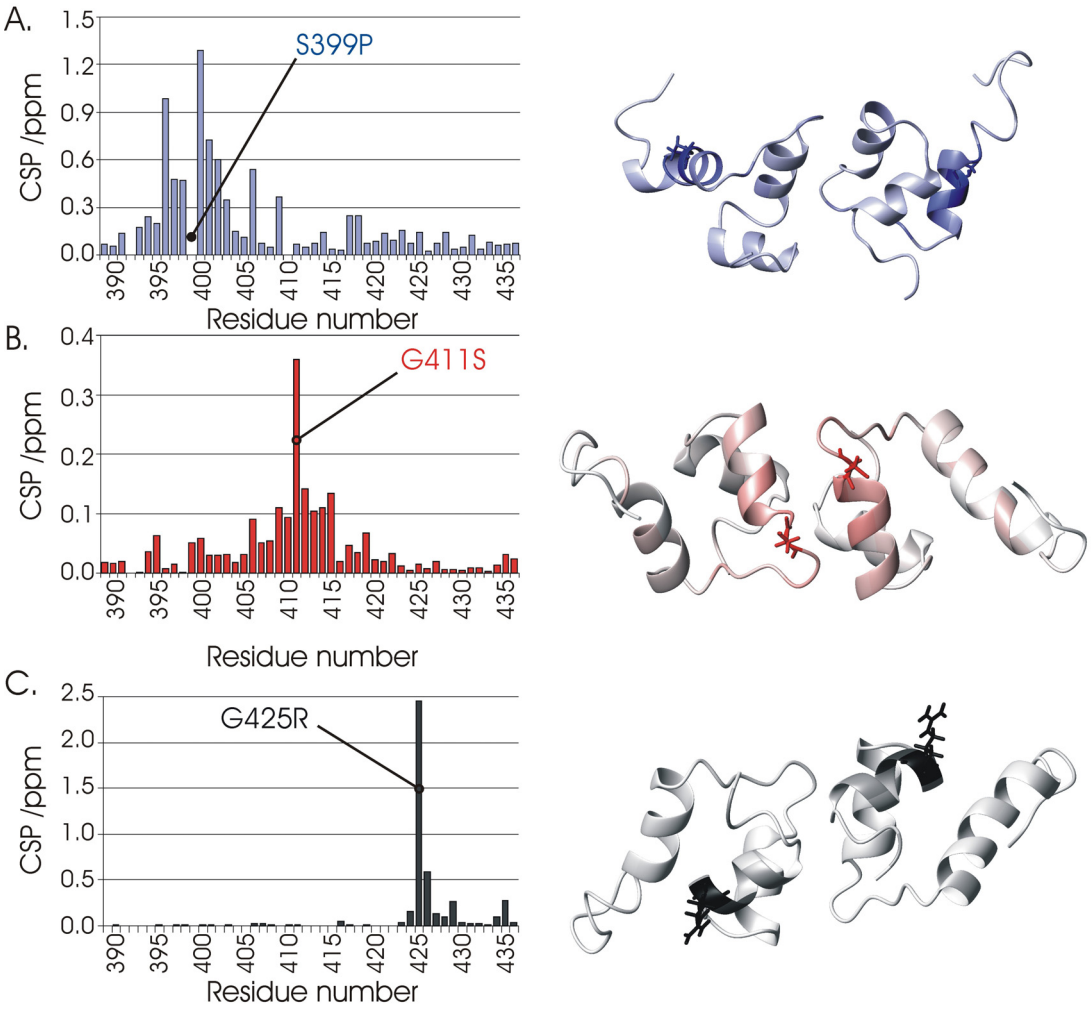



**Figure 6-29: Residue specific CSP caused by the PDB mutations (A) S399P ,(B) G411S and (C)G425R. Large CSPs are mapped onto a ribbon diagram of the p62-UBA dimer (RPDB ID: 2KNV).**

The large CSP along the entirety of helix 1 for the S399P mutation may be explained by the nature of the mutation. Ser399 is positioned in the centre of helix 1 and its side chain projects out into solution. Replacing the Ser side chain with a helix-breaking Pro may result in breaking of helix 1 with a loss of structure N-terminal to the inserted proline. This arises due to two key points, the first being the replacing of backbone NH in serine with the side chain carbon atom of Pro, this results in both a breaking of the hydrogen bond network of the helix, and provides a steric block preventing the correct helical turn requiring the helix to bend or break. Regardless of the precise structural implications, the insertion of a proline into helix 1 would give rise to significant destabilisation which is consistent with the observed $T_m$ which is considerably reduced in S399P.

Changes in conformation towards the N-terminal half of helix 1 is supported by significant chemical shift changes for Leu417 and Gln418 in helix 2 which have side chains which pack against the N-terminal portion of helix 1 (Figure 6-29). The amide proton and $^{15}N$ chemical shifts of helix 1 of S399P are inconsistent with the loss of structure showing little or no loss of CSD (Figure 6-30). This is consistent with the observed NOE patterns which show maintenance of the characteristic helical NH-NH correlations which remain strong in residues adjacent to the mutation site, consistent with stable helical structure. The only expected NOE that is not visible in the NOESY spectrum is the $NH_i$ to $NH_{i\pm3}$ correlations between residue in the range Ser397 - Met401 suggesting an increase in the expected distance between the helical turn following and preceding the mutation site (Figure 6-30). This is somewhat contradicted by the Hα chemical shifts which show markedly reduced CSD for the helical turn preceding the mutation site consistent with helix breaking or helical instability caused by Pro399 (Figure 6-30).

The maintenance of CSD for the backbone NH and $^{15}N$ chemical shifts and the NOE patterns of helix 1 indicate that helix 1 remains reasonably well folded although the high CSPs seen in helix 1, resulting from the S399P mutation suggests that this structure may be significantly different to that seen in the wt protein. The reduced Hα CSD for the loop preceding the mutation site suggests a reduction in helical character which is consistent with the observed reduction in the intensity of the far-UV CD spectrum of S399P. The best explanation for these characteristics would be a disruption of the helical turn containing the Pro399 side chain resulting in a potential bend or destabilisation of the N-terminal portion of helix 1. This is consistent with the observed pattern of CSD, the loss of $NH_I$ to $NH_{I\pm3}$ NOE correlations between residues Ser397 - Met401, the observed CSPs at the C-terminus of helix 2, and the reduction in $T_m$ observed by equilibrium unfolding experiments. Both a bend in helix 1 and a general loss of structure is consistent with the observed data and this would need to be verified by full structural characterisation.

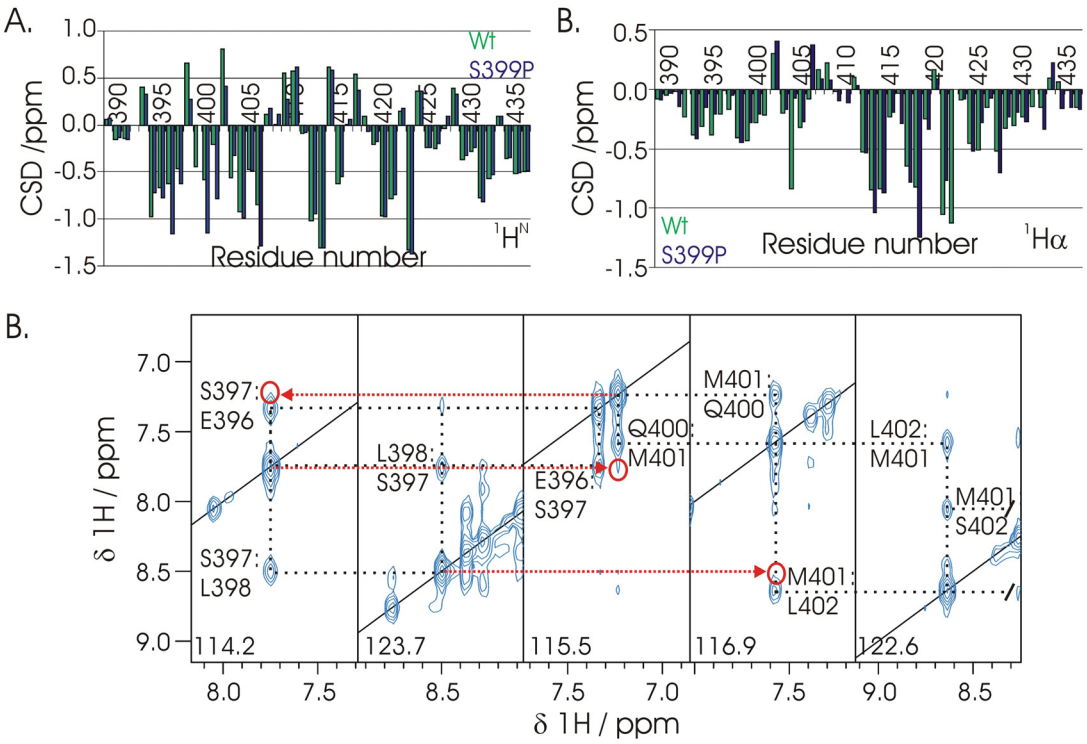



**Figure 6-30: CSD for wt p62-UBA (Green) and S399P (Blue) including; (A) $^1H^N$ and (B) $^1H\alpha$ CSD. (C) Slices through the $^1H$-$^{15}N$-HSQC-NOESY of S399P showing strong helical NOE, the missing $NH_I$ to $NH_{I\pm3}$ correlations are indicated in red.**

While these large structural perturbations may suggest a large scale conformational change which may well disrupt the ability of the p62-UBA to perform its function, the large effects do not involve residues thought to be involved in Ub recognition suggesting that the Ub binding site remains intact. Therefore while the combination of these structural changes and the subsequent lack of stability may contribute to the S399P disease mechanism the details of that mechanism remain elusive.

## 6.7 Conclusions

All previous studies of the p62 UBA both structural and biochemical, have worked on the assumption that the p62-UBA was monomeric in solution. Using a combination of NMR and complementary biophysical approaches the p62-UBA has been shown to form a symmetrical dimer in solution with a $K_{dim}$ of around 7 μM. The dimer interface is largely hydrophobic in nature resulting in an interaction which appears to be entropically driven.

The dimerisation interface of the p62-UBA consists primarily of helix two and the C-terminal portion of helix 3 as defined by both the determined structure and by the observed dimerisation induced CSPs. To date there are three other confirmed UBA homodimers in the RPDB consisting of the UBA domains of the protein doublesex from Drosophila melanogaster[232,233], the UBA domain of the Cbl-b Ub ligase and the UBA domain of the c-Cbl Ub ligase[234-236]. Despite the later two being closely related, all three form distinctly different dimer interfaces and all are distinct from the dimer interface of the p62-UBA (Figure 6-31). Although the dimer interfaces of the four current UBA dimers are significantly different, and many other UBAs are known to be monomeric in solution, dimerisation may be a feature of many UBA domains which have simply yet to be identified as dimeric.

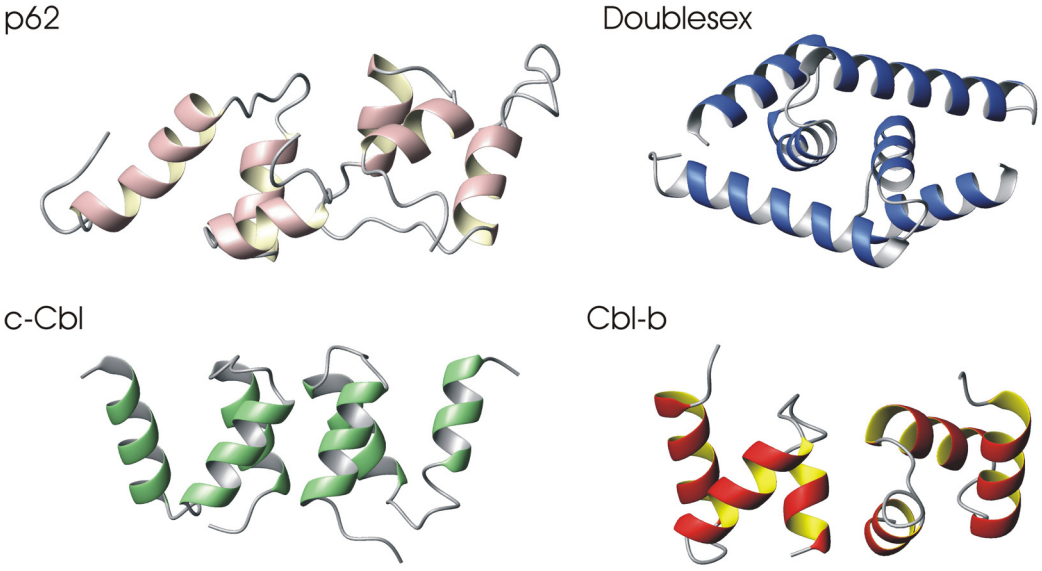



**Figure 6-31: Comparison of the 4 known homo dimeric UBA domain including the p62-UBA (Pale red), doublesex (Blue), c-Cbl (Green) and Cbl-b (Red) (RPDB ID: 2KNV, 1ZV1, 2OO9 and 2OOA).**

Of the three previously identified UBA dimers only that of Cbl-B possess fully functional Ub binding, the UBA dimer of double sex shows weak recognition of Ub while despite high sequence similarity to Cbl-b, c-Cbl has lost its ability to bind to Ub due to disruption of its hydrophobic patch by a mutation of the key hydrophobic residue at position 12 to a Glu[234]. Classically UBA domains mediate interactions with Ub utilising a hydrophobic patch centred on the so called 'MGF' motif. The examples presented here suggest that UBA function may have adapted under evolutionary pressure to form dimers, it is therefore possible that under the

same evolutionary pressures the Ub binding ability of a dimeric UBA may no longer be the required function of the UBA. For example in c-Cbl, dimerisation facilitates its phosphorylation by kinases[235] while in double sex the UBA dimerisation mediates transcription factor dimerisation to facilitate DNA recognition[232].

Regardless of the function of the p62-UBA dimer, the p62-UBA dimerisation equilibrium provides another feature of the p62-UBA which may be affected by mutation, in particular by PDB mutations, in order to regulate function. Of the PDB mutants investigated here only marginal changes in dimer affinity are observed with the PDB mutants S399P and G411S showing mildly reduced dimer affinities and the PDB mutant G425R showing mildly stronger dimerisation affinity.

The effects of the PDB mutants S399P, G411S, and G425R on the structure and properties of the p62-UBA have been examined. Both the G411S and G425R mutations appear to have little effect on structure, with chemical shift changes only seen in the immediate vicinity of the mutation site. The mutant S399P appears to have larger implications for structure causing a destabilisation of helix 1. This structural perturbation results in reduction in the thermal stability of both the dimer and monomer species by around 25 K which may contribute to the mechanism by which S399P contributes to PDB by destabilising the p62 protein[208]. The G411S similarly destabilises the domain but to a lesser extent, however the destabilisation of these PDB mutants may contribute to PDB in both case.

The formation of p62-UBA dimers does however raise a key point, the majority of patients with PDB carry only one mutant allele and one wt, assuming equal protein expression of the two PDB mutants, the p62-UBA must therefore exist in three forms, the homodimer of each mutant and double the population of the heterodimer. Any examination of the effects of PDB mutants on the p62-UBA must therefore consider the effects of those mutants in the form of both homo- and hetero-dimers. Results presented here using ESI-MS have confirmed the formation of heterodimers but the structural implications are still unknown.

# Chapter 7

# 7 Characterisation of Ubiquitin Binding by the p62-UBA Domain: Effects of Paget's Disease Mutations

## 7.1 Introduction

### 7.1.1 The UBA and Ubiquitin Recognition

As its name suggests, the UBA domain interacts non-covalently with Ub, and was in fact the first identified UBD[18]. Now almost 15 years after the initial identification of the UBA domain as a protein common to the ubiquitination system, the structure and function of a wide variety of UBDs have been characterised[16,17]. The majority of UBA domain (although not all) form non covalent complexes with Ub. However the original view that UBA domains, as a family, are primarily associated with the UPS has since been shown not to be the case with UBA domains associated with wide ranging processes including kinase regulation and autophagy[16,17].

The interaction between UBA domains and Ub has been characterised for many cases and a number of complexes between UBA domain and mono or poly-Ub are available in the RPDB. Although the orientation between the UBA domain and Ub varies between different complexes the overall interaction interface is consistent. The key interaction centres on the hydrophobic β-sheet surface of Ub including Ile44 and Val70, interacting with a hydrophobic patch of the UBA involving the so called 'MGF' loop (Figure 7-1)[17]. The affinities of UBA domains for Ub vary largely within UBA domains as it does across UBDs with mono-Ub affinities reported ranging from 0.7 mM to 15 μM[17,148].

One of the most interesting areas of Ub mediated signalling is the study of poly-Ub chain linkage specificities. A number of UBA domains have demonstrated specificity for Lys48 linked poly-Ub. In the case of both the UBA2 of hHR23A and the UBA of Mud1, Lys48 chain linkage specificity is achieved by the simultaneous contact of two adjacent Ub molecules (within the same poly-Ub chain) with one

UBA molecule, where the Ub units essential adopt a closed conformation with the UBA sandwiched in-between the two (Figure 7-1)[22,23].

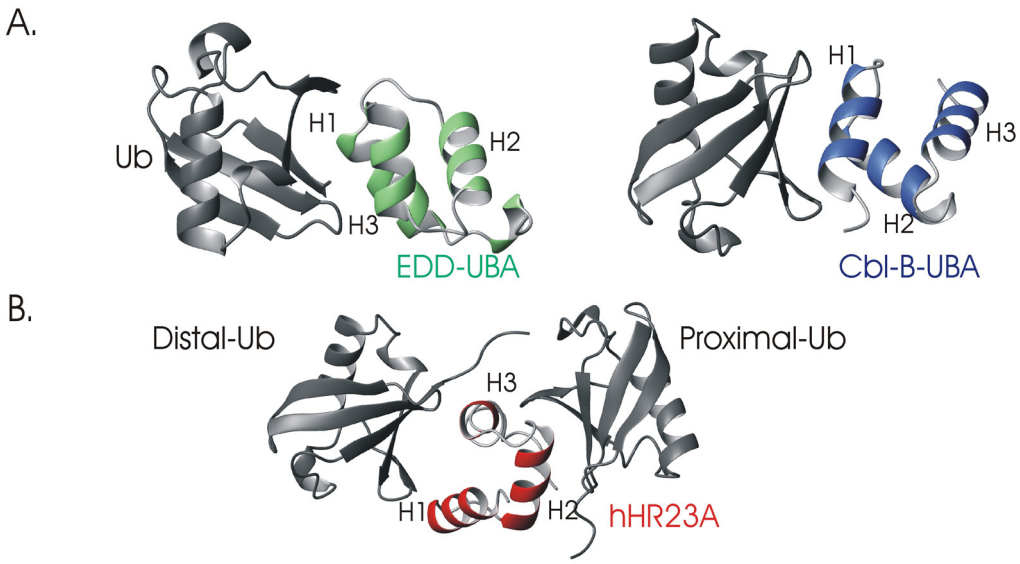



**Figure 7-1: Complexes of UBA domains with mono-Ub (Grey) including; (A) UBA of EDD (Green) and CBL-B (Blue). (B) The UBA2 of hHR23A (Red) and Lys48 linked di-Ub (Grey) (RPDB ID: 1QHO, 2OOB and 1ZO6).**

Selectivity for other chain linkages including Lys63 have been proposed for UBA domains, but no complexes detailing the structural origins of the selectivity have been presented. The observation of specificity for Lys63 for some UBA domains has been recently called into question as an artefact of the GST-pull down experiment used to examine the affinity. GST forms a strong dimer resulting in artificial pairing of their associated UBA domains in GST-pull down assays, this positioning of UBA domains leads to chelate/avidity effects when binding to the Lys63 poly-Ub chains which are known to form an extended conformation in solution[25,237]. Although this suggests that the observed Lys63 linkage specificity is artifactual, in some cases it also provides a mechanistic explanation for potential Lys63 specificity in proteins which naturally oligomerise and contain UBA domains (or other UBD).

The p62-UBA has demonstrated no specificity for any specific chain linkage in pull down assays and has been implicated in signalling pathways involving both Lys48 and Lys63 linked chains. Previous reports have mapped the p62 interaction with Ub to the classical UBA binding, Ile44 centred binding patch on Ub with an affinity of 0.7 mM, one of the weakest UBA interactions reported[61,148]. Despite the

apparent weak nature of the interaction, the UBA domain of p62 appears an absolute requirement for many p62 functions including the regulation of TRAF-6, the formation of cytosolic speckles, and receptor trafficking of TrkA.

### 7.1.2 Ubiquitin Recognition and Paget's Disease of Bone

The p62 protein is mutated in around 30 % of cases of familial PDB and 10 % of sporadic PDB cases. The overwhelming majority of p62 mutations are located within the UBA domain, demonstrating a clear functional role of the p62-UBA in PDB. Disruption of RANK-L mediated NF-κB activation using mice knock out of the genes for Ajuba, p62, TRAF6, RANK, or RANK-L leads to bone abnormalities reminiscent of PDB. Although the precise mechanisms by which p62 modulates NF-κB activity have yet to be fully elucidated, a number of the potential p62 mediated processes are dependent on a functional UBA interaction with Ub, including the key interaction and activation of TRAF6[190]. This has led to speculation that the most likely mechanism by which PDB mutations disrupt p62 function, and contribute to PDB, is by modulating Ub recognition. A number of biochemical studies have investigated the hypothesis that PDB mutants disrupt Ub recognition by the p62-UBA with some success. A number of identified PDB mutants including M404V, M404T, and G425R appear to ablate binding to poly-Ub in pull down assays, leading to the proposal that loss of Ub recognition is a unifying mechanism by which PDB mutants contribute to PDB[202,208,228].

### 7.1.3 Summary and Aims

The interaction of Ub with UBA domains was the first identified non-covalent complex with Ub and underpins many key interactions in biology. In the case of the p62-UBA domain, despite forming an apparently weak complex with Ub, it is necessary for a number of key interactions of p62 suggesting that the proposed weak interaction may be misleading. All previous studies of the p62-UBA and its interaction with Ub were performed and analysed under the assumption that the p62-UBA was monomeric in solution, the reality that the p62-UBA forms a dimer in solution may have considerable effects on the interpretation of previous studies. Many of the previous binding studies were recorded at relatively high concentrations in the range of high μM to mM where the dimer would predominate. Considering that like the interaction of the p62-UBA with Ub, the dimerisation interface is

comprised of a large hydrophobic surface, the potential for overlapping binding sites is high. The binding of Ub to the p62-UBA may therefore be heavily dependent on the interplay between dimerisation and Ub binding.

In order to gain insight into the important interaction of Ub and p62, the interaction of the p62-UBA with Ub has been investigated using NMR and complementary techniques. The proposed effects of PDB mutants on Ub recognition has been analysed quantitatively for mono-Ub using the PDB mutants S399P, G411S, and G425R in order to investigate their effects on the stability of the interaction and to understand the molecular origins of any observed changes.

## 7.2 p62-UBA Dimerisation and Ub Binding Compete

Previous studies of the recognition of Ub by the p62-UBA domain observed that the Ub binding conformation of the p62-UBA is not the conformation that is observed at high concentration by NMR[36,180]. The observed species by NMR was noted to be in equilibrium with the Ub binding species indicated by the observation that the Ub binding equilibrium in titrations involving $^{15}$N-p62-UBA is in slow exchange with secondary fast exchange shifts observed in the Ub binding species[36,180]. The recent discovery that the conformation of the p62-UBA observed at high concentration in the NMR is a dimeric species led to the proposition that the previously identified Ub binding species was that of the monomeric species[61].

### 7.2.1 Low Concentration NMR Titrations

In order to determine whether the Ub binding species of the p62-UBA was the monomeric species or the dimeric species, Ub was titrated into $^{15}$N-p62-UBA at low concentration (10 μM) by Dr M. J. Pandya, Prof M. P. Williamson and Dr C J. Craven at the University of Sheffield. $^{1}$H-$^{15}$N-HSQC of p62_UBA at 10 μM contains peaks corresponding to both the monomeric and dimeric species at approximately equal intensities. Addition of Ub therefore allows identification of the preferred Ub binding conformation.

Titrations of unlabeled-Ub into $^{15}$N-p62-UBA under these conditions results in concentration dependent chemical shift changes in the monomeric species only, consistent with a binding event which is in fast exchange (Figure 7-2). The slow exchange event seen in high concentration titrations is not seen for the monomeric species, but is represented in this case by a decrease in the observed intensities of the

dimeric species with a corresponding increase in the intensity of the monomer coupled to a fast exchange shift. At 200 µM Ub the dimeric species is no longer visible, with all signal intensity shifted to the ubiquitin bound monomer (Figure 7-2). This is consistent with a monomer-dimer equilibrium in which Ub binds only to the monomeric UBA domain.

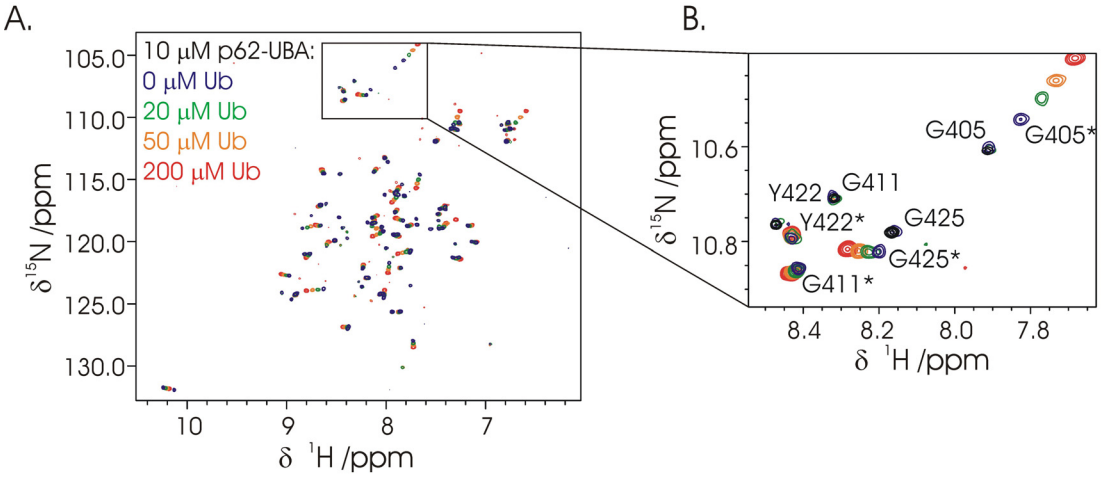



**Figure 7-2: (A) $^{1}$H-$^{15}$N-HSQC of $^{15}$N-p62-UBA at 10 µM in the presence of 0 (Blue), 20 (Green), 50 (Orange) and 200 µM (Red) (B). A portion has been enlarged and labelled for clarity, the monomeric species is labelled with a "*".**

The large CSP for the monomeric p62-UBA species upon addition of Ub indicates that the monomer is the Ub binding conformation. The lack of any CSP for the dimeric species of the p62-UBA suggests that the dimeric species is unable to bind Ub rather than just preferred Ub binding to the monomer. This suggests that the dimerisation of the p62-UBA and Ub recognition by the p62-UBA are mutually exclusive events[61].

The use of low concentration titrations allows isolation of the CSPs associated with dimerisation and Ub binding. The CSPs associated with binding of Ub to the p62-UBA are generally smaller in magnitude than those associated with dimerisation but affect overlapping sets of resonances, most notably around helix 3. Plotting the binding induced CSPs onto the structure of the p62-UBA dimer clearly predicts a large overlap in binding patches, where many of the binding induced CSPs are buried at the dimer interface (Figure 6-4). This is perhaps not surprising considering both the p62-UBA dimerisation and the p62-UBA interaction with Ub are proposed to utilise a hydrophobic surface on the p62-UBA[36,61,180].

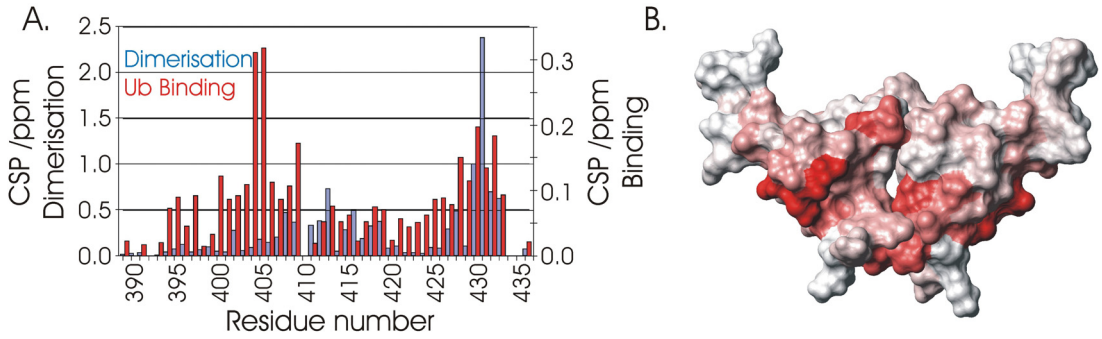



**Figure 7-3: (A) Residue specific CSP for the p62-UBA dimerisation and binding to Ub. (B) Large CSP for binding to Ub, plotted onto the p62-UBA dimer structure in red (RPDB ID: 2KNV).**

### 7.2.2 ESI-MS Titrations

In order to confirm the stoichiometry of the interaction of the p62-UBA with Ub, native ESI-MS was performed. ESI-MS were recorded on 1 µM p62-UBA mixed with 0.2 – 3 µM Ub. In all cases mass ion peaks were observed corresponding to monomeric and dimeric p62-UBA, monomeric Ub, and a 1 to 1 complex of p62-UBA with Ub (Figure 7-4). No 2-to-1 or 2-to-2 complexes of any kind were seen in the concentration range examined by ESI-MS precluding Ub binding to dimeric p62-UBA. In this case the concentration of p62-UBA was maintained at 1 µM to minimise dimer formation. However a repeat of the experiment at higher concentrations (3 µM p62-UBA) saw no higher order complexes of Ub and p62-UBA consistent with the NMR observations.

The relative peak intensities observed in the ESI-MS spectrum can be assumed to vary linearly with the solution populations allowing determination of the concentrations of the various species observed in the mass spectrum[45]. While the observed mass peaks easily identify the 1 to 1 stoichiometry the calculated concentrations of each species can be used to calculate binding affinity. $K_d$ can be calculated from the gradient of a plots of [P:L]/[P] against [L] where [P] represents the concentration of free protein (in this case p62-UBA), L represents the concentration of free ligand (in this case Ub) and [P:L] represents the concentration of their one to one complex. In the case of the p62-UBA the overlapping monomer and dimer peaks (Discussed in section 6.2.2) make accurate determination of the concentration of free p62-UBA difficult leading to errors in calculation of the various species. However using mass peaks corresponding to Ub and the p62-UBA-Ub 1 to 1

complex, the concentration of these two species was determined and used to estimate the concentration of free p62-UBA. Under these conditions a $K_d$ for the p62-UBA and Ub complex of 22 ± 15 µM was estimated (Figure 7-4). The large errors associated with the determined $K_d$ are the product of inaccuracies in the calculation of free p62-UBA and relatively poor signal to noise for the 1 to 1 complex for some ions.

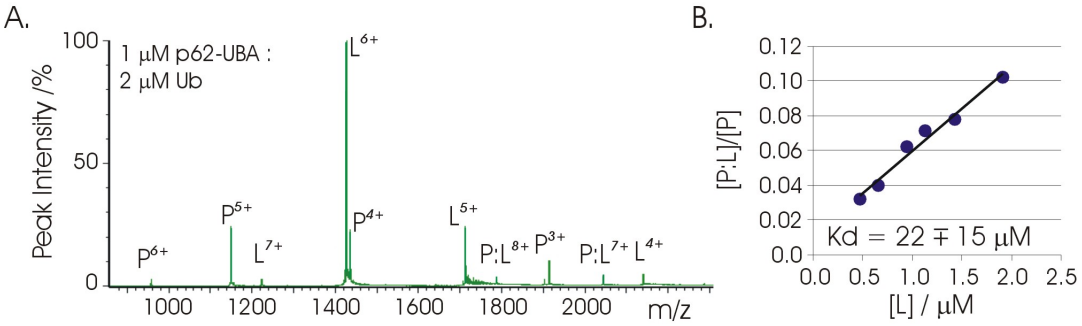



**Figure 7-4: (A) ESI-MS of 1 µM p62-UBA mixed with 2 µM mono-Ub. (B) Plots of [P:L]/[P] Vs. [L] derived from the ESI-MS titration of p62-UBA and Ub.**

### 7.2.3 Summary

Using NMR and Biophysical approaches it has been demonstrated that dimerisation and Ub binding by the p62-UBA are competitive processes. The affinity of the p62-UBA monomer with Ub, suggested by ESI-MS, is relatively high for a UBA domain. This high affinity has been masked in previous studies by competition with the dimeric form of the p62-UBA, and this will be discussed in the next section.

While other UBA domains have been reported to form dimmers, this represents the first example of a dimer which prevents the interaction with Ub providing a novel mechanism for control of Ub recognition by a UBA domain[61,233,235].

## 7.3 Identifying the p62-UBA Binding Surface of Ub

As stated previously, several of the potential mechanisms by which PDB mutants may disrupt the function of p62 focus on disruption of Ub binding. In order to better understand the interaction with Ub and provide a basis to examine the possibilities of disruption of the binding interface by PDB mutations, titrations were performed in which unlabeled p62-UBA was titrated into $^{15}$N-Ub. The addition of up to 6 mM p62-UBA to 1 mM $^{15}$N-Ub resulted in concentration dependent chemical

shift changes for many of the Ub resonances, consistent with a binding event which is in fast exchange (Figure 7-5). The fast exchange seen for the majority of resonances allowed tracking of the bound shifts directly back to the previously assigned free Ub facilitating the assignment of the fully bound state. All resonances, with the exception of three, were in fast exchange, however Arg42, Lys48, and Leu71 all shift and broaden as p62-UBA is titrated into Ub, with complete signal loss by 0.2 mM p62-UBA consistent with intermediate exchange. All three peaks had undergone considerable CSP by 0.2 mM before broadening out suggesting that the fully saturated CSP would have been considerable.

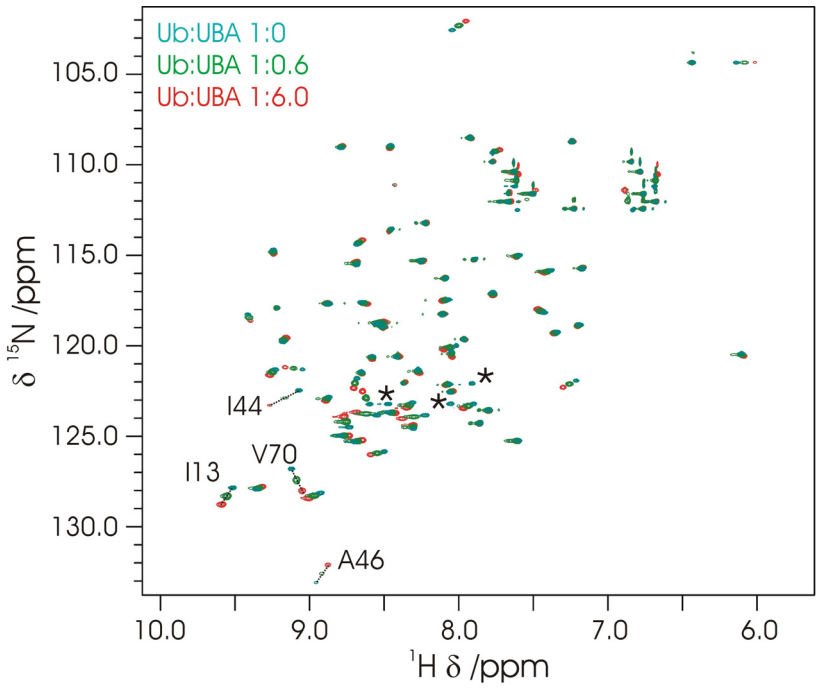



**Figure 7-5: $^{1}$H-$^{15}$N-HSQC of $^{15}$N-Ub in presence of 0 (Blue), 0.6 (Green) and 6 mM (Red) p62-UBA. Residues which broaden during the titration are marked with a '*'.**

The large CSP may explain the peak broadening in these cases. The exchange regime is related to the ratio of the rate constant of the exchange process to the chemical shift separation in Hz. The intermediate exchange regime of these three resonances may suggest either a reduced rate constant for these residues, or a particularly large chemical shift separation, given the large CSPs predicted for these three resonances the latter would be expected to be the cause[145]. However for typical intermediate exchange the resonances for these three residues would be expected to sharpen again at saturation returning measurable signal intensity, however this is not

the case and in fact these three resonances do not reappear in the measured concentration range.

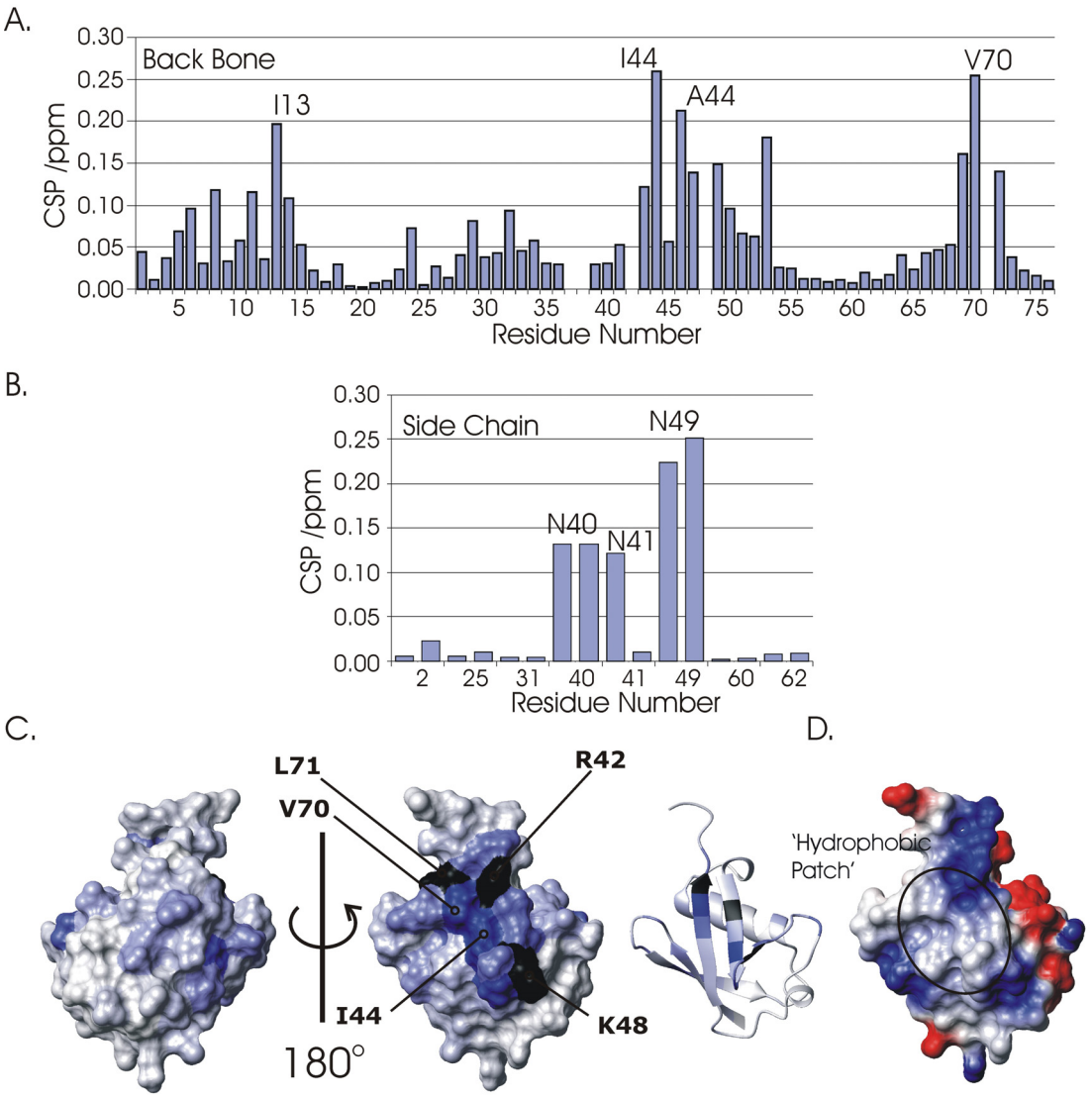



**Figure 7-6: Residue specific CSPs for the addition of p62-UBA to $^{15}$Ub for; (A) Backbone NH (B) and side chain NH/NH$_2$. (C) Large CSPs plotted onto the structure of Ub in blue (RPDB ID: 1UBQ) Broadened residues are coloured black. (D) The electrostatic surface potential of Ub showing the hydrophobic patch.**

The measured CSPs for $^{15}$N-Ub affect a large portion of Ub, with significant perturbations seen for the backbone NH of Leu8, Lys11, Ile13, Leu43, Ile44, Ala46, Gly47, Asn49, Gly53, Leu69, Val70, and Arg72 and for the side chain NH$_2$ of

Asn40, Asn41, and Asn49. When combined with the three residues showing significant line broadening of Arg42, Lys48, and Leu71 they cluster to form a relatively localised region on the β-sheet surface of Ub, with the hydrophobic Ile44 and Val70 pair at its centre (Figure 7-6). This is consistent with other UBA domains binding to the canonical Ile44 binding patch[17,22,36,180].

The observed chemical shift at a given concentration of binding partner for a system in fast exchange is a population weighted average between the chemical shifts of free and bound conformations. The concentration dependence of the chemical shifts can therefore be used to generate equilibrium saturation binding curves by plotting chemical shift or CSP, as a function of partner concentration. The measured saturation binding curve can be fitted to the appropriate binding model in order to determine equilibrium constants for the measured interaction. In the case of the p62-UBA previous studies had fitted the fast exchange shifts in Ub titrations to a simple 1 to 1 binding model and obtained a $K_d$ of around 0.5-0.7 mM[36,180]. However the dimerisation of the p62-UBA, which is now known to compete with Ub binding by the p62-UBA, complicates this in that the observed $K_d$ of 0.5-0.7 mM is actually the product of two competing equilibria and is better described as an apparent $K_d$.

The apparent $K_d$ was determined by fitting the observed concentration dependent CSP to a simple 1 to 1 binding model. All residues showing significant CSP (>0.05) and giving a clean binding curve (a total of 31 curves) were fit individually using the program IGOR Pro™ (V 5.0.5.7 WaveMetrics) to obtain an apparent $K_d$, the full set of curves were then globally fitted to a single $K_d$ using the program IGOR Pro™ (V 5.0.5.7 WaveMetrics) yielding an apparent $K_d$ of 741 ± 80 μM (Figure 7-7).

However this only reflects the observed $K_d$ at the concentrations used in the NMR experiment and does not give an accurate measurement of the $K_d$ of the p62-UBA for Ub. The program DynaFit is designed for fitting complex equilibria, and the concentration dependence of the observed CSPs was successfully fitted using DynaFit (V 3.28.070, Biokin)[59,60] to a competitive equilibrium model. Attempting to define both the $K_{dim}$ and $K_d$ on the basis of the binding curve alone yielded poor results. However fixing of the $K_{dim}$ value to that calculated using ITC improved the quality of the fit, yielding a $K_d$ for the monomeric p62-UBA interaction with Ub of 42 ± 3 μM (Figure 7-7), in reasonably good agreement with the value from ESI-MS.

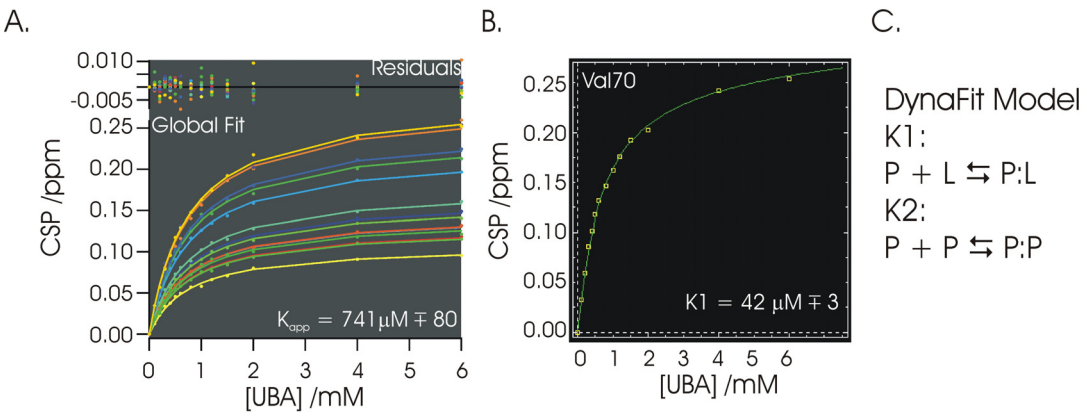



**Figure 7-7: Saturation binding curves for the p62-UBA; (A) Fit globally to a simple 1 to 1 binding model using IGORPro™. (B) Output from DynaFit[59,60] for Val70 fitting the observed CSP to a competitive binding model. (C) The binding model used by DynaFit.**

Extending the p62-UBA at the N-terminus (341-44) showed little evidence for effects on the p62-UBA dimer stability and no changes were evident in the $^{1}$H-$^{15}$N-HSQC. The binding induced CSPs for the monomeric p62-UBA are small around the N-terminal region of the domain and no region close in space to the N-terminus shows significant binding induced CSPs. It would therefore be assumed that extending the region N-terminally would not alter the Ub binding ability of the p62-UBA.

In order to confirm that extending the p62-UBA domain N-terminally by ~50 residues has no significant effect on Ub binding, the titration of p62-UBA into $^{15}$N-Ub was repeated using the p62 341-440 construct under identical conditions and concentrations. The 341-440 construct generally results in larger CSPs for Ub than those seen for the p62-UBA, in addition to this, where in the p62-UBA significant line broadening was seen only for residues Arg42, Lys48, and Leu71, the titration with the p62 341-440 results in additional line broadening for residues Ile13, Gly47, and Val70 consistent with a dramatic increase in the chemical shift separation between the free and bound state and/or a decrease in the rate constant for the binding equilibrium. In this case fitting of the concentration dependence of the binding induced CSPs to a simple 1 to 1 binding model, as described for the p62-UBA, yielded an apparent $K_d$ of 500 ± 60 µM, suggesting either an increase in the affinity of the UBA domain for Ub in the 341-440 construct or a decrease in dimer stability. The dimer stability for the 341-440 construct is not significantly reduced

and fitting of the observed data to a competitive binding model using the program DynaFit yielded a $K_d$ of 20 ± 4 µM for the interaction of p62 341-440 with Ub, a decrease in the observed $K_d$ of 42 ± 3 µM observed for the p62-UBA.

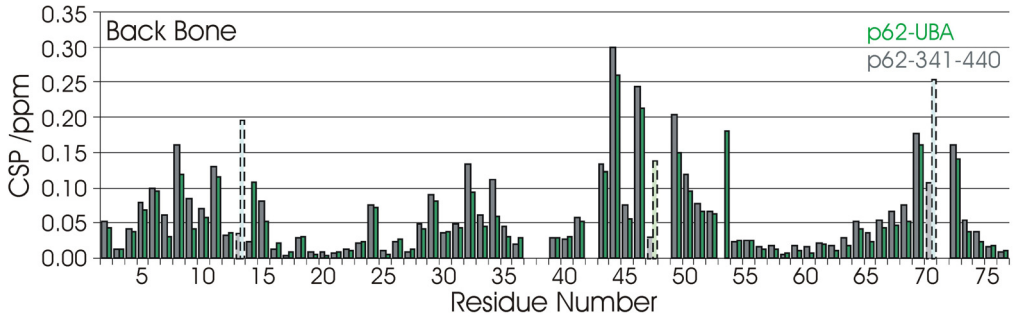



**Figure 7-8: Residue specific CSPs for binding of the p62-UBA (Green) and p62 341-440 (Grey) to $^{15}$N-Ub, residues which are measurable for the p62-UBA but exchange broadened for the 341-440 titration are faded with broken lines.**

**Table 7-1: Apparent $K_d$ and monomeric $K_d$ for the interaction of p62-UBA and p62-341-440 with Ub.**

| | p62-UBA | p62 341-440 |
|---|---|---|
| Apparent $K_d$ (±) / µM | 741 (80) | 500 (60) |
| Monomeric $K_d$ (±) /µM | 42 (3) | 20 (4) |

The apparent increase in affinity of p62 residues 341-440 compared to the p62-UBA construct is small with a difference in binding energy of around 1.8 kJ/mol.

### 7.3.1 Summary

NMR titrations using unlabeled p62-UBA titrated into $^{15}$N-Ub reveals that the p62-UBA interacts with the expected hydrophobic patch on the surface of Ub with Ile44 and Val70 at its centre. The interaction of the p62-UBA with Ub is in fast exchange on the NMR time scale[145], the resulting concentration dependence of the chemical shift can therefore be fitted to an appropriate binding model. Fitting of the observed binding data revealed a reasonably high affinity interaction between the p62-UBA and Ub. This high affinity is modulated at high concentrations by competition for the p62-UBA with homodimerisation, resulting in a much weaker apparent interaction by NMR.

## 7.4 Effects of Paget's Disease of Bone Mutations on Ub recognition

There are a number of potential mechanisms by which PDB mutants may interfere with the function of the p62-UBA, the effects of PDB mutants on structure, stability and on dimerisation have been investigated in the previous chapter. Another potential mechanism by which PDB mutants may disrupt p62-UBA function is by direct disruption of Ub recognition which is central to the majority of p62 functions[190,191,193,196,201]. In order to investigate the potential affect of PDB mutations on Ub recognition a combination of biochemical and biophysical tools have been employed to characterise Ub recognition by the PDB mutants S399P, G411S, and G425R.

### 7.4.1 Pull down assays

Previous investigations by the group of Dr Robert Layfield on the effects of Paget's disease mutations on the ability of the p62-UBA to bind to mono- and poly-Ub have centred on 'pull-down' assays, whereby mixtures containing poly-Ub of different lengths (but the same linkage type) are incubated with p62 attached to Sepharose beads (either directly or via GST), which are then washed and analysed by SDS-PAGE to analyse any bound proteins. Such 'pull-down' assays provide semi-quantitative measures of the affinity based on the amount of bound protein indicated by the intensity on an SDS-PAGE gel and visualised by Western bloting[180,202].

Pull down assays with either Lys48- or Lys63-linked poly-Ub chains with wt p62 and a selection of Paget's disease mutations including P387L, P392L, S399P, M404V, M404T, G411S, and G425R were performed by Dr J Cavey[180,202,228]. The p62-UBA shows no apparent selectivity for Lys48- over Lys63-linked chains, with strong wt p62 binding observed for di-Ub and a variety of longer poly-Ub chains. However no mono-Ub binding could be detected and the amount of poly-Ub bound varied between the examined PDB mutants (Figure 7-9). A complete attenuation of poly-Ub binding was observed for the PDB mutants M404V, M404T, and G425R while P392L, S399P, and G411S show only marginal reductions in poly-Ub binding as seen in previous studies[180,202,228].

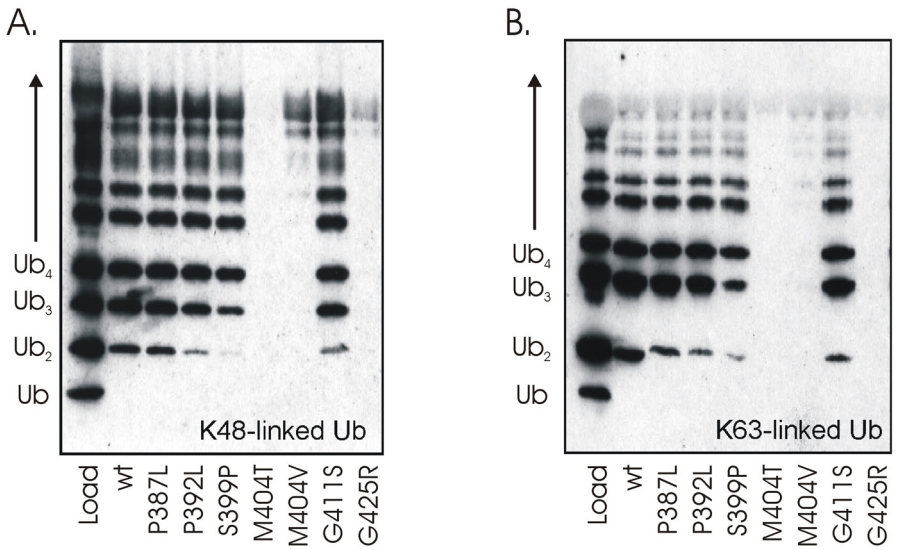



**Figure 7-9: Pull down experiments of p62 and various PDB mutants with (A) Lys-48 and (B) Lys63-linked poly-Ub chains.**

The disruption of Ub recognition by the M404V and M404T mutants is unsurprising, since Met404 forms the central Met residue to the classical 'MGF' binding motif[148,238,239]. However the molecular basis for the proposed attenuation of Ub binding is unclear for other PDB related mutations.

### 7.4.2 NMR Studies of the Effects on Mono-Ub Binding

In order to examine the effects of PDB mutations on the binding interface between Ub and the p62-UBA, and to get quantitative measurements of the effects on the affinity for mono-Ub, titrations were performed with a 1 mM sample of $^{15}$N-Ub as before. As already seen in previous chapters, PDB mutations within the p62-UBA cause significant CSPs which can be seen across large portions of the domain (Section 6.6.2). Examining potential effects on the interface using titrations involving $^{15}$N-p62-UBA would therefore be based on dissection of common CSPs for the mutation site and binding (Section 7.4.3). Titrations involving $^{15}$N-Ub on the other hand may simplify this approach; only direct changes to the Ub binding interface of the p62-UBA would be expected to alter the pattern of observed CSPs caused by the intermolecular contacts between Ub and the p62-UBA giving a potentially clearer picture of interface effects. As with both approaches, this can not distinguish between direct contacts of a mutation site with the interface, and indirect effects

caused by altering the conformation of adjacent residues which make contacts with the interface.

$^{15}$N-Ub titrations were performed using the PDB mutants, P392L, S399P, G411S, and G425R. As with the wt-UBA titration, concentration dependent chemical shift changes were observed, consistent with a binding event which is in fast exchange on the NMR time scale[145]. The magnitudes of the CSPs observed at the highest studied concentration (6 mM UBA) vary, with all PDB mutants recording generally smaller CSPs compared to wild type with the exception of G411S which gives larger CSPs (Figure 7-10). In the case of G425R the CSPs are particularly low, this is consistent with a low degree of binding to G425R, resulting in the chemical shifts remaining close to the free shifts in the measured concentration range. The line broadening seen for residues Arg42, Lys48, and Leu71 (and Ile13, Gly47 and Val70 for p62 341-440) is reduced in all PDB mutants, with a return of the signal intensity at saturation for Arg42, Lys48, and Leu71. As expected these residues record CSP values which are consistently high compared to other residues (Figure 7-10). The reduced line broadening may reflect a subtle reduction in the difference between the free and bound chemical shifts for PDB mutants. However G411S possesses consistently larger CSPs than wt suggesting this not to be the cases. It may therefore be the case that it is in fact the rate constants which are altered in these cases. If the investigated PDB mutants have reduced affinities (as suggested by pull down assays) the life time of the bound species may be expected to have decreased resulting in more ‘fast exchange’ like properties consistent with the reduction in line broadening.

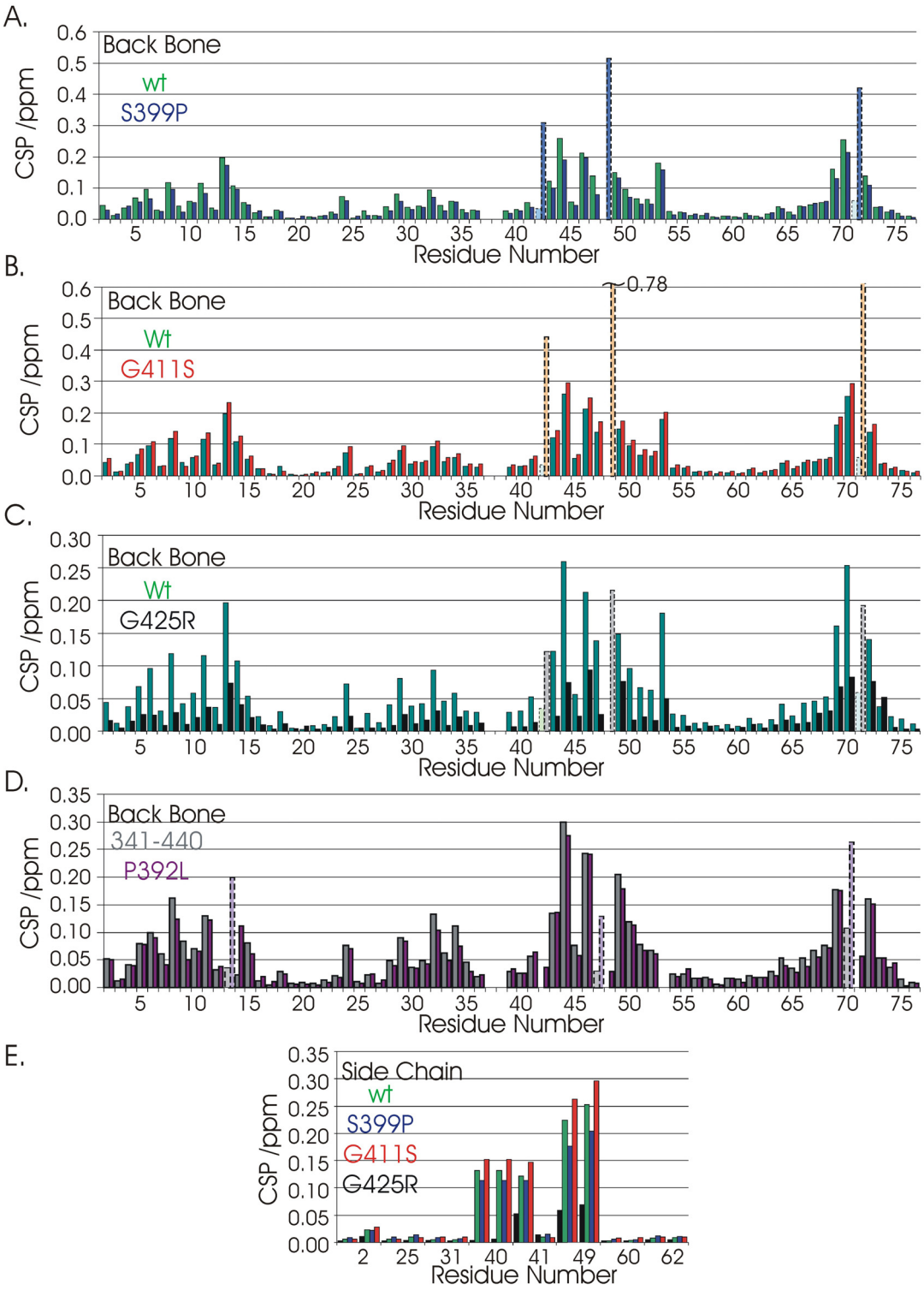



**Figure 7-10: CSPs for binding of the p62-UBA (Green) or p62-341-440 (Grey) to $^{15}$N-Ub compared with the PDB mutants P392L (Purple) S399P (Blue), G411S (Red), and G425R (Black). CSPs are shown for; (A-D) Backbone NH. (D) Side chain NH/NH$_2$. Residues which are measurable for the PDB mutants but exchange broadened for the wt titration are faded with broken lines.**

In all cases the overall pattern of CSPs remained similar to the wt-UBA suggesting that all PDB mutants bind to the same binding surface as the wt protein. However in the case of the S399P and G425R mutants, there are some noticeable differences in the chemical shift perturbations profiles. By comparing the wt-UBA titrations to PDB mutant titrations, it becomes clear that in most cases the expected 'saturation' chemical shifts for the bound state will be similar, however in a number of Ub residues with S399P and G425R this does not appear to be the case. For S399P, changes are visible for residues Leu43, Ala46, Gly47, and Gln49. For G425R changes are visible for Ub-Arg72, Gln49 and the side chain $NH_2$ of Gln40 (Figure 7-11) this is particularly obvious in the case of the side chain of Gln40 which previously gave large CSPs shows no CSP in titrations with the G425R mutant (Figure 7-10).

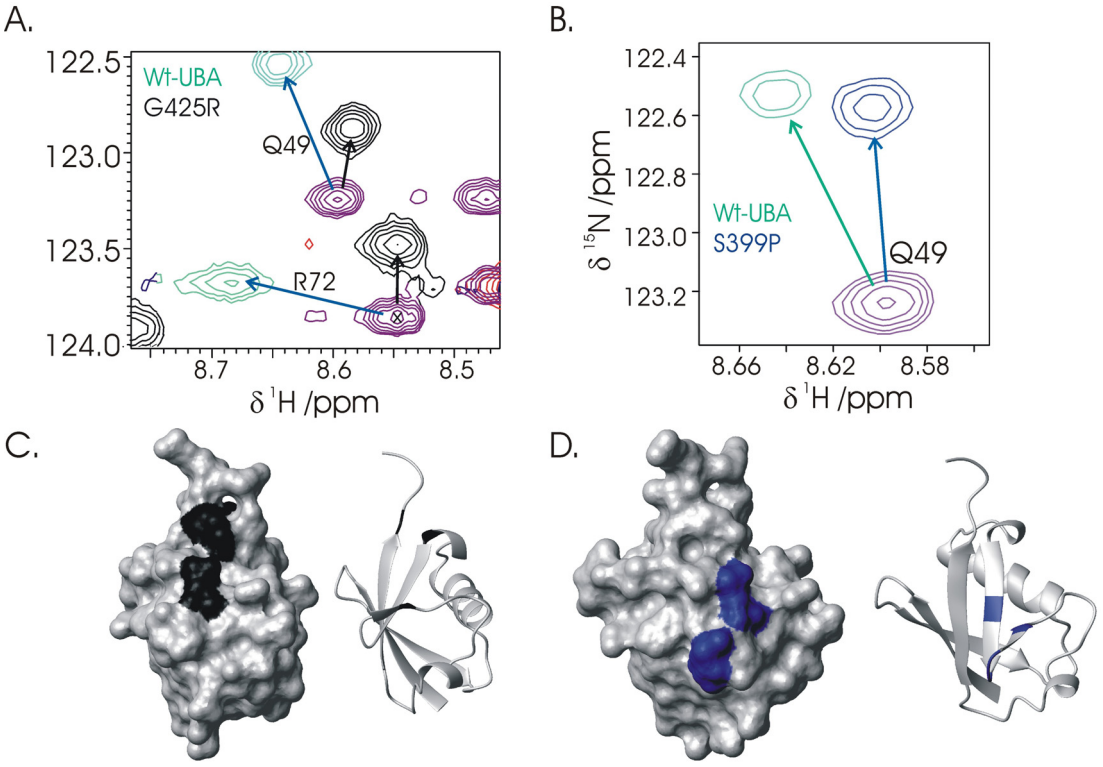



**Figure 7-11: Sections of $^{1}H$-$^{15}N$-HSQC for $^{15}N$-Ub containing; (A) Gln49 and Arg72 in the presence of 0mM (purple) and 6mM of p62-UBA-Wt (Green) and G425R (Black). (B) Gln49 in the presence of 0 mM (purple) and 6 mM of p62-UBA-Wt (Green) and S399P (Blue). (C) Effected residues for G425R (C) and S399P (D) are plotted onto the structure of Ub (RPDB ID 1UBQ).**

The subtle changes in the chemical shifts of Ub residues in titrations with S399P and G425R compared to wt p62-UBA most likely do not suggest gross changes in binding mode, but more likely mutation site induced differences in interface complementarities. If this is the case, then the position of these changes should correspond to a patch that lies close to the mutation site (or to residues affected by the mutation sites) in the complex. Mapping the location of the changes onto the structure forms a small patch in both cases, consistent with this notion (Figure 7-11). Unfortunately no model for the complex is currently available, rationalising these interface changes is therefore difficult.

As for wt-UBA titrations, the observed concentration dependence of the CSPs were fit to an apparent $K_d$ by the global fitting of multiple curves to a one to one binding model using IGOR Pro™ and to a competitive binding model utilising the ITC derived values of dimer affinities using DynaFit[59,60] (Figure 7-12). The determined affinity constants are summarised in Table 7-2.

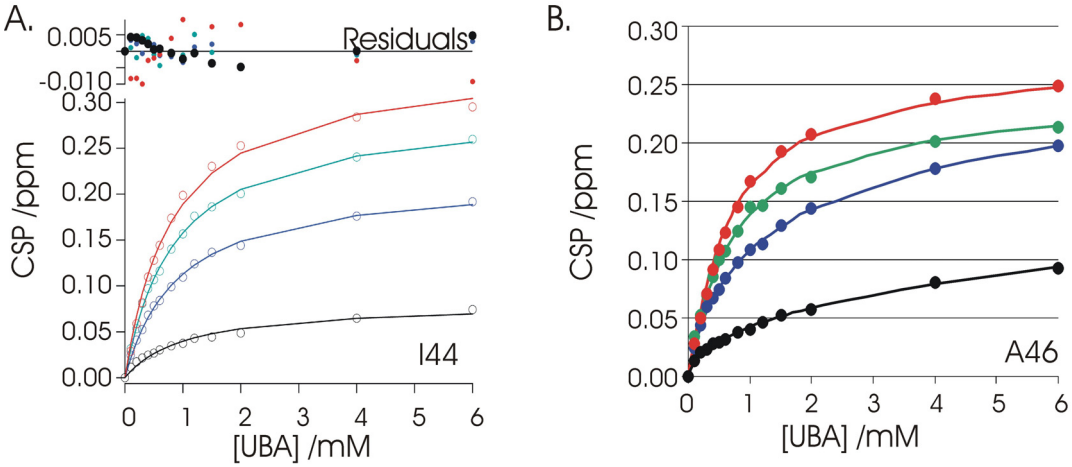



**Figure 7-12: (A) Fits of Ile44 to a one to one binding model using IGOR Pro™ (B) Fits of Ala46 to a competitive binding model taken from DynaFit. In both cases p62-UBA (Green), 341-440 (Grey), S399P (Blue), G411S (Red), G425R (Black) and P392L (Purple) are included.**

**Table 7-2: Apparent $K_d$ and determined actual $K_d$ for the interaction of p62-UBA and PDB mutants with Ub. ΔG of binding was determined from the measured affinity constants, the ΔΔG for PDB mutants were also calculated.**

| | Apparent $K_d$ (±) /μM | Actual $K_d$ (±) /μM | $\Delta G_{298\ mono}$ (±) / kJ/mol | $\Delta\Delta G_{298\ mono}$ (±) / kJ/mol |
|---|---|---|---|---|
| p62-UBA | 741 (80) | 42 (3) | -24.96 (0.15) | - |
| S399P | 991 (90) | 140 (16) | -21.98 (0.23) | 2.98 (0.28) |
| G411S | 800 (80) | 43 (4) | -24.95 (0.08) | 0.01 (0.16) |
| G425R | 1430 (200) | 340 (100) | -19.77 (0.86) | 5.19 (0.87) |
| p62-341-440 | 500 (60) | 20 (4) | -26.7 (0.3) | - |
| P392L | 658 (60) | 35 (3) | -25.43 (0.19) | 1.32 (0.24) |

All of the PDB mutants studied here show attenuation of binding to poly-Ub by pull down assays (Figure 7-9), and moderate reductions in affinity to mono-Ub by NMR titrations at high concentration when fitted to an apparent $K_d$ based on a simple one to one binding model (Table 7-2). Changes in the determined apparent $K_d$ are moderate and would represent modest changes in binding energy (< 1.5 kJ/mol). However when the observed CSPs are fit to the appropriate binding model consisting of the two competing equilibria of p62-UBA dimerisation with the formation of a p62-UBA-Ub complex, the actual $K_d$ for the interaction of the monomeric p62-UBA and its mutants can be determined. Reduction in the affinity of p62-UBA for Ub is seen for P392L, S399P, and G425R. The reduction in affinity of P392L is only moderate, with a reduction in binding energy of around 1.3 kJ/mol. For S399P this effect is greater measuring a difference in binding energy of around 3 kJ/mol. G425R shows the strongest reduction in binding affinity, with a $K_d$ of 340 ± 100 μM, an increase in $K_d$ of almost 10-fold representing a decrease in binding energy of around 5 kJ/mol. This is consistent with the trends observed in pull down assays, which show a complete loss of poly-Ub binding for G425R, with clear reductions in poly-Ub binding for S399P and moderate reductions in poly-Ub binding by P392L. The observed strong reductions in the affinity for Ub in the mutants S399P and G425R are consistent with the observed chemical shift changes in titrations with $^{15}$N-Ub, which suggest changes to the interface between the p62-UBA and Ub. The reduction in affinity seen for the P392L is within error, the P392L mutant is located at the

extreme N-terminus of the UBA domain which shows little or no binding induced CSP in titrations with $^{15}$N-p62-UBA, suggesting no involvement in Ub binding for the N-terminal region or in the following helix 1.

The G411S PDB mutation appears to show little or no effect on Ub recognition despite a reduction in the apparent $K_d$. In pull down assays the G411S mutant shows only moderate reductions in affinity for poly-Ub and it is possible that potential effects of the G411S mutation are only manifested in poly-Ub bound p62-UBA. This is consistent with the lack of clear changes in the binding interface between wt-UBA and G411S.

Although the binding curves for the interaction of $^{15}$N-Ub with the p62-UBA and its mutants appear to saturate, the fits in DynaFit suggest that the titrations may not have reached saturation. The shape of the curves is instead a product of the competing equilibria between Ub binding and UBA dimerisation. The lack of saturation is consistent with the observation that those residues showing significant line broadening remain undetected at apparent saturation. The previous observation that the line broadening is reduced in PDB mutants may well be the result of an increased rate of exchange as suggested above, however the G411S mutant has a $K_d$ which is very close to that of the wt-UBA suggesting that this may not be the case. The return of signal intensity of the line broadened residues in the G411S mutant may instead be explained by the observed reduction in dimer stability (Section 6.6.1) which would result in an increase in the concentration of available monomer at a given concentration of protein, pushing the equilibrium in the direction of the p62-UBA complex with Ub. The G411S mutant at high concentration may therefore be closer to saturation than the wt-UBA, resulting in the expected sharpening of the broadened residues, this would also explain the increase in CSP observed for G411S compared to wt titrations.

### 7.4.3 Structural Effects of the PDB Mutants

Comparison of the wt p62-UBA $^{1}$H-$^{15}$N-HSQC to those for PDB mutants of the p62-UBA in the absence of Ub, identified regions of the UBA which are affected by each of the mutations (Section 6.6.2). This was repeated for the PDB mutants in the presence of large quantities of Ub (6 mM) in order to examine the effects of Paget's disease mutants on the monomeric Ub-bound state of the p62-UBA. Ub was titrated up to 6 mM (up to 8 for G424R) into a 1 mM samples of $^{15}$N-p62-UBA-PDB

mutants including S399P, G411S, and G425R. In all cases the same slow exchange binding event was seen with CSP patterns and magnitudes between the free dimer and bound monomer which were indistinguishable from the wt protein, suggesting the same overall dissociation and binding process utilising the same or similar interfaces (Figure 7-13).

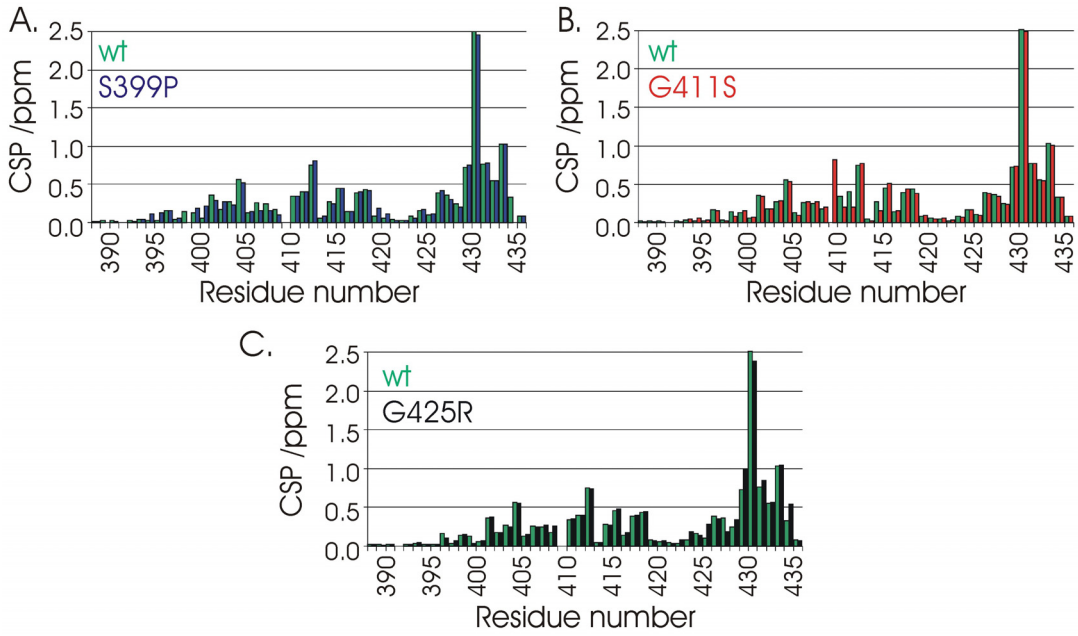



**Figure 7-13: Residue specific CSPs for coupled dissociation and binding for the p62-UBA wt (Green) compared to; (A) S399P (Blue). (B) G411S (Red). (C) G425R (Black).**

The pattern of CSPs caused by the PDB mutation (PDB mutant compared to wt-UBA) is essentially the same in the free dimeric state as in the Ub-bound monomer state. This suggests that the local contacts around the mutation site are unchanged between the monomer and dimer states. This is potentially contradictory to the idea of a gross conformational change during dimer dissociation which has previously been proposed for the p62-UBA[180].

Comparison of the Ub binding induced CSP (for wt p62-UBA) to the CSPs resulting from the incorporation of PDB mutants gives an indication of potential effects on Ub binding residues caused by the PDB mutations which could affect binding. Although binding induced CSPs are reasonably spread out over the entire UBA domain, mutation induced CSPs for G411S appear to only affect residues showing small binding induced CSPs, consistent with the $^{15}$N-Ub titrations which showed little or no changes at the interface (Figure 7-14).

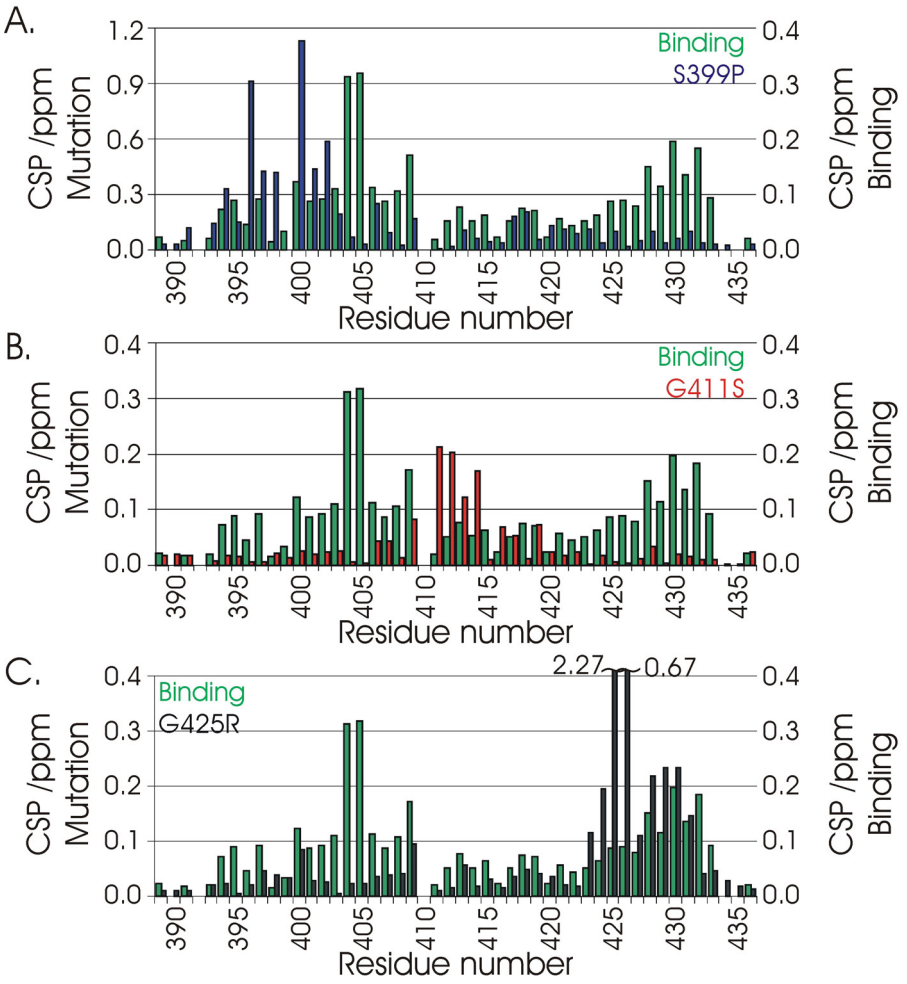



**Figure 7-14: Residue specific CSPs for the binding of Ub to $^{15}$N-p62-UBA (Green) compared to bound states CSPs caused by the PDB mutations; (A) S399P (Blue). (B) G411S (Red). (C) G425R (Black).**

As seen for the dimeric state, the S399P mutation causes significant CSPs across the entire domain which are considerably larger than the binding induced CSPs (Figure 7-14). Although the particularly large CSPs do not overlay with the binding induced CSPs there are considerable CSPs seen for many of the residues involved in binding, suggesting that changes in conformation which may affect the Ub binding site are possible. This is consistent with the changes in interface suggested in $^{15}$N-Ub titrations.

As with S399P the CSPs induced by the G425R mutant are much larger than those seen for binding, however in the case of G425R these are more localised to residues surrounding the mutation site (Figure 7-14). In the monomeric bound state the observed CSPs for the helical turn following the mutation site (Section 6.6.2)

appear extended to the next helical turn compared to the dimer. Relatively large binding induced CSPs are seen along the majority of helix 3 which contains the G425R mutation. Therefore the CSPs caused by binding and the mutation overlap in this case, providing evidence for direct interference with the interface as indicated by titrations with $^{15}$N-Ub. The extra CSPs observed for the helical turns following the mutation site may be an artefact of the secondary binding shifts as mentioned above. Large binding CSPs are seen for these residues and the lack of saturation in the G425R titration may contribute to these CSPs. However those residues immediately following the mutation site have magnitudes which are inconsistent with the majority of the contribution originating from binding induced CSP suggesting that the overlap, in this region at least, is not an artefact.

### 7.4.4 ESI-MS titrations of p62 UBA

In order to confirm the stoichiometry of the interaction of PDB mutants with Ub ESI-MS titrations were recorded on the PDB mutants S399P, G411S, and G425R and compared to that recorded for the wt UBA. ESI-MS spectrum were recorded on 1 μM p62-UBA PDB mutant with 0.2 – 3 μM Ub. In all cases mass ion peaks were observed corresponding to monomeric and dimeric p62-UBA, monomeric Ub, and a 1 to 1 complex of p62-UBA with Ub. Again no 2 to 1 or 2 to 2 complexes of any kind were seen in the concentration range examined by ESI-MS.

### 7.4.5 Summary

Using a combination of biochemical and biophysical approaches the effect of PDB mutations in the p62-UBA have been investigated. All of the investigated PDB mutants show attenuation of poly-Ub binding in pull-down assays leading to suggestions that an impaired ability to bind Ub is a unifying concept among PDB mutations associated with the p62-UBA[202,228]. Investigations here, where the affinity for mono-Ub has been investigated by NMR, are largely in line with this observation. The PDB mutants S399P and G425R all show reduced binding to Ub, with the binding to G425R showing the largest reduction in affinity. However the G411S mutant which shows only marginal reduction in poly-Ub binding by pull down assays shows no measurable effect on mono-Ub binding by NMR.

Chemical shift mapping of the binding interface provides evidence for direct effects on Ub binding caused either by contact with the mutated side chain or

structural perturbations resulting from the incorporation of the mutation. No such interface perturbations were seen for P392L, the mechanism of the effects of this mutation therefore remains unclear.

## 7.5 PRE Studies of the p62-UBA complex with Ubiquitin

In order to gain insight into the structure of the p62-UBA complex with Ub, PRE based NMR experiments were performed on a 1 mM sample of $^{15}$N-p62-UBA in the presence of 6 mM Ub. Two types of PRE experiments were performed; the first was a repeat of that performed on the p62-UBA dimmer, whereby PRE caused by exposure to the soluble nitroxide radical containing compound HyTEMPO were measured in order to investigate the solvent accessible surface area (Section 6.3.1). The second PRE based experiment involved using the approach described by Fushman *et al* for determining the complex of the UBA domain of Ubiqulin-1 with Ub, in which a nitroxide radical containing compound was attached to Ub in order to measure intermolecular PRE in the p62-UBA.

### 7.5.1 Solution PRE Experiments

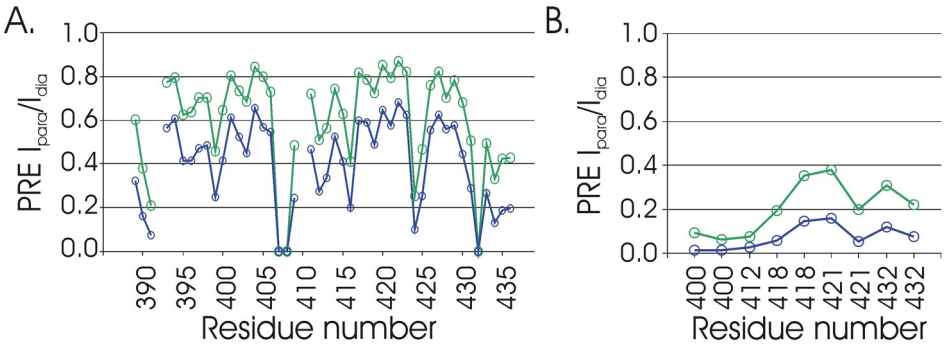



**Figure 7-15: Residue specific PRE values observed for the p62-UBA in complex with Ub caused by the soluble spin label HyTEMPO at 15 mM (Green) and 60 mM (Blue) for (A) Backbone NH (B) Side chain NH/NH$_2$ .**

In order to identify residues with significant exposure to solvent, $^{1}$H-$^{15}$N-HSQC were recorded on a 1 mM sample of p62-UBA mixed with 6 mM Ub in the presence of HyTEMPO at various concentrations between 15 and 60 mM. PRE was determined as the ratio of the $^{1}$H-$^{15}$N-HSQC peak intensities in the presence and absence of HyTEMPO. Large portions of the p62-UBA in complex with Ub show large PRE effects including Glu389, Ala390, Asp391, Ser399, Ser407, Asp408,

Trp412, Leu416, Ile424, Ala425, Ile431, Gln432, Tyr433, Ser434, Lys435, and His436 as well as all of the measured side chain NH and $NH_2$ (Figure 7-15).

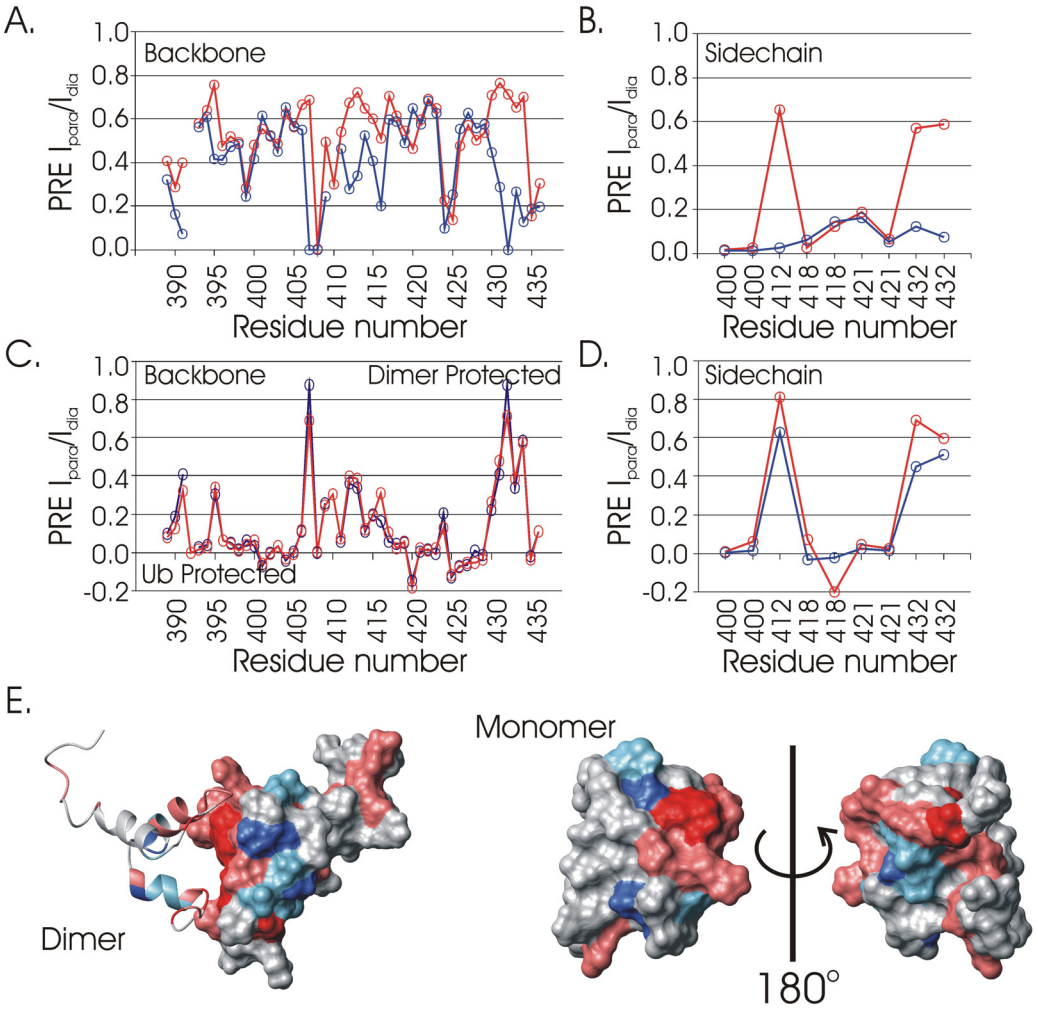



**Figure 7-16: (A-B) Residue specific PREs for the p62-UBA as a free dimer (Red) or as a Ub bound monomer (Blue) caused by 60 mM HyTEMPO. (C-D) The difference in PRE values between the free dimer and Ub bound monomer species of p62-UBA in the caused by HyTEMPO at 15 mM (Blue) or 60 mM (Red), negative values are seen for residues showing more solvent protection in the Ub bound monomer state; positive values are more protected in the free dimmer. (E) Residues which are more protected in the free p62-UBA dimer (Red) or the Ub bound monomer (Blue) plotted onto the p62-UBA dimer or Ub bound monomer (RPDB ID: 2KNV and 2JY8).**

The PRE effects observed with HyTEMPO show clear changes when compared to the free p62-UBA dimmer, with many more residues showing large PRE effects in the Ub bound monomer state than seen for the free dimeric species

(Figure 7-16). Subtraction of the observed PRE values for the Ub bound monomeric p62 from those for the free p62-UBA dimer allows identification of residues which are more solvent 'protected' in the dimer species or in the Ub-bound monomer species (Figure 7-16). Many residues show considerably more protection in the free dimeric state than in the Ub-bound monomer state. Considerably weaker PRE effects are seen in the free dimeric p62-UBA for residues Ser407, Gln432, Ser434, I431, and the side chain resonances of Trp412, and Gln432. Increased protection in the Ub bound monomer state is much less pronounced with only Met401, Lys420, and Gly425 showing significantly increased protection, however much weaker potential protection is also seen in residues Met404, Ala426, Ala427, Leu428 Asp429 and Leu418.

The observed differences are consistent with strong protection from the solvent caused by the relatively stable dimer interface of the p62-UBA (Figure 7-16). Protection caused by the binding interface between the p62-UBA and Ub are difficult to identify due to the large protective effects seen for the dimer interface which mask the apparent weak protection provide by the Ub binding interface. More information could be obtained if PRE measurements were available for the free monomer species, although this is not available at this time.

The residues showing marginal protection from solvent in the Ub bound state are close in the structure to those for the dimer binding site, again providing evidence for overlapping dimerisation and Ub binding surfaces on the UBA. Interestingly the observed residues showing protection from solvent in the Ub bound state include Gly425, supporting the binding induced CSPs and data for the G425R mutant suggesting that this residue, and therefore the G425R PDB mutant, are located at the binding interface.

### 7.5.2 Intermolecular Paramagnetic Relaxation Enhancement

Previous attempts to obtain unambiguous intermolecular distance restraints between Ub and the p62-UBA for subsequent structure generation had been unsuccessful. In a previous study by Fushman *et al,* where restraints for a complex between Ub and the UBA of ubiquilin-1 had been similarly lacking, a complex structure had been generated based on RDC restraints and intermolecular PRE measurements[152]. The PRE measurements used by Fushman *et al* were provided by the spin label MTSL (A chemical analogue of HyTEMPO) covalently attached to Ub

via a cysteine engineered into position 48. The MTSL coupled K48C mutant of Ub was then titrated into $^{15}$N-p62-UBA until saturation, $^{1}$H-$^{15}$N-HSQC were then recorded followed by reduction of MTSL with ascorbate, before recording a second $^{1}$H-$^{15}$N-HSQC in order to measure specific intermolecular PRE the p62-UBA resulting from MTSL bound to the Ub in complex with the UBA[99,152]. In principle the recorded PRE can be used to calculate distance restraints between the unpaired electron of MTSL and the measured backbone NH in an identical way to that done routinely for NOE measurements[68,152,216,218].

Significant signal attenuation was seen for all residues without exception with the weakest PRE effect measured for residues Leu417 and Tyr422 which had signal intensities only reduced to 60% of that seen for diamagnetic MTSL coupled K48C-Ub (Figure 7-17). Mapping the observed CSPs onto the surface of the p62-UBA monomer show an irregular pattern of PRE intensities inconsistent with the trends expected for intermolecular PREs between a stable complex[68,152](Figure 7-17). In order to compensate for potential solution effects from the high concentration of spin-labelled protein in solution, the solution PRE values obtained using HyTEMPO were subtracted from the data but showed no improvement.

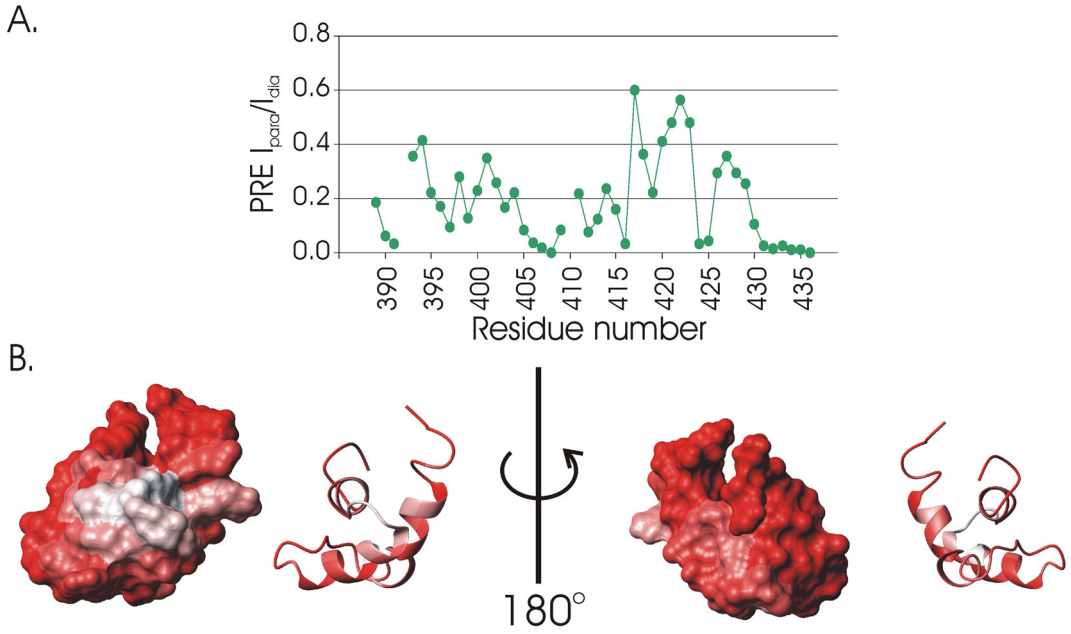



**Figure 7-17: (A) Plot of residue specific PRE effects for the p62-UBA in complex with MTSL-K48C-Ub. (B) Large PREs mapped onto the surface of the p62-UBA monomer structure in red (RPDB ID: 2JY8).**

Despite the lack of useful intermolecular PRE obtained from titrations using MTSL coupled K48C-Ub, potentially useful information was obtained in the form of intermolecular CSPs caused by the mutation of Lys48 to MTSL-Cys48, which is CSPs caused by the mutation of the Lys48 residue of Ub measured in the $^{1}H$-$^{15}N$-HSQC of the p62-UBA. Fushman *et al* demonstrated no structural perturbations resulting from the mutation of Lys48 to MTSL-Cys48, therefore the observed CSPs in the $^{15}N$-p62-UBA most likely originate from close cross-interface contacts involving the mutation site. These intermolecular CSPs are relatively localised with only residues Tyr433 and His436 showing significant CSPs, consistent with a close proximity of Lys48 of Ub to the C-terminal portion of the p62-UBA domain in the complex of the p62-UBA and Ub.

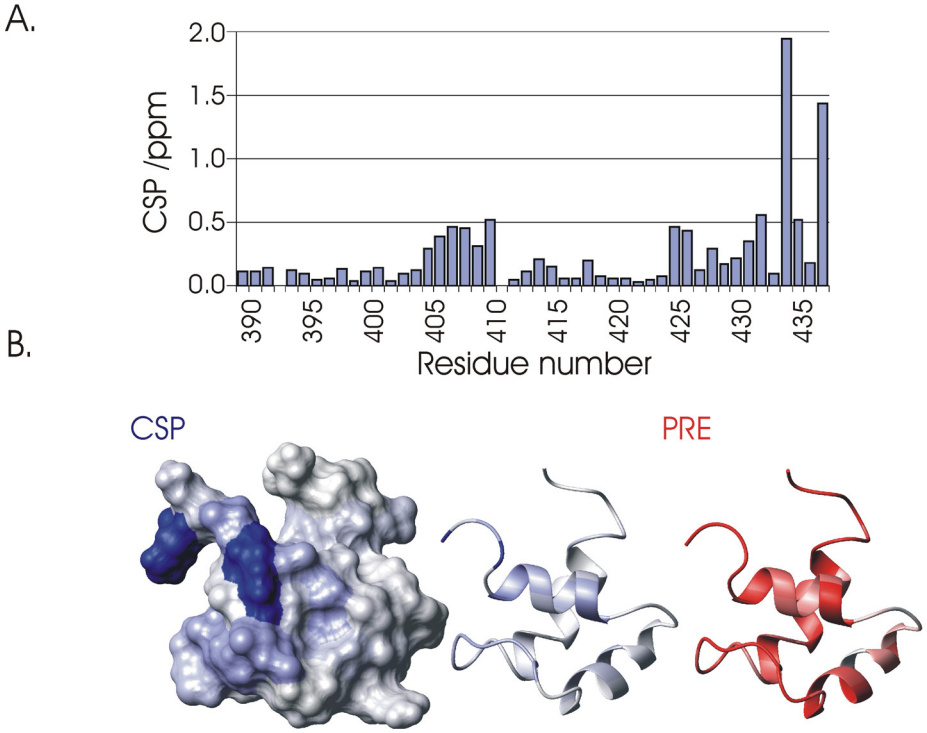



**Figure 7-18: (A) Plot of residue specific intermolecular CSP for the p62-UBA caused by the MTSL coupled K48C mutant in complex with the p62-UBA. (B) Large CSP (Blue) mapped onto the p62-UBA monomer (RPDB ID: 2JY8), a comparison with the MTSL induced intermolecular PREs (Red) is included.**

The attenuation of the majority of residues in the p62-UBA resulting from the p62-UBA complex with Ub is difficult to rationalise. One potential explanation would be to suggest a relatively non-specific interaction between the p62-UBA and

Ub, whereby the complex adopts no fixed position and is relatively loosely associated, resulting in large PRE effects for all residues in close proximity to the MTSL group in the various conformations of the 'non-specific' complex. However this seems contradictory to the relatively localised intermolecular CSPs and the relatively high affinity of the p62-UBA monomer complex with Ub, which would be expected to involve specific stabilising interactions to increase the binding energy.

Another potential explanation could be that proposed by Clore *et al* in the study of the interaction of the N-terminal domain of Enzyme I and the histidine-containing phosphocarrier protein, where intermolecular PREs were observed at non-binding site residues originating from non-specific encounter complexes which form part of the binding mechanism but are invisible by most approaches[218]. While this may account for the large scale PRE effects the magnitude of these PRE effects seems large for a non-specific, transient interaction, however a more detailed study would be required.

## 7.6 Conclusions

The observation that dimerisation and Ub binding are mutually exclusive processes suggests that dimerisation and Ub binding must compete. This may function as a novel control mechanism for the recognition of Ub by p62, and gives extra emphasis to the previously postulated potential mechanism for the disruption of p62-UBA functions by PDB mutants by altering the dimerisation affinity. PDB mutants which increase the affinity of the p62-UBA dimer will result in a reduction in the apparent affinity for Ub by reducing the concentration of monomer available for interactions with Ub.

Although the apparent $K_d$ of 0.7 mM observed in previous studies at high concentrations suggests a relatively weak interaction between the p62-UBA and Ub, this is the result of competition between dimerisation and Ub binding by the p62-UBA. Fitting the data to a competitive binding model using DynaFit reveals a much stronger interaction which had previously been masked by competition with the dimer under the examined high concentrations.

The affinity of Ile44 binding UBDs shows large variation from low mM to low μM, with variation which is as large between members of the same family as that between different families[17]. The apparent $K_d$ for the p62-UBA interaction with Ub and the actual $K_d$ represent opposite ends of the spectrum of binding affinities

reported for UBA domains, with 0.75 mM being one of the weakest reported affinities and 42 μM one of the higher affinity UBA's recorded[17]. This may support the notion that the dimer structure acts as a control mechanism, under conditions where the dimer is favoured the $K_d$ is typical of UBA domains, however under conditions where the monomer is favoured the apparent affinity is increased due to reduced dimer formation generating a 'strong' UBA. This provides a mechanism for an 'affinity switch' by mediating the conditions to favour dimer formation or dissociation. Given the low concentration of a given protein within the cell, the monomeric p62-UBA dimer may well be expected to be the dominant species, however the relatively low $K_{dim}$ of ~ 7 μM may allow mechanisms to increase the dimer population. The N-terminal PB1 domain of p62 functions as an oligomerisation domain and also interacts with a host of other proteins[174,177,240-242]. Dimerisation of the p62-UBA may therefore be induced by p62-PB1 mediated oligomerisation and/or increased p62 expression or by localisation to a specific organelle which would result in a high local concentration to facilitate both PB1 and UBA mediated homotypic intermolecular interactions. This may be reversible to favour the monomeric state perhaps by reducing cellular concentration, disrupting PB1 oligomerisation using alternative binding partners such as the aPKCs which interact with the p62-PB1 domain and may reduce oligomer formation[178,181], or directly disrupting UBA dimerisation by Ub-independent interactions of the p62-UBA. This is supported by the discovery of the apparent Ub independent interaction of the p62-UBA with MURF2[176].

This control mechanism may serve another purpose when considering the function of p62. One of the proposed roles of p62 involves shuttling of aggregate material including organelles, misfolded proteins and bacteria for degradation by autophagy[173,194]. In principle a major contributor to protein aggregation is the exposure of large hydrophobic patches caused by misfolding. In an environment rich in misfolded protein aggregates, the dimerisation mechanism may serve to conceal the hydrophobic patches of the p62-UBA domains which are not bound to Ub as a protection mechanism from non-specific interactions with misfolded cargo which could result in p62 aggregation.

Several studies have demonstrated that oligomerisation of p62 via its PB1 domain is as key to many of its functions, including TRAF6 K-63 linked poly-ubiquitination and regulation of Keap1, as the binding sites for the individual

signalling components. These functions may be mediated by the sequestration of signalling components into cytosolic speckles to mediate signalling complexes. However p62 oligomerisation may have implications for Lys63-linked poly-Ub signalling in NF-κB and autophagy pathways. A recent study by Sims *et al* of UBA domain interactions with Ub using GST-pull down assays concluded that oligomerisation (mediated by the GST-fusion) leads to linkage specific avidity effects which favour binding to Lys63-linked poly ubiquitin by presenting an array of UBA domains that enhance binding affinity through multiple UBA interactions with multiple linked Ub units through the avidity/chelate effect[237]. In the case of p62 the dimerisation of the p62-UBA adds a further level of complexity to this model whereby PB1 oligomerisation also results in increased UBA dimerisation by the same avidity effects. The avidity effects on UBA dimerisation disfavour binding to mono-Ub but favour the association of longer poly-Ub chains where multiple UBA interactions with Ub may compensate for the inhibitory effects of UBA dimerisation. The combination of UBA dimerisation and PB1 mediated oligomerisation may contribute to selective poly-Ub recognition and potentially to *in vivo* linkage specificity.

The PDB mutants P392L, S399P, G411S, and G425R have all been investigated for effects on Ub recognition by the p62-UBA as a potential contributor to the disease effects seen in these mutations. All of the examined PDB mutants attenuate poly-Ub binding, however only P392L, S399P and G425R appear to disrupt recognition of mono-Ub by the p62-UBA. NMR titrations suggest direct effects on the interface between Ub and the p62-UBA for the S399P and G425R mutants. CSP patterns and PRE measurements provide evidence that G425 and therefore the G425R mutation, is directly involved in the interface between the two proteins. The insertion of the large arginine side chain in place of a glycine may well disrupt the surface complementarity between the p62-UBA and Ub, disrupting both shape and chemical composition. On the other hand little or no binding CSPs are seen for helix 1 which contains the S399P mutation; however this mutation appears to subtly alter the structure of the p62-UBA and therefore indirectly disrupt the interaction surface for Ub binding leading to a reduction in affinity.

In poly-Ub binding by pull down assays the G425R mutant appears to disrupt recognition to a greater extent compared to S399P than the relative $K_d$s for mono-Ub might suggest. This may be the result of the differences in dimer affinity, both the

S399P and G425R mutations bind Ub with a $K_d$ in the order of 200 μM, however the S399P mutant has a reduced dimerisation affinity compared to wt p62-UBA while the G425R mutant increases the dimer affinity. The outcome at high concentration is that the S399P loss in affinity is compensated for by a reduction in dimer affinity resulting in modest apparent $K_d$ changes, while the increased dimer stability of G425R accentuates its loss of Ub binding affinity.

It is worth noting that in this study the p62-UBA and Ub originate from two different species (human and yeast), yeast ubiquitin caries 3 mutations compared to human including P19S, E24D, and A28S none of which cause significant structural changes or are located close to the bindig interface with the p62-UBA. The species variation is therefore not anticipated to affect the structure in this case.

The precise mechanism of how Paget's disease mutations within the p62-UBA result in a predisposition to Paget's disease remains unclear; however the results presented here along with previous studies within the Searle and Layfield groups at the University of Nottingham provide an ever growing insight into the molecular effects of PDB mutants.

# Chapter 8

# 8 Disrupting the p62-UBA Dimer Interface by Mutation

## 8.1 Introduction

The p62-UBA forms a high affinity dimer *in vitro* (Chapter 6) and the p62-UBA dimer must be dissociated in order to allow recognition of the p62-UBA's primary binding partner Ub (Chapter 7). Both dimerisation and Ub recognition by the p62-UBA *in vitro* have been discussed in detail in previous chapters (Chapters 6 and 7), however the *in vivo* significance of the p62-UBA dimer remains uncertain.

The dimerisation of the p62-UBA has the potential to function as a control mechanism by modulating the affinity of the p62-UBA for Ub. However the reported concentrations of an individual protein are generally low, for example the most abundant natural protein in *E. coli*, EfTu, is reported to have a cellular concentration of approximately 1 μM[243] and the concentration of ubiquitin in HeLa cells, approximately 10 μM[244]. Most proteins would therefore be expected to be at sub μM concentrations. Despite the p62-UBA dimer affinity being relatively high at around 7 μM, at these low cellular concentrations it may not be expected to be the dominant species *in vivo.*

However the conditions within the cell are radically different to those *in vitro*; protein localisation, molecular crowding and the presence of the cytoskeleton may all affect the behaviour and oligomerisation state of proteins, making accurate predictions of a protein's behaviour inside the cell difficult[245,246]. What's more the PB1 domain of p62 functions as an oligomerisation domain, with a $K_d$ in the order of 50 pM (based on other PB1-PB1 interactions)[247] which facilitate localisation into small cellular bodies or 'speckles' resulting in an increased local concentration which may facilitate p62-UBA dimerisation[173]. The regulation of UBA dimerisation by PB1-mediated oligomerisation may be particularly important given the evidence that a number of p62 functions require both the PB1 and UBA domains including TRAF6 mediated NF-κB regulation[190].

As p62 possesses no intrinsic measurable activity, functional studies are difficult and relying instead on indirect measurements of function. p62 has strong links with the regulation of NF-κB activity, and previous studies have shown that over expression of p62 is associated with repression of NF-κB activity and that certain PDB linked mutants of p62 may reduce the ability of p62 to repress NF-κB in luciferase reporter assays[61,204,248,249]. Luciferase reporter assays therefore provide a useful indicator of the biological function of p62, which can be compared to the *in vitro* functional data such as binding affinities or stability, facilitating a better understanding of p62 and its mutants.

### 8.1.1 Summary and aims

The *in vivo* significance of p62-UBA dimerisation remains unclear. In order to examine the functional implications of p62-UBA dimerisation, in terms of Ub binding and NF-κB activation, mutations were designed to disrupt the dimerisation. Using a combination of biophysical measurements and luciferase reporter assays the effects of reducing the p62-UBA dimerisation affinity was examined in order to gain insight into the potential *in vivo* significance of p62-UBA dimerisation.

## 8.2 Disrupting the Dimer Interface with Lysine Mutations

In a recent examination of homodimer interfaces the amino acid lysine was found to be the least favoured interface residue[214]. In order to disrupt the dimerisation of the p62-UBA, residues at the dimer interface, namely Glu409 and Gly410, were mutated to lysine (Figure 8-1). In the case of the G410K mutation, the large increase in steric bulk between the Gly and Lys side chains, as well as the preference for Lys side chains to be solvent exposed was anticipated to cause significant dimer disruption. In addition, the inserted Lys side chain would be in close proximity to the solvent exposed Arg415 in the adjacent monomer, resulting in charge repulsion which may also contribute to reduced dimer formation. In addition to the unfavourable properties of Lys at interfaces the E409K mutation was chosen to disrupt a potential intermolecular charge-charge interaction between Glu409 and Lys420 as the mutation to Lys409 would result in charge repulsion in the dimer leading to reduced dimer stability[160,161](Figure 8-1).

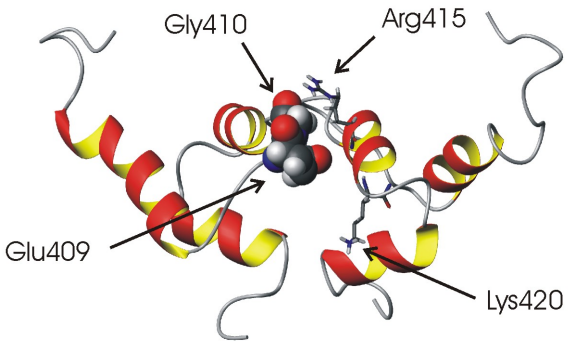



**Figure 8-1: Ribbon diagram of the p62-UBA dimer (RPDB ID 2KNV) showing the position of the mutated residues Glu409 and Gly410.**

### 8.2.1 Biophysical Characterisation of p62-UBA Lysine Mutants

The E409K and G410K mutants were successfully generated by site directed mutagenesis by Dr Ping Chen. As with the wt-UBA, SEC was utilised during the purification of the E409K and G410K mutants of the p62-UBA. During SEC it was observed that although both mutants eluted at a volume consistent with a dimeric species, the elution volume had shifted indicating a smaller mass than the wt protein (Figure 8-2).

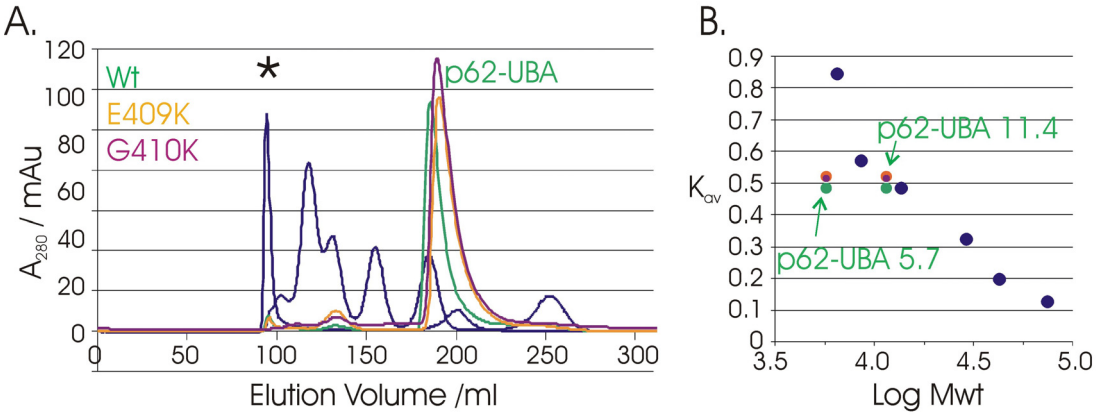



**Figure 8-2: (A) SEC trace for wt p62-UBA (Green) the dimer-disrupting mutants E409K (Orange) and G410K (Purple) and Low MWt proteins (Blue)(* denotes blue dextran). (B) The standard curve used to calculate the apparent native mass of p62-UBA. The $K_{av}$ against expected mass for monomeric and dimeric p62-UBA are included.**

This apparent reduction in size recorded for the two lysine mutants could indicate a compacting of the fold with maintenance of the oligomerisation state or

more likely, in this case, by a reduction in dimer affinity resulting in a reduction in apparent size due to higher population of monomer at equilibrium. However the mass is still consistent with a dimeric species indicating that dimerisation had not been completely ablated, but may have been only reduced.

While the observed shift in SEC elution volume gives a good indication that the p62-UBA dimerisation has been effectively reduced, a more quantitative analysis of the effects of these mutants can be provided by biophysical analysis using ITC, CD ESI-MS and NMR.

#### 8.2.1.1 ESI-MS of p62-UBA Lysine Mutants

The oligomerisation state of the dimer-disrupting mutants E409K and G410K was investigated by native ESI-MS in the concentration range 1 to 8 µM. In all cases the highest order oligomer detected was a dimer observed in low quantities (Figure 8-3). As discussed previously (Section6.2.2), although the populations of monomer and dimer can not be accurately determined the population changes in monomer and dimer will be reflected in changes in the average charge-state of the apparent monomer ions with concentration. Both the E409K and G410K mutants show similar magnitude concentration dependence in the average charge-state when compared to wt, however the charge-state values are consistently higher in the two lysine mutants consistent with either an increase in solvent exposed surface area resulting from partial unfolding, or a larger population of the monomeric species comparable to the lower concentration range of the wt p62-UBA titration (Figure 8-3).

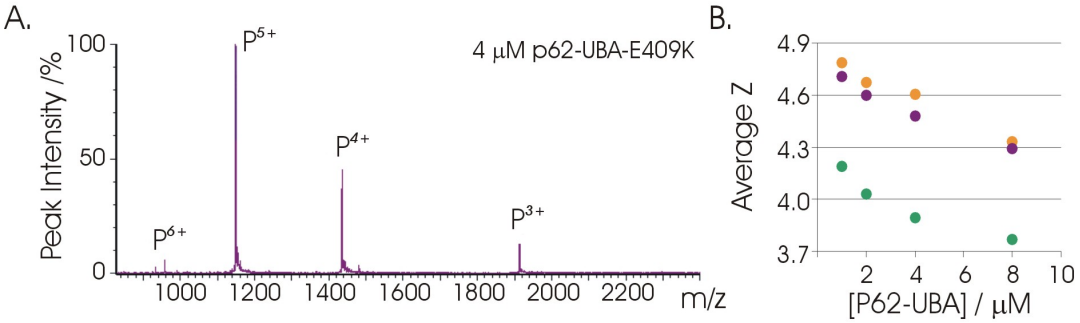



**Figure 8-3: (A) ESI-MS of 4 µM E409K. (B) Concentration dependence of the average charge-state for wt-p62-UBA (Green), E409K (Orange) and G410K (Purple).**

### 8.2.1.2 ITC of p62-UBA Lysine Mutants

In order to determine the thermodynamic parameters for the p62-UBA dimer-disrupting mutants E409K and G410K, ITC dilution experiments were recorded whereby concentrated p62-UBA (~2 mM) was diluted by sequential injections into buffer in order to observe the calorimetric response of dimer dissociation induced by rapid dilution. A series of endothermic heat pulses of decreasing intensity were observed resulting from gradually decreasing dissociation to monomer with increasing protein concentration within the calorimetric cell, until such point where saturation occurs and dissociation is no longer observed (Figure 6-6). The resulting dissociation isotherm of the integrated heat response was fitted to a dimer dissociation model to give a $K_{dim}$ and ΔH at 298 K, from this the $\Delta G_{298}$ and ΔS were determined (Table 6-2).

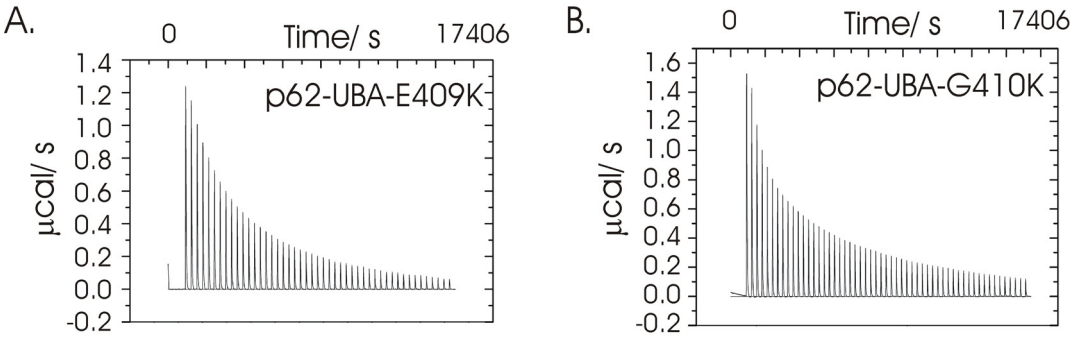



**Figure 8-4: ITC dilution thermal output for; (A) the p62-UBA-E409K (B) G410K.**

**Table 8-1: Thermodynamic parameters for dimer-disrupting mutants of p62-UBA derived from ITC.**

| | $K_{dim}$ (±) / µM | ΔH (±) / kJ/mol | ΔS (±)/ J/mol/K | $\Delta G_{298}$ (±) / kJ/mol |
|---|---|---|---|---|
| p62-UBA | 7 (1) | 6.8 (3) | -75 (5) | 29.4 (1) |
| E409K | 143 (10) | 6.5 (3) | -53 (5) | 21.9 (1) |
| G410K | 90 (15) | 13.6 (4) | -35 (10) | 23.0 (1.4) |

The fits in this case were poor compared to the wt protein, which suggests that the dimer dissociation model is inadequate for describing the response possibly suggesting dimers of dimer or tetramers. However this is inconsistent with either the

ESI-MS or SEC which both suggest the presence of a dimer with no evidence for higher order complexes. The wt protein was shown to aggregate at high concentrations[36], given the high concentrations used in the ITC in this instance, and the sensitivity of the ITC response, it may be likely that the poor fits result from low level aggregation at high concentration.

#### 8.2.1.3 CD studies of dimer-disrupting mutants of the p62-UBA

In order to investigate the effect of the introduced mutations on the stability of the p62-UBA domain, equilibrium unfolding experiments were performed monitored by far-UV CD spectroscopy on the mutants E409K and G410K. As with the studies of the wt p62-UBA the ellipticity at 222 nm was recorded as a function of temperature in the range 283 – 368 K and the apparent $T_m$ for the transition was calculated at a range of concentrations from 1 – 200 μM.

The observed CD spectrum for the examined mutants gave similar classical helical CD spectrum with no large differences between the observed spectrum and that seen for wt p62-UBA at high or at low concentrations (Figure 8-5). Both mutants give strong negative bands at 222 nm and 208 nm and as observed for the wt p62-UBA the ratio of $\theta_{222}/\theta_{208}$ shows concentration dependence. However both lysine mutants show significantly weaker concentration dependence than the wt, with no ratio of over 0.95 measured at any point. Assuming that the observed change in $\theta_{222}/\theta_{208}$ ratio is associated with the dimer to monomer population shift, the reduction in value and concentration dependence may indicate either a decrease in the dimer affinity so that the population is no longer shifting significantly in the measured concentration range, or that the tertiary structure of these mutants is similar in terms of helical packing and/or pitch in the monomer to that seen in the dimer which results in the reduced concentration dependence in the $\theta_{222}/\theta_{208}$ ratio[67,168,212,213](Figure 8-5).

The thermal stability of both lysine mutants was measured as a function of temperature and showed weak concentration dependence. The recorded difference in apparent $T_m$ between 1 and 200 μM is around 2 and 4 K for the E409K and G410K mutants respectively, this is considerably smaller than the 8 K observed for wt p62-UBA (Table 8-2). This is consistent with a reduction in variation of the $\theta_{222}/\theta_{208}$ ratio and ITC which suggests a decrease in dimer affinity, therefore in the concentration range measured in these experiments the population changes and the associated

changes in $T_m$ are considerably reduced compared to those observed in titrations with p62-UBA wt.

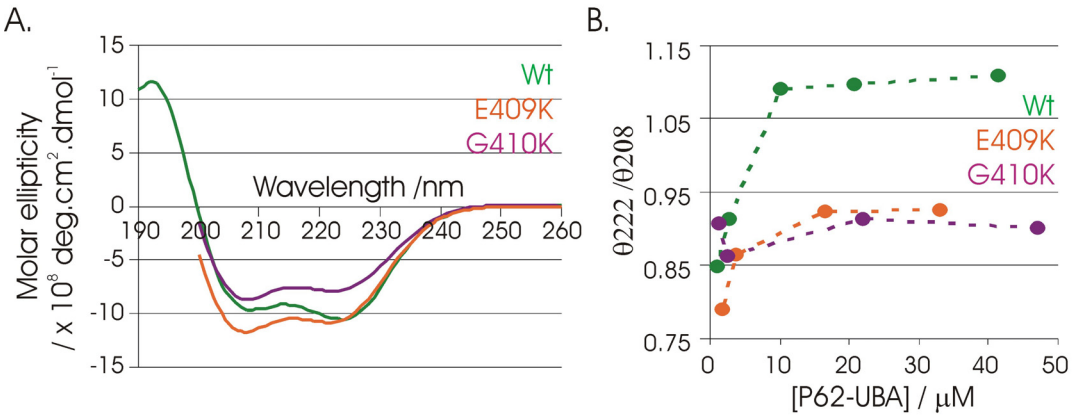



**Figure 8-5: (A) Far-UV CD spectrum for the p62-UBA-Wt (Green) and the dimer-disrupting mutants E409K (Orange) and G410K (Purple). (B) Plots of $\theta_{222}/\theta_{208}$ as a function of concentration.**

Both the E409K and G410K mutants show decreased thermal stability compared to wt, with monomeric $T_m$ that are around 9 and 13 K less stable than the wt UBA for E409K and G410K respectively (Table 8-2). While these are significant reductions which may have the potential for detrimental effects on function, they are both more stable than the two PDB mutants S399P and G411S and would be over 90 % folded at cellular temperature[231].

**Table 8-2: Thermodynamic parameters for the thermal unfolding of the p62-UBA mutants E409K and G410K derived from far-UV CD.**

| | $T_m$ (±)/ K | $\Delta T_m$ / K | $\Delta G_{298K}$ (±)/ kJ/mol | $\Delta H$ (±)/ kJ/mol | $\Delta S$ (±)/ J/mol/K | $FF_{298K}$ / % | $FF_{309K}$ / % |
|---|---|---|---|---|---|---|---|
| | 'Dimer' | | | | | | |
| Wt | 347.2 (1.5) | - | 31.6 (1.8) | 229.8 (1.6) | 660 (6) | 99.9 | 98.5 |
| E409K | 333.0 (2) | -14.2 | 15.4 (2.2) | 155.9 (1.6) | 471 (8) | 98.5 | 93.1 |
| G410K | 330.2 (2) | -17.0 | 14.3 (2.4) | 159.2 (1.6) | 486 (8) | 99.0 | 94.0 |
| | 'Monomer' | | | | | | |
| Wt | 339.2 (1) | - | 24.1 (0.6) | 201.4 (2.9) | 595 (2) | 99.2 | 96.7 |
| E409K | 330.2 (2) | -9.0 | 11.6 (0.3) | 121.1 (0.8) | 367 (2) | 98.0 | 93.0 |
| G410K | 326.2 (2) | -13.0 | 12.8 (1.2) | 154.6 (0.8) | 475 (2) | 97.5 | 91.0 |

Given the reduced dimer affinity determined from ITC, supported by ESI-MS and the CD data, the presented 'dimer' values in Table 8-2 for the two dimer-disrupting mutants is unlikely to represent the true dimer values as at the measured concentration a significant population of monomer remains.

### 8.2.2 NMR Characterisation of p62-UBA Lysine Mutations: Observation of an Alternative Dimer

The recorded biophysical data including ITC, ESI-MS and far-UV CD is consistent with the two incorporated loop 1 mutants resulting in disruption of the dimer interface with E409K being the weaker dimer of the two. In order to characterise the structural effects of the incorporation of the two dimer-disrupting mutants $^{1}$H-$^{15}$N-HSQC were recorded on both $^{15}$N-p62-UBA-E409K and $^{15}$N-p62-UBA-G410K at a range of concentrations between 1 – 0.1 mM. At high concentrations both mutants showed significantly more complicated spectra when compared to wt p62-UBA, consistent with multiple species at equilibrium (Figure 8-6). Assignment of the $^{1}$H-$^{15}$N-HSQC of both mutants was achieved using sequential connectivities using $^{1}$H-$^{15}$N-HSQC-NOESY aided by comparison with wt-p62-UBA dimer and monomer, and by chemical exchange cross peaks in $^{1}$H-$^{15}$N-HSQC-NOESY and $^{1}$H-$^{15}$N-HSQC-TOCSY. The assignment revealed the presence of three species detectable at equilibrium for both mutants; including a species with chemical shifts which are comparable to the wt dimer, a species consistent with the monomeric p62-UBA, and a third unknown species which shows similarly large chemical shifts differences when compared to either the dimeric or monomeric wt-p62-UBA. The chemical shifts are consistent with the same third species detected in both E409K and G410K with it being the most populated species at equilibrium at high concentration of G410K but only detectable at low levels for the E409K mutant at high or low concentration.

The population of all three species was seen to vary as the concentration was reduced from ~ 1 mM to ~0.1 mM. $^{1}$H-$^{15}$N-HSQC were recorded at several concentrations and the relative intensities of residues which show little or no overlap in all three species were used to estimate the concentration of each population as a function of total protein concentration (Figure 8-6).

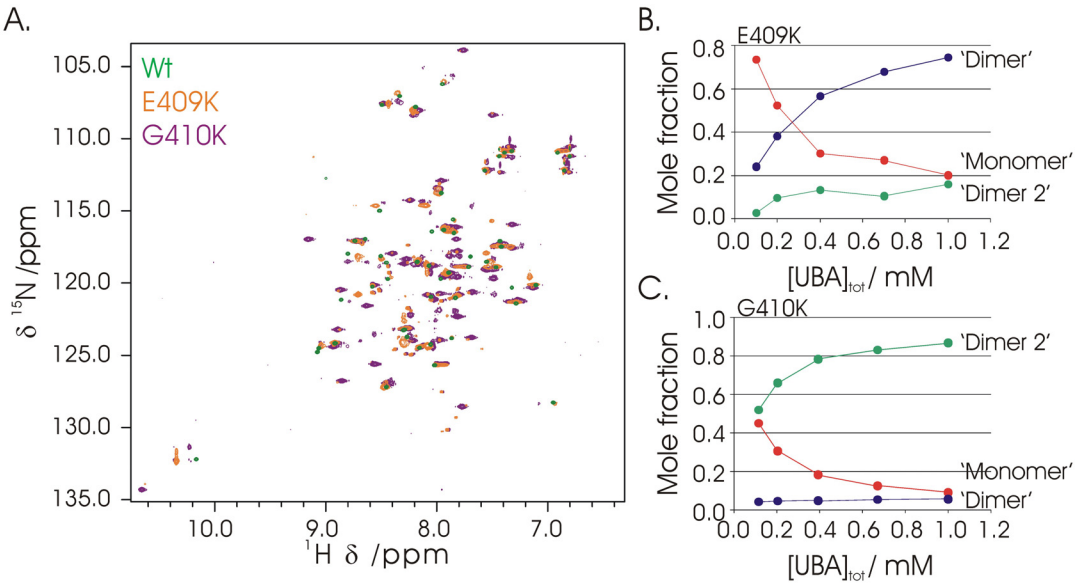



**Figure 8-6: (A) $^{1}$H-$^{15}$N-HSQC of 1 mM $^{15}$N-wt-p62-UBA (Green) and the dimer-disrupting mutants E409K (Orange) and G410K (Purple). (B-C) Concentration dependence of the populations of the three species present in the two dimer-disrupting mutants.**

The concentration dependence of the observed species followed the same trend in both mutants; the species with chemical shifts similar to those of the p62-UBA monomer increased in intensity relative to the other two species as the concentration was decreased consistent with a monomeric species, the two other species both decreased in intensity as the concentration was decreased suggesting that both are dimeric or higher order species. The oligomerisation state of the two higher order states was presumed to be dimer based on the following evidence from other techniques: The ESI-MS spectrum shows no evidence of oligomers larger than a dimer at concentrations up to 8 µM. Although both of these mutants would be expected to be predominantly monomeric in this concentration range, the ESI-MS would under normal circumstances detect low concentrations of dimer and other higher order oligomers should they be present at equilibrium. The low population of dimer and lack of any higher order mass ions therefore provides good evidence for the lack of higher order species. The second piece of evidence for the lack of higher order structures comes from SEC, both the E409K and G410K mutants have a marginally increased elution volume relative to wt-p62-UBA. This is inconsistent with the formation of oligomers larger than those seen for wt which has been shown

to form dimers in solution. What's more the observed elution volume is consistent with a mass of around 13 KDa consistent with a dimer, the shift in elution volume for the two lysine mutants is consistent with the reduced affinity in the dimer affinity as discussed previously.

The chemical shifts of the first dimer species are consistent with the wt-p62-UBA with chemical shift perturbations arising from the mutation site, this species is termed from now on dimer1. The second dimer species, known from now on as dimer2, has chemical shifts which are distinct from either the monomeric or dimeric species and therefore represents an alternative dimmer. Only a single set of peaks are present for dimer2 suggesting that although the structure is likely to be different to the wt-dimer it maintains the C2 symmetry seen for wt. In the case of the E409K mutant dimer1 is the dominant species at equilibrium with dimer2 representing less than 20 % of the observed intensity at 1 mM, however the opposite is seen for the G410K mutant where dimer 2 accounts for almost 90% of the signal intensity while dimer1 represents less than 5%. The population of the observed species at various concentrations was used to estimate the dimer affinities of the observed species assuming that the two dimer species are described by independent equilibrium and that one is not an on pathway intermediate in the formation of the other. The concentration of the two dimeric species were combined to estimate the observed dimer affinity of monomer to dimer* (dimer1 + dimer2) (Table 8-3).

**Table 8-3: Table of affinity constants for the dimerisation of p62-UBA and the two dimer-disrupting mutants E409K and G410K determined by NMR.**

| | $K_{dim}$ Dimer* (±)/ μM | $K_{dim}$ Dimer1 (±)/ μM | $K_{dim}$ Dimer2 (±)/ μM |
|---|---|---|---|
| wt-p62-UBA | 6 (3)[61] | - | - |
| E409K | 190 (40) | 220 (50) | 1200 (100) |
| G410K | 80 (40) | 1800 (200) | 86 (40) |

* denotes sum $K_{dim}$ for the equilibrium between monomer and the sum of the dimer species.

The $K_{dim}$ determined by NMR are consistent with those previously determined by ITC, with both dimer mutants reducing the affinity by more than an order of magnitude from wt demonstrating that the desired effect of the mutations in disrupting the dimer interface was achieved.

In the case of the E409K mutant, significant CSPs are observed in both the dimer and monomer compared to the wt protein. In the case of the dimer the CSPs are wholly consistent with direct contact with the mutation site either intramolecular or intermolecular across the dimer interface, consistent with the expected repulsion between the incorporated Lys side chain and Lys420 and/or Arg415 in the adjacent monomer (Figure 8-7). In the case of the monomer the observed CSPs are restricted to those in the immediate proximity of the mutated side chain, consistent with little or no significant perturbations to the domain structure resulting from the Lys409 side chain (Figure 8-7).

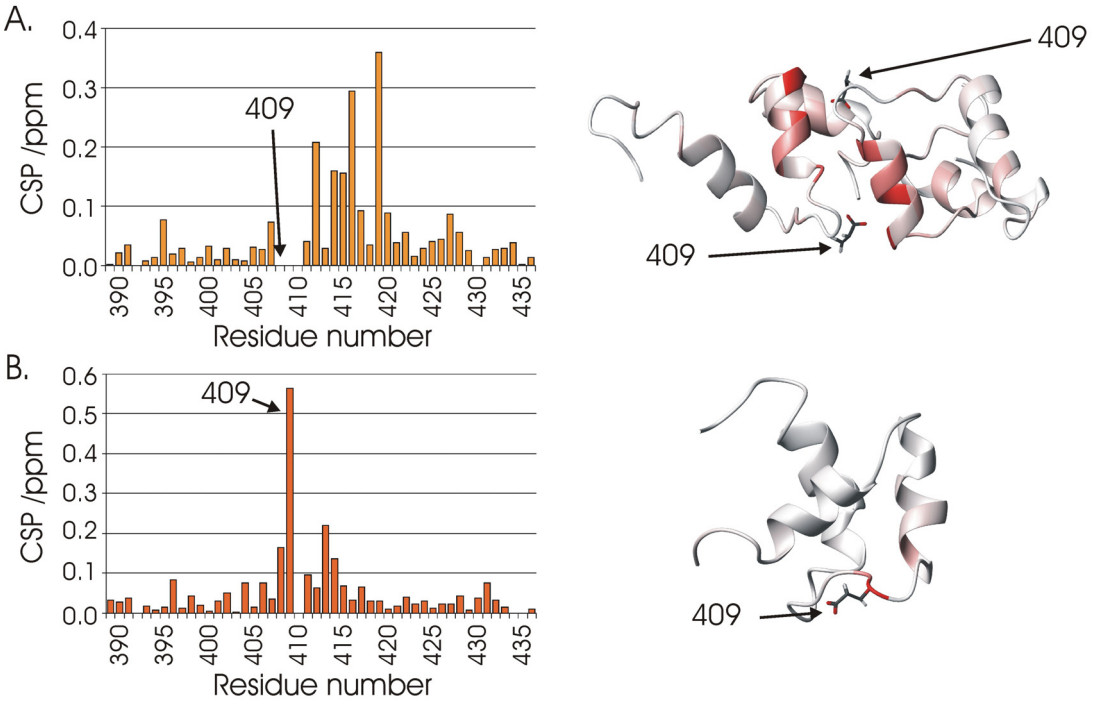



**Figure 8-7: Residue specific CSPs resulting from the E409K mutant for; (A) the p62-UBA dimer (B) p62-UBA monomer. The relevant CSPs are plotted onto the dimer or monomer structures of the p62-UBA (RPDB ID: 2KNV and 2JY8).**

However this is quite the opposite in the case of the G410K mutant where the monomer species shows only CSPs localised to the mutation site but in the dimeric species the incorporation of Lys410 appears to result in the formation of an alternative dimer structure (Figure 8-8). The chemical shifts of dimer2 shows similar magnitude difference when compared to the wt-dimer or wt-monomer, this is consistent with a fold which is different to that of the wt-dimer but not a dimer of the proposed alternatively packed UBA fold suggested for the monomeric p62-UBA[180]

(Figure 8-8). The presence of multiple species of the G410K mutant at equilibrium results in a large number of exchange correlations in $^{1}H$-$^{15}N$-HSQC-NOESY which prevents the generation of unambiguous NOE derived distance restraints for a full structural characterisation of dimer2. In order to gain insight into potential changes in secondary structure of dimer2 the CSD was examined for the $^{1}H\alpha$, $^{1}H^{N}$, and $^{15}N^{H}$ chemical shifts and compared to wt-p62-UBA in its dimeric or Ub bound monomeric state (Figure 8-8). The CSDs of dimer2 are similar in magnitude and sign to that for both of the wt states suggesting neither significant change in local secondary structure across the three helices or local unfolding. The only significant changes in CSD are seen in loop 1 where the mutation is situated, consistent with subtle changes surrounding the mutation site.

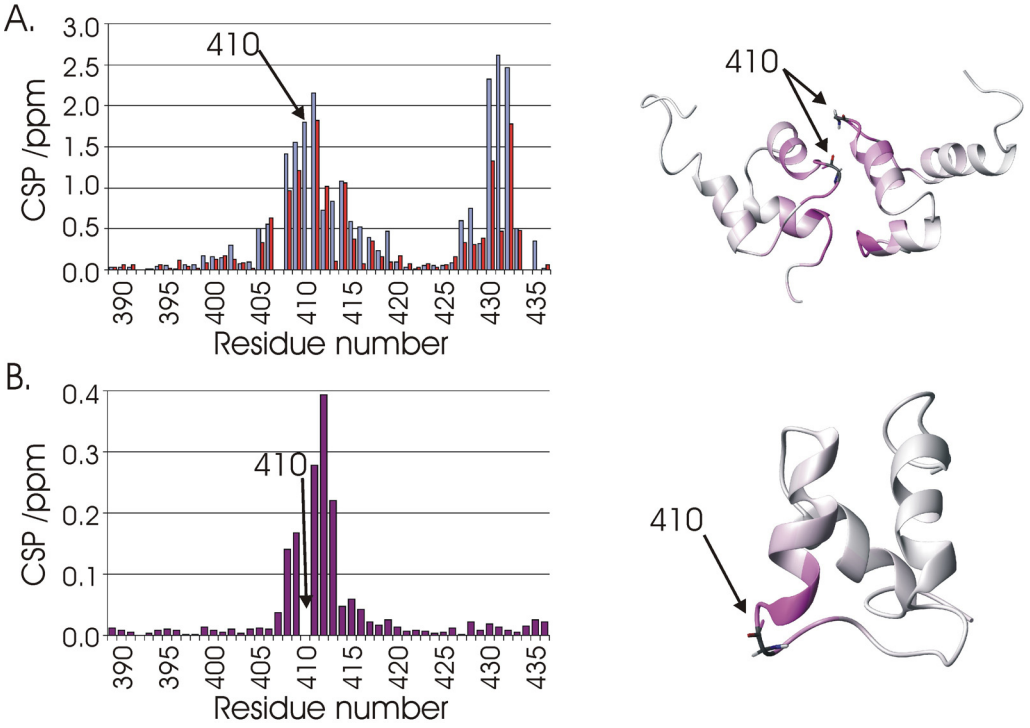



**Figure 8-8: Residue specific CSPs for the G410K dimer compared to; (A) p62-UBA dimer (Blue) and monomer (Red). (B) G410K monomer compared to wt monomer. The relevant CSPs plotted onto the dimer or monomer structures of the p62-UBA are included (RPDB ID: 2KNV and 2JY8).**

The large CSPs between the wt dimer and the G410K dimer2 appear to be localised to the dimer interface (Figure 8-8) suggesting that the differences between dimer1 and dimer2 may be in the orientation of the two monomer units consistent

with only small changes in CSD, however a more detailed structural investigation is required.

The nature of dimer2 remains unclear, while a full structural characterisation of dimer2 may be possible using backbone $^{13}C$ chemical shift derived dihedral angle restraints and RDCs, it was deemed outside of the remit of this investigation.

The dimerisation induced CSPs between dimer1 measured for wt and E409K and for dimer2 of G410K, are remarkably similar with large CSPs localised to helix 2 and the C-terminal end of helix 3, suggesting that both dimer1 and dimer2 utilise a similar dimerisation interface (Figure 8-9). Although large CSPs are seen for dimer1 and dimer2, in the case of dimer2 the largest CSPs are seen for helix1 with fewer large CSPs in helix 3, this is the opposite to dimer1 where the larger CSPs are seen in helix3 (Figure 8-9). The changes in the magnitude of the observed CSPs are difficult to interpret without a detailed structural characterisation, however this is consistent with the notion that the difference in the two dimers lies in the orientation of the two monomer units in the dimer interface.

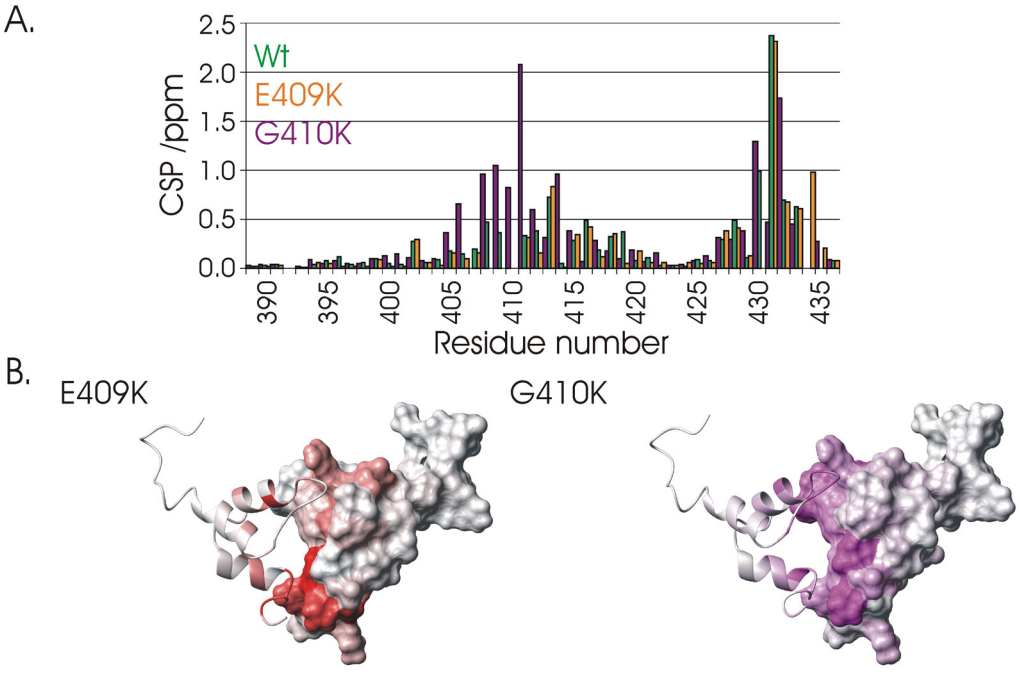



**Figure 8-9: (A) Residue specific CSPs for the dissociation of dimer2 of G410K (Purple), dimer1 of E409K (Orange) and wt dimer (Green). (B) Large CSPs plotted onto the structure of the p62-UBA dimer E409K (Red) and G410K (Purple) (RPDB ID: 2KNV).**

### 8.2.3 Summary

In order to investigate the functional significance of the dimerisation of the p62-UBA two separate mutations were engineered into loop 1 of the p62-UBA aimed at disrupting the dimer interface and therefore reducing the dimerisation affinity. The two mutants E409K and G410K were characterised using ITC, ESI-MS, far-UV CD, and NMR in order to examine their dimer affinity, stoichiometry, stability and structure. Both mutants showed a $K_{dim}$ which was at least an order of magnitude weaker than that observed for the wt protein giving $K_{dim}$ of 165 ± 40 and 90 ± 30 μM for the E409K and G410K mutants respectively compared to 7 ± 2 μM for the wt-p62-UBA.

Interestingly both mutations show evidence for the presence of a non wt-like dimer at equilibrium which is currently uncharacterised. However only a small population of this alternative dimer is measured for the E409K mutant and the monomer species which is proposed to be the biologically active species has chemical shifts which are consistent with no large structural changes resulting from either the incorporation of E409K or G410K mutations. Despite the formation of a non-native dimer species the reduced dimer affinity p62-UBA mutants provide good candidates for examination of the functional implications of disrupting the dimerisation of the p62-UBA.

## 8.3 Effects of Dimer Disruption on the Function of p62: Ub recognition

The two loop one mutants E409K and G410K were designed to disrupt p62-UBA dimerisation with the intention of examining the subsequent effects on the function of p62. The two functions of particular interest to this investigation are the interaction of the p62-UBA domain with its key binding partner Ub *in vitro,* and the ability of p62 to down regulate NF-κB *in vivo* [204,249]. These functions are considered to be of particular interest due to experimental links to the role of p62 and PDB for both UBA domain and non UBA domain mutations.

### 8.3.1 The Effect of Dimer Disruption of Ub Binding

#### 8.3.1.1 $^{15}$N-UBA Titrations with Ub

In order to investigate the effects of the dimer-disrupting mutants on the ability of the UBA domain to bind to Ub, unlabelled Ub was titrated into 1 mM $^{15}$N-UBA-E409K or G410K and the response measured using $^{1}$H-$^{15}$N-HSQC. In both cases the two dimeric species decrease in intensity with increasing concentration of Ub, while the monomeric species both increased in intensity and shifts with increasing concentration of Ub (Figure 8-10).

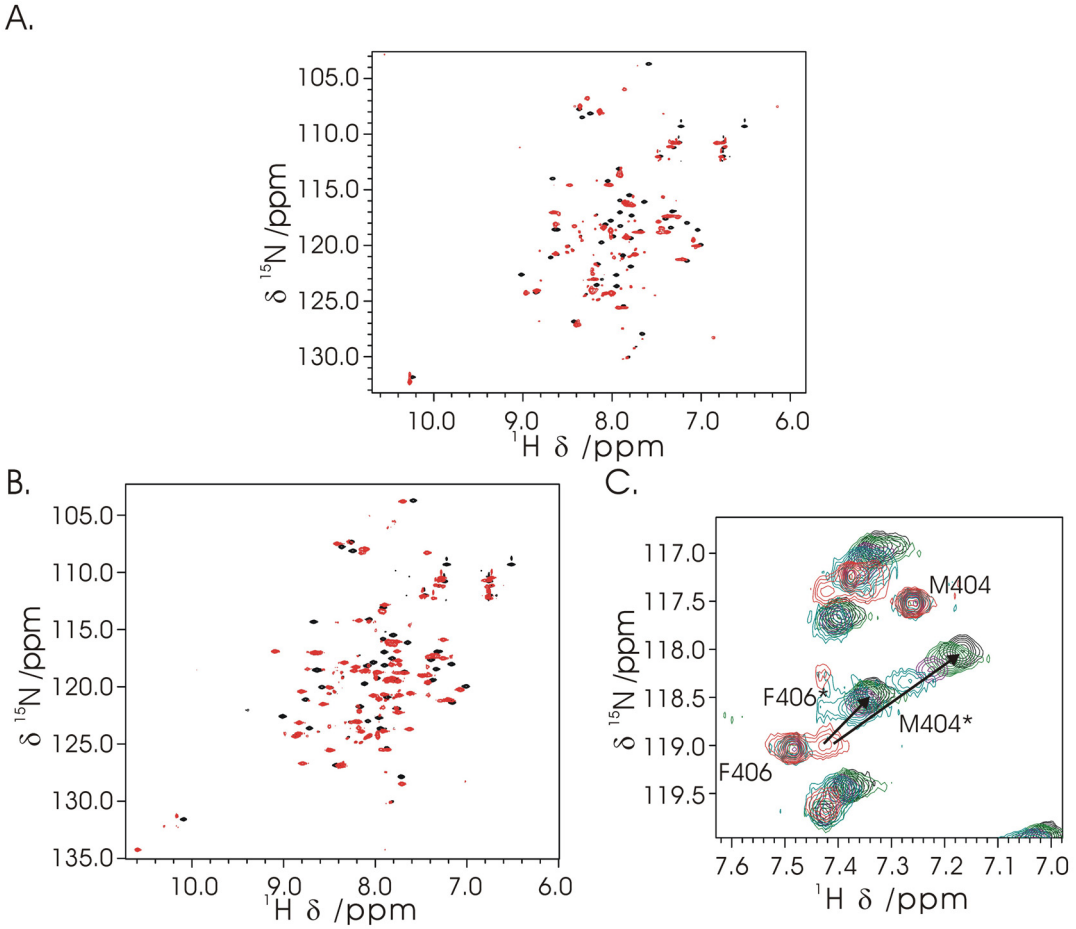



**Figure 8-10: $^{1}$H-$^{15}$N-HSQC of; (A) 1mM $^{15}$N-UBA-E409K. (B) 1mM $^{15}$N-UBA-G410K (B) in the presence of 0 (Red) and 6 mM (Black) Ub. (C) A highlighted section of $^{1}$H-$^{15}$N-HSQC of 1 mM G410K in the presence of 0 (Red), 0.6 (Teal), 1 (Purple), 2 (Green) and 6 mM (Black) Ub; '*' indicates monomer peaks.**

The relative intensities of residues which show little or no overlap in all three species in the $^{1}$H-$^{15}$N-HSQC were used to estimate the concentration of each UBA

population as a function of Ub concentration. The mole fraction of monomeric UBA reaches 1 at a Ub concentration of 1 mM (1 to 1 concentration of Ub to UBA) for E409K and 2 mM for G410K, under near identical conditions the wt-UBA reaches a monomeric mole fraction of 1 at around 5 mM Ub (Figure 8-11).

Binding induced chemical shifts are seen only for the monomeric species of both mutants which appear to saturate between 1 and 2 mM Ub, the magnitude of the observed chemical shift changes show the same chemical shift perturbations as those seen for wt-p62-UBA with large chemical shift perturbations seen for Met404 and Gly405 in loop 1 and in helix 3 (Figure 8-11).

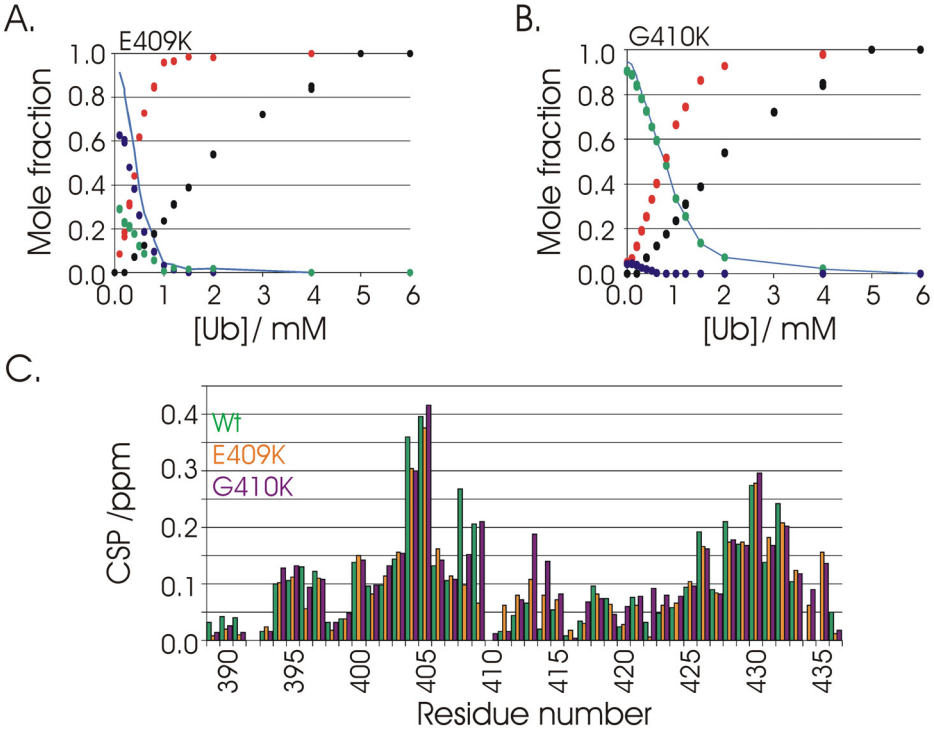



**Figure 8-11: Mole fraction of dimer1 (Blue) dimer2 (Green) and monomer (Red) as a function of Ub concentration for; (A) p62-UBA-E409K. (B) G410K. The total dimer fraction is represented by a sold blue line and the corresponding monomer population for wt-p62-UBA is included (Black). (C) The binding induced CSPs for wt-p62-UBA (Green), E409K (Orange) and G410K (Purple).**

The observed reduction in populations of the two dimeric species with a concurrent increase in the population of monomeric species with increasing concentration of Ub, combined with chemical shift perturbations observed for the monomeric species only, is consistent with Ub binding to the monomeric species of

both dimer-disrupting mutants with no detectable interaction with either of the dimer species. Both of the dimer-disrupting mutants saturate at lower Ub concentrations than the wt UBA suggesting an increase in the apparent affinity at the concentration used in the NMR experiments.

An increase in the apparent affinity for Ub for the two dimer disrupted mutants is consistent with the competitive model, whereby reduction in the dimer affinity results in an increase in the apparent affinity for Ub as a result of reduced competition between dimerisation and Ub binding. This similarly explains the observation that the monomer population of E409K, which has a lower dimer affinity than G410K, also dominates at lower Ub concentration than observed for G410K. The lack of significant changes in the binding induced CSPs between wt-p62-UBA, E409K and G410K, with the exception of residues immediately adjacent to the mutation site, suggests no significant changes in the binding interface caused by either of the dimer-disrupting mutants.

#### 8.3.1.2 $^{15}$N-Ub titration with p62-UBA-E409K

In order to examine the effect of the E409K mutant on the Ub binding interface and on the apparent affinity for Ub, unlabelled p62-UBA-E409K was titrated into 1 mM $^{15}$N-Ub and the response monitored using $^{1}$H-$^{15}$N-HSQC (Figure 8-12). The binding of the p62-UBA-E409K mutants resulted in concentration dependent chemical shift changes which saturate between 1 and 2 mM. The final CSPs are similar to those observed for binding to wt-p62-UBA with marginally larger magnitude CSPs observed for the E409K mutant (Figure 8-12). The line broadening seen in titrations with wt-p62-UBA for residues Arg42, Lys48, and Leu71 is seen in the titration with E409K but are less severe, and full peak intensity is restored at concentrations of Ub above 1 molar equivalent.

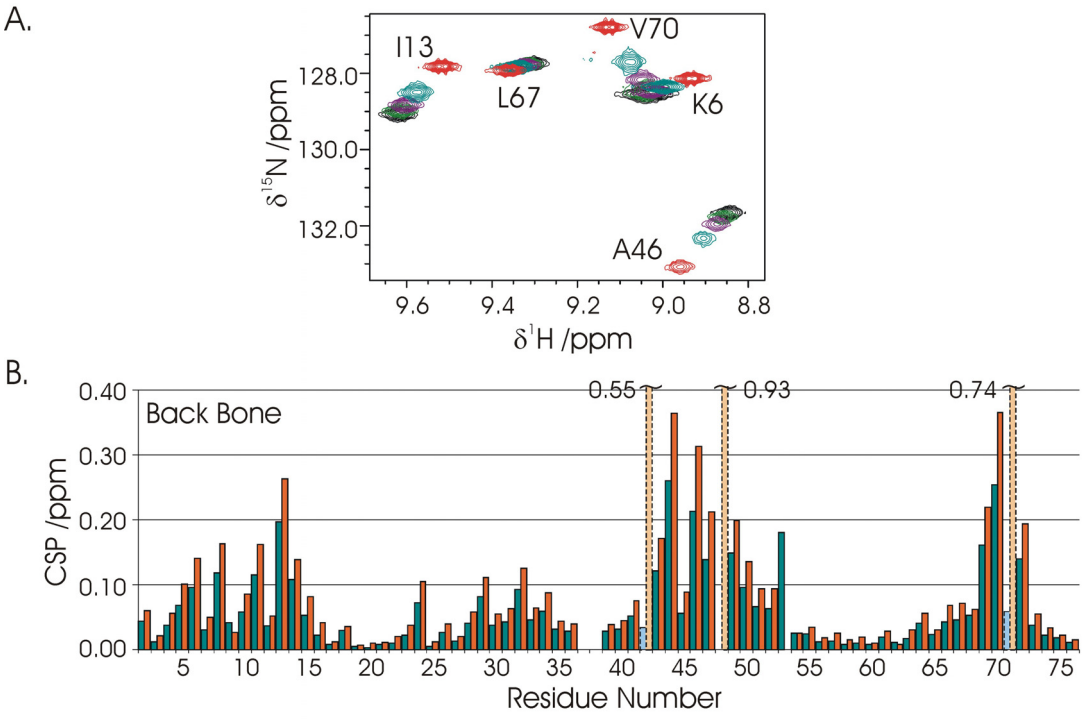



**Figure 8-12: (A) Section of $^{1}$H-$^{15}$N-HSQC of 1 mM $^{15}$N-Ub in the presence of 0 (Red) 0.6 (Teal) 1 (Purple) 2(Green) and 6 mM (Black) p62-UBA-E409K. (B) Binding induced CSP for $^{15}$N-Ub bound to wt-p62-UBA (Green) and the E409K mutant (Orange) residues which broaden during the titration are represented by faded dashed lines.**

The binding induced CSPs as a function of the concentration of the p62-UBA were used to calculate an apparent $K_d$ at high concentration by fitting to a one to one binding model in IGOR Pro™ 5.05A, and used to estimate the actual monomer $K_d$ by fitting to a competitive binding model using the ITC and NMR determined values of $K_{dim}$ using DynaFit (V 3.28.070, Biokin)[59,60](Table 8-4). The p62-UBA-E409K mutant gives an apparent $K_d$ at 1 mM of 486 ± 30 µM for the interaction with Ub which is significantly higher affinity than that observed for the wt protein of 741 ± 80 µM. On the other hand the estimated actual $K_d$ is close to the wt value recording a $K_d$ of 32 ± 3 µM for the E409K mutant compared to 42 ± 2 µM for the wt protein (Table 8-4).

**Table 8-4: Apparent and actual dissociation constants for the interaction of Ub with p62-UBA-E409K.**

| | Apparent $K_d$ (±) /μM | Actual $K_d$ (±) /μM | $\Delta G_{298\ mono}$ (±) / kJ/mol | $\Delta\Delta G_{298\ mono}$ (±) / kJ/mol |
|---|---|---|---|---|
| p62-UBA | 741 (80) | 42 (3) | -24.96 (0.15) | - |
| E409K | 486 (30) | 32 (3) | -25.6 (0.20) | 0.61 (0.21) |

The similarity of the binding induced CSPs between Ub binding to the wt protein or to the E409K mutant suggests that the binding of the monomeric form of the p62-UBA to Ub is unaffected by the Lys409 side chain and that it is unlikely to make direct contact with Ub as little or no clear CSPs are observed. The increased CSP magnitude and reduction in line broadening is most likely the product of the E409K mutant reaching saturation (or close to) where the wt-protein does not. The lack of CSPs between binding to wt compared to the E409K mutant, supports the notion of an indirect effect mediated by disrupting the dimer interface rather than by increasing the affinity for Ub directly by additional contacts mediated by the Lys409 side chain. This is supported by the observation that the apparent $K_d$ at high UBA concentrations is reduced in the E409K mutant compared to the wt protein with little corresponding affinity increase in actual affinity of the p62-UBA monomer for Ub (Table 8-4).

Overall NMR titrations demonstrate that, as one would expect, the disruption of p62-UBA dimerisation results in a subsequent increase in the apparent affinity of the UBA domain for Ub. This is consistent with the proposed competitive binding model and supports the notion that the dimerisation of the p62-UBA may function to regulate the affinity of the UBA for Ub. While the mutants selected here are situated in a region of loop1 not thought to be intimately involved in Ub recognition, a number of dimer interface residues are common between Ub recognition and dimerisation, allowing for the possibility of generating future mutants which may both disrupt dimerisation and Ub recognition.

### 8.3.2 The Effect of Dimer Disruption on NF-κB activity

One of the best characterised roles of p62 is the regulation of signalling pathways leading to the activation of NF-κB[204]. Previous reports have demonstrated

that over-expression of p62 results in repression of NF-κB activity in luciferase reporter assays, and that impaired p62 function resulting from PDB mutants leads to a degree of recovery of NF-κB activity. In order to investigate the effects of the two dimer-disrupting mutants E409K and G410K on the ability of p62 to mediate regulation of NF-κB activity, NF-κB luciferase reporter assays were performed in U20S cells by the group of Dr Robert Layfield[61] using wt-p62, the two dimer-disrupting mutants E409K and G410K, and the PDB mutant G425R which has both increased dimer affinity and reduced Ub binding.

Over expression of wt-p62 resulted in a decrease in NF-κB activity to around 25 % of that seen without p62, over expression of the two dimer-disrupting mutants E409K and G410K resulted in further reduction of NF-κB activity down to around 17 % while over expression of the PDB mutant G425R results in partial recovery of NF-κB activity compared to wt-p62 to around 38 % (Figure 8-13). The recovery of NF-κB activity compared to wt-p62 for the G425R suggests a reduction in the ability of p62 to modulate the signalling pathways associated with NF-κB activation which is consistent with previous reports[204,248]. The opposite is seen for the two dimer-disrupting mutants suggesting that impaired dimerisation of the p62-UBA increases the ability of p62 to regulate signalling events associated with NF-κB.

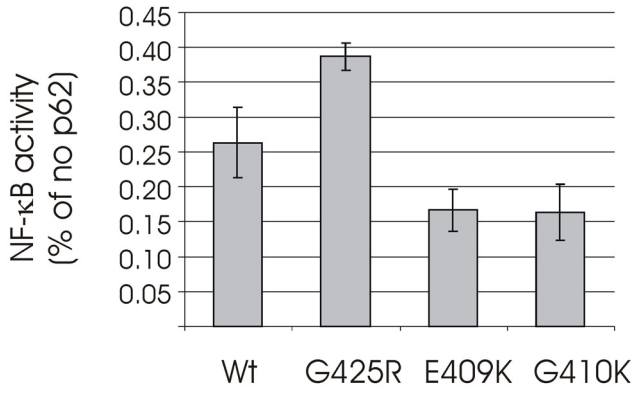



**Figure 8-13: Effects on relative basal NF-κB in U20S cells co transfected with p62, the dimer-disrupting mutants E409K and G410K and the PDD mutant G410K measured using a luciferase reporter gene.**

The increased down regulation of NF-κB *in vivo* by the over expression of p62 containing the UBA dimerisation disrupting mutations E409K or G410K compared to wt-p62, suggests an increase in the function of p62 as a result of

reduced dimerisation. The potential implication of this is that p62-UBA dimerisation negatively regulates the ability of p62 to regulate NF-κB.

## 8.4 Conclusions

The p62-UBA dimer has been effectively destabilised by the incorporation of lysine residues at key positions in loop 1 of the p62-UBA which forms part of the dimer interface. Using a combination of biophysical measurements and NMR the p62-UBA mutants E409K and G410K were characterised and showed to reduce the dimer affinity compared to wt-p62 by over an order of magnitude in both cases. The functional implications of dimer disruption on the p62 protein were assessed using NMR to study the ability of the p62-UBA to interact with its key binding partner Ub, and using luciferase reporter assays to assess the ability of p62 to down regulate NF-κB activation. Both Ub binding and modulation of NF-κB activity are increased as a result of reduced dimerisation of the p62-UBA, demonstrating clear functional consequences of UBA dimerisation in p62. The increased binding to Ub in the dimer disrupted p62-UBA mutants is wholly consistent with the previously outlined model of competition between UBA dimerisation and Ub binding, adding to the body of evidence for the proposed competing equilibria model.

The role by which impaired UBA dimerisation increases p62 function remains unclear, the results presented here, and in previous studies, demonstrate a clear correlation between the ability of p62 to interact with Ub and its ability to down regulate NF-κB[202,248]. However in this case it is not possible to determine whether the effects are mediated via Ub recognition or by a Ub independent interaction which is modulated by dimerisation. The potential Ub dependent and Ub independent functions of p62-UBA dimerisation may be dissected by a comparison with the effects of mutations which affect both dimerisation and Ub recognition, just dimerisation and just Ub recognition. Ub is intimately involved in many aspects of NF-κB regulation involving many different Ub-receptors and a variety of Ub chain linkages[12,27,33,123], the notion of Ub recognition by p62 as key to the ability of p62 to modulate signalling events associated with NF-κB activation therefore seems particularly attractive.

Although the degree of p62-UBA dimerisation *in vivo* is unknown, the use of mutations which disrupt dimerisation results in clear functional effects *in vivo*, providing evidence for a functional population of dimeric p62-UBA *in vivo*. The

potential roles of the p62-UBA dimerisation are numerous; the results here support the proposed role as a control mechanism for regulating the function of p62 both in the effective affinity for Ub and in the ability of p62 to regulate NF-κB. However it is possible that the p62-UBA dimerisation may be involved in Ub independent interactions involving either the p62-UBA monomer or utilising surfaces specific to the p62-UBA dimer, to date such interactions have not been identified although a Ub-independent interaction with MURF2 has been proposed[176].

# Chapter 9

# 9 Investigating a Ub Mediated Ternary complex with ZNF216 and p62

## 9.1 Introduction

### 9.1.1 Interpretation of Ubiquitination by UBDs

Protein ubiquitination has been the focus of intense research for the past 30 years, it is central to many cellular processes including protein quality control, DNA repair, epigenetics, and NF-κB signalling[2,7,159]. The versatility of ubiquitination in comparison to other post-translational modifications such as phosphorylation or acetylation is regularly attributed to two key factors[1,2]. The first is the size of Ub, the majority of post-translational modifications involve the addition of a relatively small chemical group such as phosphate, on the other hand Ub is a small protein presenting around 450 $Å^2$ of solvent exposed surface area, providing a large area for interaction with binding partners[1,2,159]. The second is the ability of ubiquitin to form poly-Ub chains by the formation of peptide bonds between its C-terminus and any one of its seven lysine side chains or its N-terminus. This has the potential to form a hugely diverse range of polymers of different lengths and of different chain linkages[2].

Although reports have suggested that the addition of Ub and poly-Ub chains can affect the intrinsic stability of the protein to which it is attached[13], it is generally accepted that Ub functions via interaction with Ub-receptors containing UBDs. The majority of proteins bind to only a small number of binding partners, however ubiquitin is an outstanding exception to this observation[159]. To date there are around 20 different UBD families with over 200 proteins in the human genome predicted to contain at least one UBD[14]. UBDs are diverse in secondary structure content, size, and in their affinity for Ub providing the potential for huge structural variety in Ub-UBD pairings.

Despite the versatility of chain linkage types and lengths, poly-Ub chains are still essentially repeating units of the same protein and it is therefore of particular interest how UBDs successfully interpret ubiquitination to mediate the correct

process based on simple differences in chain linkage type or length of poly-Ub. Many UBDs have shown exquisite selectivity for chain linkages mediated via specific contacts with the linker regions or by linkage specific interfaces generated by contacts between two connected Ub molecules[2,24,25]. The ability of UBDs to effectively interpret the ubiquitination code is largely attributed to such chain linkage selectivity but other factors play an important role. These factors may include Ub-receptor expression levels and/or co-localisation of the target protein and the Ub-receptor to the same machinery, such as seen in the 'endosomal sorting complex required for transport' (ESCRT) machinery, where the recruitment of Lys63-tagged cell surface receptors is facilitated by a host of membrane-associated UBD-containing proteins[250].

### 9.1.2 Ub Mediated Ternary Complexes

However one aspect of Ub mediated signalling which remains unclear is the mechanism by which Ub-receptors pass ubiquitinated proteins from one UBD-containing protein to the next. Such processes must occur in the UPS in order to transfer ubiquitinated cargo from proteasomal shuttling proteins to the proteasome, but the same is also true of the majority of Ub mediated signalling pathways which often contain multiple Ub-receptors[16,17,250]. There is also the possibility for decisions of fate, for example ubiquitinated receptors can be internalised for degradation by the proteasome or lysosome, or deubiquitinated and returned to the membrane. Such fate decisions may well be controlled by the selective passing of the ubiquitinated cargo from one UBD-containing protein to another, highlighting the need for an improved understanding of the mechanism[20,250].

The overwhelming majority of UBDs interact with ubiquitin via the canonical Ile44 centred hydrophobic patch, which most likely limits any proposed mechanism of passing ubiquitinated substrates from one UBD to the next, to those involving competitive displacement of the ubiquitinated substrate from one UBD to another with either higher affinity for Ub or of considerably higher local concentration[20]. However the recent emergence of UBDs which utilise other parts of the 450 $Å^2$ surface area of Ub, including binding sites centred on Leu8, Asp58, and the C-terminus[2,15,16], opens up the possibility of a 'hand-off' mechanism, whereby a second UBD interacts with a binary complex between a UBD and Ub to form a ternary complex, followed by dissociation of the first UBD[20]. Evidence for such a

mechanism is clear in the crystal structure of Rabex-5 in complex with Ub, Rabex-5 contains two UBDs a Znf_A20 and an MIU which bind to the Asp58 and the Ile44 centred patches and can be seen contacting the same molecule of Ub simultaneously to form a ternary complex in the crystal lattice[20,21].

However the key to Ub's ability to recognise a diverse range of UBDs, each with high specificity, has been attributed to the intrinsic flexibility of the solvent exposed regions of Ub[2,159], with the binding mechanism proposed to be dominated by conformational selection with only residual induced fit[251,252]. It is therefore possible that the binding of one UBD may limit the conformational space accessible to Ub leading to impaired binding to a second UBDs at a distant binding site to the first, preventing the formation of ternary complexes.

### 9.1.3 Summary and Aims

Ub signalling is mediated by the formation of non-covalent interactions of Ub with UBDs and interpretation of the ubiquitination code may largely be mediated by linkage specificity of many of UBDs. However the surface area of Ub features a variety of potential binding sites, giving rise to the possibility of the formation of ternary complexes between different UBDs which bind to different parts of the Ub surface. This has been demonstrated in the crystal lattice of the Rabex-5 complex with Ub which clearly shows the formation of a ternary complex between Ub and two molecules of Rabex-5 bound via different UBDs[20,21].

The implications of Ub mediated ternary complexes have yet to be explored but would add an extra level of complexity to ubiquitin mediated signalling with potential implications for the proposed 'hand-off' mechanism, the general mechanism of ubiquitin recognition, as well as implications for the function of Ub-receptors. Ub-receptors may be co-localised via Ub with other Ub-receptors which are involved in similar processes or signalling pathways but for which there is no binary interaction. The formation of ternary complexes also has the potential to facilitate specificity for poly-Ub chain linkages or even chain linkage conformations resulting from different poly-Ub chain lengths.

In order to assess the potential for ternary complex formations by UBDs other than the MIU and Znf_A20 of Rabex-5, 22 structures representing single members of the 22 known UBD families were examined for their ability to form a ternary complex with Ub. The predicted compatible pair of a Znf_A20 domain and a UBA

domain was tested using two domains from biologically-related proteins in the form of the ZNF216-Znf_A20 domain and the p62-UBA domain. Both p62 and ZNF216 function as proteasomal shuttling factors, both are linked to NF-κB regulation and share several common interaction partners linked to NF-κB regulation including TRAF6 and RIP[113,116,117,178,181]. The potentially biologically relevant ternary complex between the Znf_A20 and p62-UBA was assessed using a combination of biophysical and biochemical approaches.

## 9.2 Predicting Ternary Complexes from Binary Complex Models

### 9.2.1 UBD Ternary Complexes

In order to evaluate the possibility of Ub mediated UBD pairings, representative members of the 22 known UBD families so far characterised were obtained from UBD-Ub structures deposited in the protein data bank (RPDB). Each structure was overlaid using Ub residues 2-72 in MOLMOL[64] and the resulting ternary complex was analysed for the presence of van der Waals clashes and was rated from highly likely to highly unlikely to form a ternary complex on steric grounds (Figure 9-1). The examined domains contain single Ub binding sites, both the complex for the MPN and Josephin domains present UBD bound to Di-Ub, in these cases the proximal Ub was used for the comparison, in both cases the distal Ub was more open for interaction with UBDs. The represented domain families used in this investigation, their RPDB ID, and main Ub binding loci are summarised in Table 9-1.

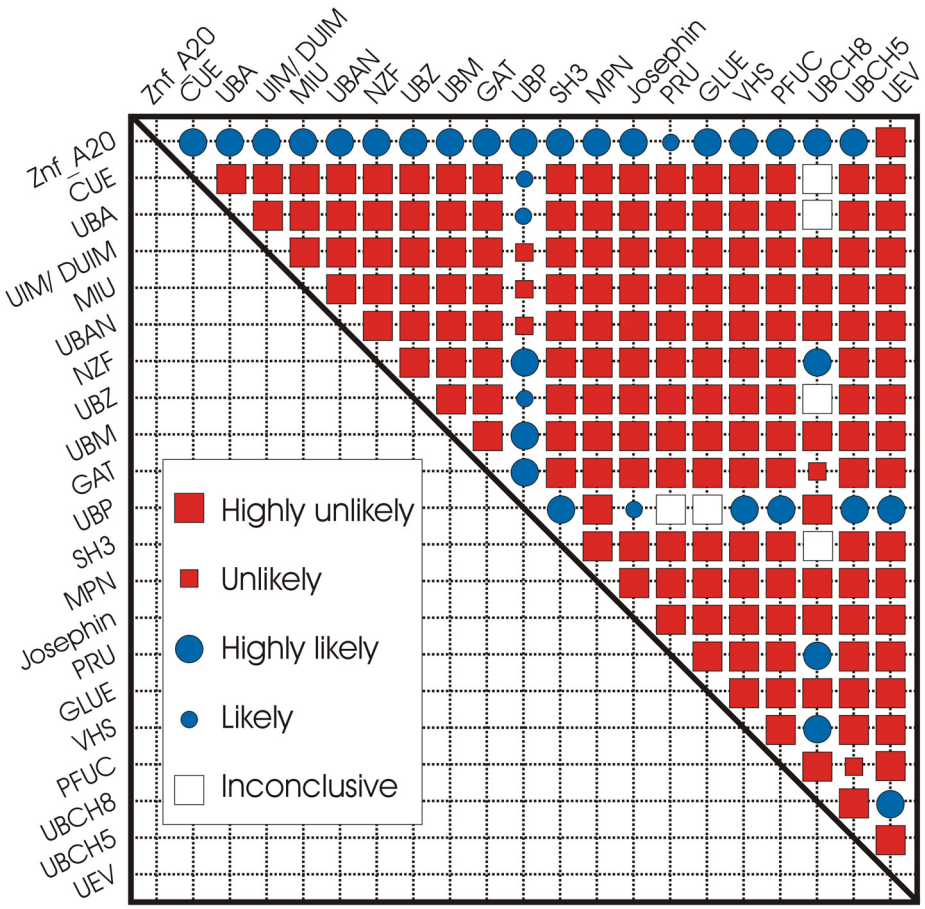



**Figure 9-1: Likelihood of UBD pairs to form a ternary complex with Ub based on steric grounds, ternary complex formation was classed as highly likely, likely , inconclusive, unlikely or highly unlikely.**

The backbone RMSD for Ub across all 22 complexes was low (BB RMSD = 0.57 Å) consistent with previous reports[159,251,252] which suggest that UBD-specific conformational preferences are not a significant additional consideration when considering potential ternary complex formation. As anticipated, the majority of UBD pairings compete for the common Ile44 centred patch, however a total of 35 pairings out of a possible 210 UBD-pairs are rated likely or highly likely to form ternary complex. A number of representative structures from the different categories are highlighted in Figure 9-2. Interestingly no UBD family appears restricted to binary interactions, with all being classed as likely to form at least one ternary complex. However no UBD can a form ternary complex with all UBD families with the Znf_A20 appearing the most versatile, rated likely to form ternary complexes with 20 of the 21 other known families (Figure 9-1).

**Table 9-1: Summary of the Ub complex structures used to predict potential ternary complex formation, domains belonging to E2 and E3 ligases or DUBs are labelled.**

| Domain | RPDB ID | Binding patch | Domain | RPDB ID | Binding patch |
|---|---|---|---|---|---|
| UIM | 1Q0W | Ile44 | PRU | 2Z59 | Ile44 |
| MIU | 2C7N | Ile44 | UBCH8 (E3) | 2KJH | Leu71/Gly76 |
| DUIM | 2D3G | Ile44 | GLUE | 2DX5 | Ile44 |
| UBAN | 2ZVN/2ZVO | Ile44 | PFUC | 2K8B | Ile44 |
| CUE | 1OTR | Ile44 | GAT | 1YD8 | Ile44 |
| UBA | 2DEN | Ile44 | UBM | 2KHW | Leu8 |
| UBZ | Not deposited[99] | Ile44 | SH3 | 2JT4 | Ile44 |
| NZF | 1Q5W | Ile44 | UBCH5 (E2) | 2FUH | Ile44 |
| Znf_A20 | 2C7N | Asp58 | VHS | 3LDZ | Ile44 |
| UEV | 1S1Q | Ile44/ Gln62 | MPN/JAB1 (DUB) | 2ZNV | Leu71 |
| UBP/PAZ (DUB) | 2G45 | C-terminus | Josephin (DUB) | 2JRI | Ile44 |

The formation of ternary complexes is dependent on UBDs interacting at different positions on the surface of Ub, however despite being a prerequisite for ternary complex formation two separate binding patches does not appear to be sufficient and other factors appear to play a key role. The size of the UBD binding at a given site appears to be a key determinant; for example the long helix of the MIU domain, which binds to the canonical Ile44-centred binding site, appears unlikely to form a ternary complex with the UBP domain, despite binding at distinct binding sites. In this case the length of the MIU helix causes significant clashes with the large UBP domain which would prevent ternary complex formation. Much smaller Ile44 binders, such as the NZF or SH3 domains, have no such clashes and would be predicted to form ternary complexes with UBP domains.

Interestingly the previously mentioned ESCRT system, which is thought to involve the passing of ubiquitinated cargo from one protein to the next, seems to utilise Ile44 binding UBDs including UIM, UEV, and GLUE domains, suggesting that in this case the 'hand-off' mechanism is unlikely and that competitive displacement is the likely mechanism[250].

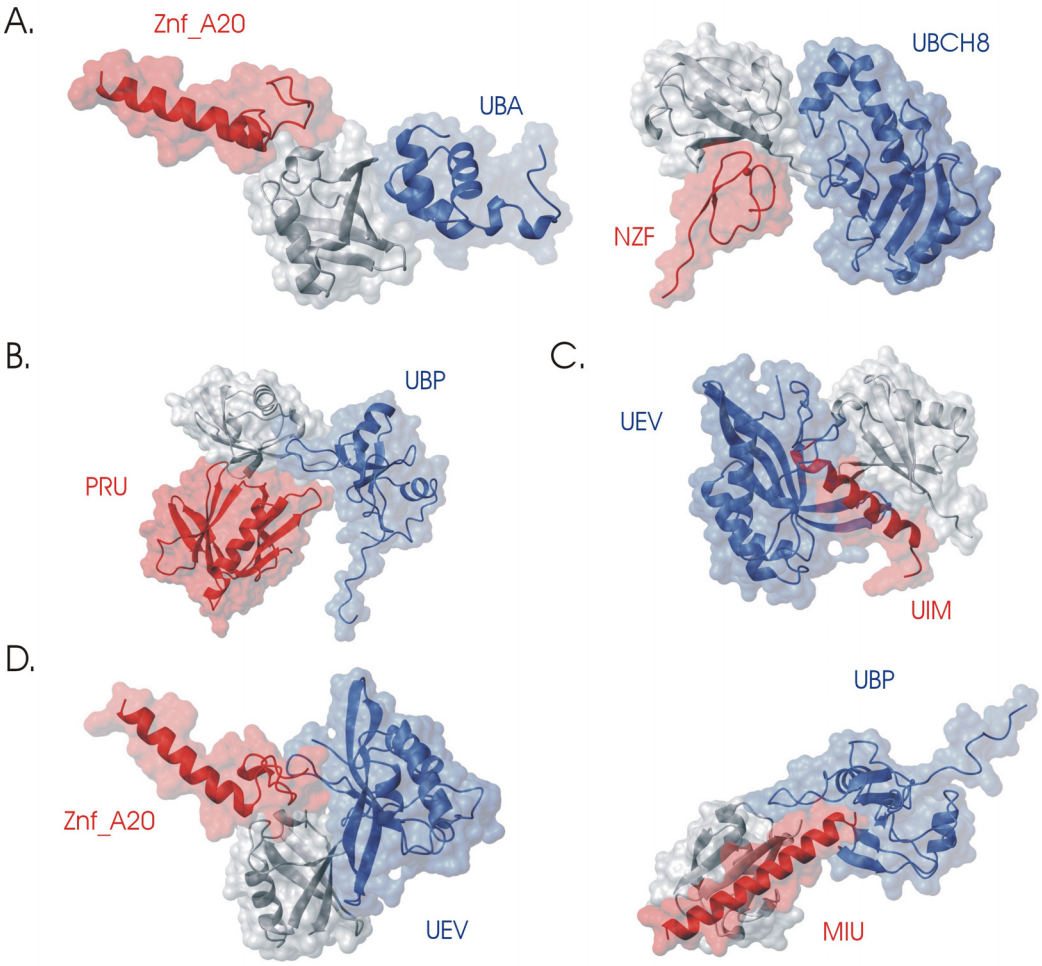



**Figure 9-2: Representative ternary complex models deemed; (A) Highly likely (B) Likely (C) Highly unlikely (D) or unlikely. Classifications are based on steric clashes.**

From examination of the 22 UBD families for which there are binary complexes in the RPDB, it becomes apparent that despite the discovery in recent years of UBDs which bind away from the classical hydrophobic binding patch, a large portion of the Ub surface remains untouched by UBDs including the α-helix and the so called α-β groove which is so often used in UBL interactions with their

binding partners (Figure 9-3)[14]. The high conservation of the Ub primary sequence, including its surface residues, suggests a degree of evolutionary pressure to remain invariant, this is particularly interesting for surface residues, which contribute little to the overall stability of the protein[14,253]. This conservation perhaps suggests that UBDs which employ these unutilised regions may exist but have yet to be discovered. It is worth noting that many of the lysines used in the formation of poly-Ub chains reside in these untouched regions potentially as a means to keep them accessible. The function of many of these lysines, remains unclear with the more common Lys11, Lys48, and Lys63 residing in close proximity to the UBD interactions sites.

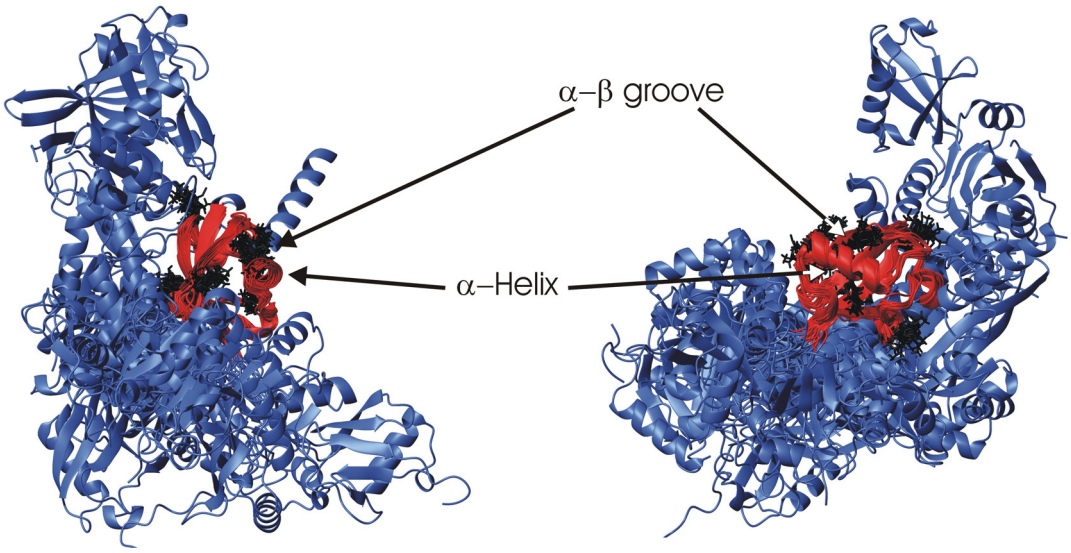



**Figure 9-3: Overlay of all 22 UBD complexes UBDs are coloured blue, Ub in red, Ub's lysine residues are highlighted in black.**

### 9.2.2 Higher Order Complexes with Poly-Ub

The formation of ternary complexes on mono-Ub could have potential implications for poly-Ub; the binding of the Znf_A20 domain to poly-Ub can be affected by the conformation of the poly-Ub chain, with the Znf_A20 seemingly preferring the closed conformation of K48-linked poly-Ub (Section 4.3.12). The influence of chain link specific interactions of some UBDs with poly-Ub has the potential to influence poly-Ub chain conformation, which may have knock on effects on the formation of ternary complexes.

In order to investigate the effects of UBD-bound poly-Ub on ternary complex formation, the Znf_A20 domain (RPDB ID: 2C7N) was analysed for its ability to

interact with UBD bound poly-Ub chains from the RPDB. The Znf_A20 was chosen for its proposed ability to form ternary complexes with most other UBDs, and its relevance to the study presented here.

A number of structures of free poly-Ub chains have been published to date including Lys48, Lys63, Lys11, Lys6, and linear di-Ub, the Znf_A20 would be expected to bind all poly-Ub chain linkage types when free in solution (Section 4.3.12). A number of poly-Ub complexes with UBDs are available including Lys48-linked di-Ub in complex with UBA, UIM, and Josephin domains (RPDB ID: 1Z06, 2KDE and 2JRI), Lys63-linked di-Ub in complex with UBAN, UIM, NZF, and MPN (3JSV, 3A1Q, 2A9K and 2ZNV) domains, and linear di-Ub in complex with a UBAN domain (2ZVO).

In the majority of cases the binding of the Znf_A20 to either the proximal or the distal Ub is free from steric clashes with the adjacent Ub, their bound UBD, or the Znf_A20 bound to the adjacent Ub permitting simultaneous occupancy of both Ub in the di-Ub molecule (Figure 9-4). However there are two notable exceptions. The binding of Lys63 di-Ub to the NZF domain of TAB2 appears to hold the adjacent Lys63-linked Ub units in such a conformation as to cause steric clashes between Znf_A20 bound to adjacent Ub units. The binding of Znf_A20 to either the proximal or the distal Ub is not hindered in isolation but the simultaneous occupancy of both would appear to be restricted. This in principle could work both ways, whereby the binding of Znf_A20 to both Ub units in a Lys63 linked di-Ub unit could prevent it forming a conformation capable of binding to the NZF domain. The second exception arises from the proximity of the Znf_A20 binding site to Lys48, the UIM domain binds at the Ile44 patch, however it is made up of a single long helix which extends towards the C-terminus of Ub, in the distal-Ub the proximity of the UIM to the C-terminus of Ub brings it close to Lys48 of the proximal Ub. The result is that the UIM bound to the distal Ub causes considerable clashes with the Znf_A20 bound to the proximal Ub. This therefore suggests that UIM-bound poly-Ub would be limited in its ability to bind Znf_A20 domains. This is not seen for Lys63-linked di-Ub in complex with a UIM, where both Ub are free to bind to Znf_A20 simultaneously.

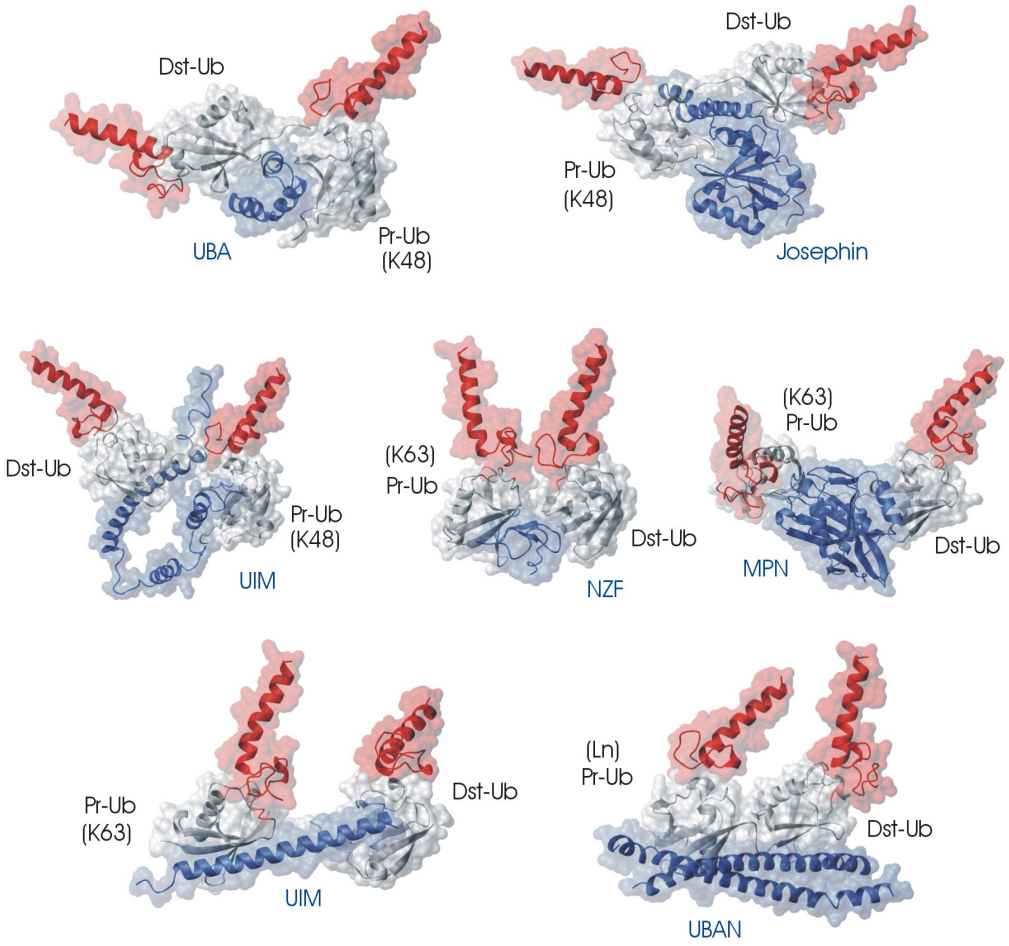



**Figure 9-4: Selection of Di-Ub (Grey) in complex with a variety of UBD (Blue) examined for their ability to accommodate the Znf_A20 (Red). Di-Ub complexes are linked via Lys48 (K48), Lys63 (K63) or the N-terminus (Ln), the proximal (Pr) and distal (Dst) Ub are labelled[16].**

The case of UIM inhibition of Znf_A20 binding in poly-Ub chains, or vice versa, is an interesting one, particularly when considering that the structure used in this investigation is Lys48-di-Ub in complex with the UIM of the proteasomal subunit S5a, which functions as one of the two key Ub-receptors of the proteasome complex. The ZNF216 protein, which binds to Ub via its Znf_A20 domain, functions as a Ub shuttle protein, carrying ubiquitinated proteins to the proteasome[254]. The mechanism of transfer of the ubiquitinated substrate from ZNF216 to the proteasome may well involve both the S5a-UIM and the Znf_A20 of ZNF216, which the examination here suggests may involve both ternary complex formation and displacement of the two UBD types binding to poly-Ub[113,254].

The Znf_A20 appears able to form a ternary complex with the majority of domains bound to mono and di-Ub, this includes those which function as Ub-ligases and DUBs, several reports have suggested that the Znf_A20 domain in some cases may possess ubiquitin ligase activity, however this has been disputed in other reports which suggest that the Znf_A20 may simply act as an adapter for catalytic ligases, the results presented here which suggest co-interaction with Ub ligases and DUBs would be consistent with either hypothesis but particularly the proposed function as an adapter[20,21,32,113,122].

Currently unpublished results from the group of Dr Robert Layfield in the school of Biochemistry at the University of Nottingham have demonstrated that the Znf_A20 domain of ZNF216 can be found associated with free cellular Di-Ub. The Znf_A20 ability to bind to both Ub units in the complexes of di-Ub bound to the Josephin and MPN domains both of which possess DUB activity may well be relevant to this observation. The binding of Ub to both of these domain is incompatible with ternary (or higher order) complex formation with the majority of UBDs (Figure 9-1) suggesting that they would be displaced prior to removal of the poly-Ub chain by the activity of the DUB. The Znf_A20 on the other hand does not need to be removed in order for the DUB to bind leading to free-Ub, or poly-Ub associated with Znf_A20. The biological implication of this Znf_A20-associated poly-Ub chain is unknown but could act to sequester free poly-Ub, or localise it to a desired location. This may be particularly important given the recent observation that free Lys63-linked poly-Ub chains can stimulate autophosphorylation and activation of TAK1 leading to activation of IKK and therefore activation of NF-κB[29]. This activation is mediated by an interaction of the TAB2-NZF domain with the free poly-Ub chain, as shown above there is the possibility of competition between Znf_A20 and NZF for binding to Lys63-linked poly-Ub chains which may be relevant in this case, however this would need considerable further investigation.

### 9.2.3 Summary

Using comparisons of available binary complexes of UBDs with Ub taken from the RPDB, it appears to be highly likely that many UBD pairs have the ability to form ternary complexes with Ub. This represents the first comprehensive investigation into the degree of overlap between UBD binding sites on Ub and has a number of implications for the regulation of Ub mediated signalling. From the

limited number of available UBD complexes with poly-Ub chains there is evidence for the ability of one UBD to modulate the binding of another UBD to the same chain (where ternary complexes to mono Ub are possible). Of the 22 investigated domains the Znf_A20 appears to be the most versatile being deemed highly likely to form ternary complexes with 20 of the 21 other domains.

## 9.3 The Znf_A20 and p62-UBA Form a Ternary Complex with Ub

Comparison of the available Ub complexes with binding partners, predicted the Znf_A20 to form ternary complexes with the majority of UBD families with the only exception being the UEV domain. Although the overall backbone RMSD for Ub across the studied complexes is low at 0.57 Å, the adaptability of ubiquitin is thought to be key to its ability to bind to its large number of structurally distinct binding partners and the interaction of Ub with UBDs may influence this adaptability[159,251,252]. The use of binary complexes therefore only gives an indication of compatible UBDs for ternary complex formation for further examination *in vitro* or *in vivo*.

One of the identified compatible pairs of interest, is that formed between the Znf_A20 domain and the UBA domain, on simple steric grounds these two represent one of the best examples of a pair rated highly likely to form a Ub mediated ternary complex. The two domains are separated by 11 Å at their closest point and have termini which are directed away from the other UBD making steric clashes highly unlikely. However as discussed above other factors may affect the binding of these two domains, both domains make contacts close to the loop between β3 and β4, and the loop between β4 and the $3_{10}$ helix. These loops are two of Ub's variable regions thought to be key to both the conformational selection and induced fit models of Ub recognition[159,251]. The interaction of the UBA and Znf_A20 with this loop may lead to changes in the available conformations at the two binding sites altering the binding abilities. However it is worth noting that the same studies which identify the β3 - β4 loop as variant have also highlighted the $3_{10}$ helix region as one showing relative invariance to structural changes during recognition although none of these studies included the Znf_A20 complex.

Another factor which is hard to anticipate is the effect of the bound UBD on the interfacial water structure at either binding site which have been proposed to play an important role in UBD recognition by Ub, which could lead to unexpected results in ternary complex investigations.

The two proteins studied in this thesis provide an ideal platform on which to test the predicted ternary complex between the Znf_A20 and a UBA domain. Both the Znf_A20 of ZNF216 and the p62-UBA have had their interaction with ubiquitin studied in detail in this thesis and in previous work with a variety of biophysical measurements available for both[61,180]. What's more the ZNF216 and p62 proteins have a clear functional connection suggesting their ability to form a ternary complex may be biologically relevant. Both are known to function as proteasomal shuttling proteins, have been shown to modulate NF-κB activation and the two share a number of binding partners relevant to NF-κB signalling including TRAF6 and RIP[113,116,117,178,181].

### 9.3.1 NMR and Biophysical Analysis

#### 9.3.1.1 Comparison of the Binary Complex Chemical Shifts

The question of whether or not the binding sites for the p62-UBA and the Znf_A20 can be occupied simultaneously is one which is ideally suited to NMR due the sensitivity of the chemical shift to its local environment giving information on a per residue basis, facilitated by a prior knowledge of the chemical shifts of Ub saturated with both the p62-UBA and the Znf_A20 from previous studies (Chapters 4 and 7).

Titrations of both Znf_A20 and the p62-UBA with $^{15}$N-Ub have previously identified those residues of Ub which are affected by the two different binding events monitored by changes in chemical shift. A comparison of the two chemical shift patterns show that significant chemical shift perturbations are seen for different sets of residues for the two UBDs (Figure 9-5). However a subset of residues with significant perturbations is common between the two binding events, these residues include residues Lys48 to Gly53 forming the β4-$3_{10}$ loop (Figure 9-5).

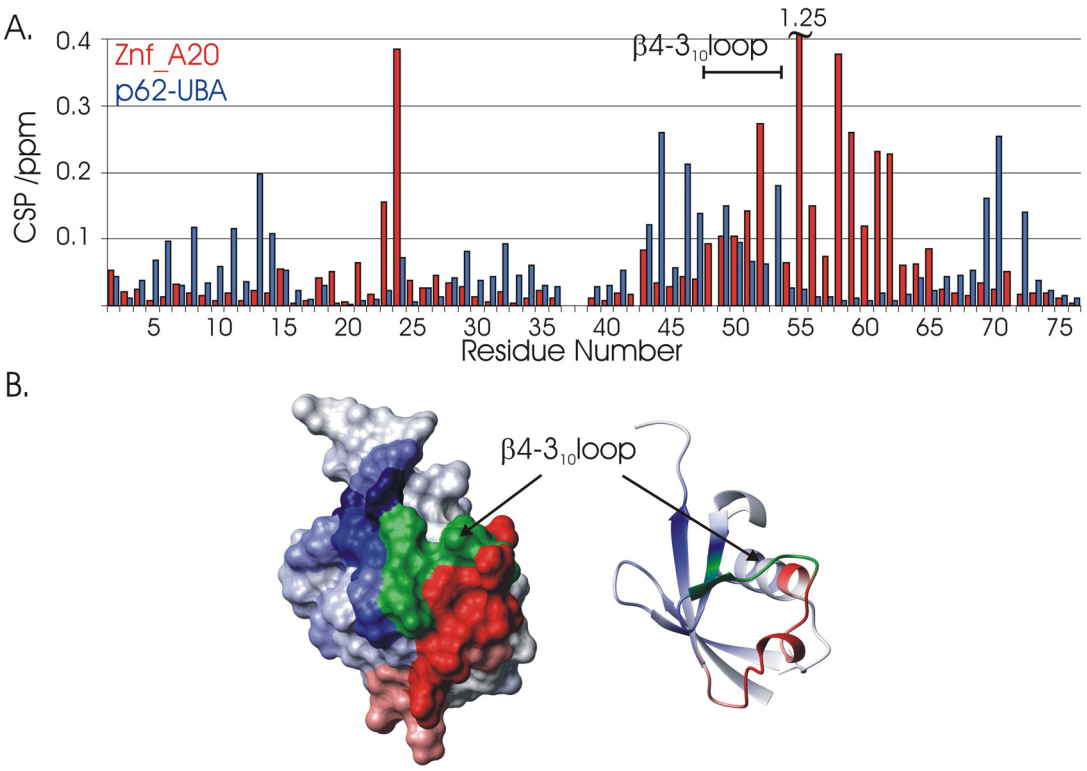



**Figure 9-5: (A) Residue specific CSP for $^{15}$N-Ub bound to Znf_A20 (Red) or the p62-UBA (Blue). (B) The two CSPs were mapped onto the UB structure (RPDB ID: 1UBQ), overlapping shifts are coloured green.**

The p62-UBA causes significant CSPs over a large residue range consistent with the concept of small structural changes across the whole β-sheet surface as seen for other Ile44 UBDs[61,159,251]. By comparison the binding to the Znf_A20 causes relatively localised CSP consistent with the observation that the $3_{10}$ helix region is fairly stable and relatively inflexible. The observation of common CSPs across the β4 - $3_{10}$ helix loop is consistent with the previous suggestion that both domains contact this region of Ub, and therefore that binding at one site may lead to conformational changes which affect the other binding site, highlighting the need for experimental validation of ternary complex formation.

#### 9.3.1.2 Ternary Complex NMR Titration

The intermediate exchange line broadening seen during the titration of Znf_A20 with Ub could potentially mask subtle fast exchange effects from the p62-UBA interaction if the titration is performed by first saturating $^{15}$N-Ub with p62-UBA followed by titrating in Znf_A20. To avoid this 1 mM $^{15}$N-Ub was first saturated with 2 mM Znf_A20 followed by step wise addition of 0.5 equivalents of

p62-UBA to a total of 4 mM p62-UBA to monitor the binding at the two sites (Figure 9-6).

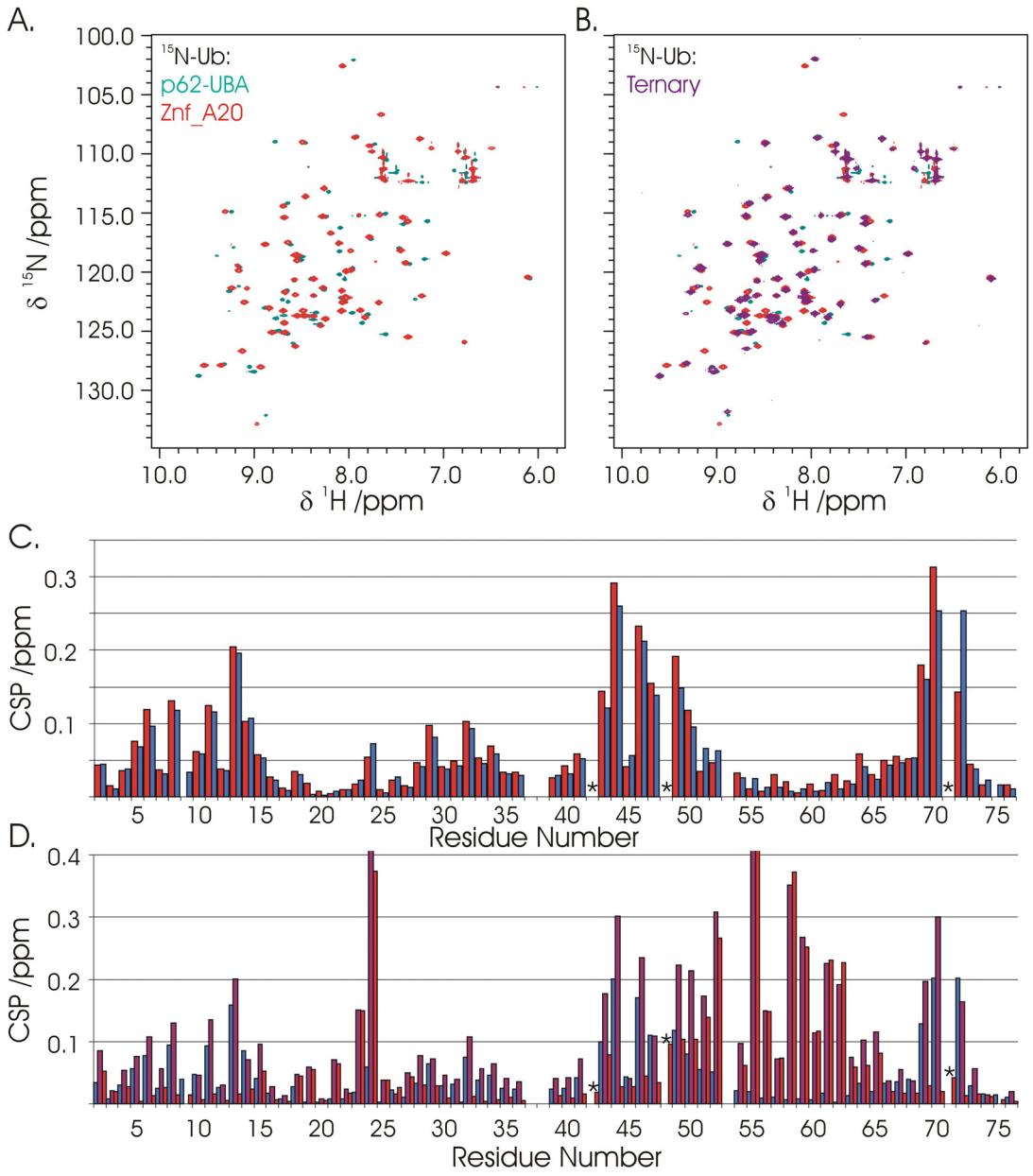



**Figure 9-6: (A-B) $^{1}$H-$^{15}$N-HSQC of 1mM $^{15}$N-Ub with 2 mM Znf_A20 (Red), 4 mM p62-UBA (Blue) or Both (Purple). (C) Residue specific CSP for the addition of saturating quantities of p62-UBA to $^{15}$N-Ub free (Blue) or saturated with Znf_A20 (Red). (D) Residue specific CSP for $^{15}$N-Ub with saturating amounts of Znf_A20 (Red), p62-UBA (Blue) or Both (Purple).**

The addition of p62-UBA to $^{15}$N-Ub saturated with Znf_A20 showed concentration dependent chemical shift changes consistent with a binding event in

fast exchange on the NMR time scale. The line broadening for Arg42, Lys48, and Leu71 evident for the binary titration was also seen. The magnitude of the chemical shift changes upon addition of p62-UBA were consistent with those seen for the addition of p62-UBA to free $^{15}$N-Ub, suggesting unperturbed binding by the p62-UBA.The final CSPs compared to free $^{15}$N-Ub is well represented by the sum of the CSPs for $^{15}$N-Ub binding to the Znf_A20 and the p62-UBA (Figure 9-6).

Due to the sensitivity of the chemical shift the binding of p62-UBA and Znf_A20 both cause at least small CSPs to the majority of residues, which may complicate the assigning of the two sites as saturated as the two could be in fast exchange, resulting in bound-like chemical shifts without simultaneous binding. However by taking subsets of resonances originating from the two binding sites which only show CSPs during titrations with one UBD but not the other one can assess the degree of binding to the two sites with minimal effects from binding to the other site. For the p62-UBA binding site residues Ile13 and Val70 were selected and for the Znf_A20 binding site Asp58 and Asn62, all of which show large CSP during binding at their relevant binding sites but very little or no CSP for binding to the other. From comparison of the chemical shifts of these residues in $^{15}$N-Ub saturated with Znf_A20, p62-UBA or both it can be seen that the chemical shifts of the ternary complex titration are consistent with close to 100% saturation. In order for this to be the case both sites must be occupied simultaneously showing that ternary complex formation can be unambiguously detected between the Znf_A20, Ub, and the p62-UBA.

This is supported by the lack of any exchange broadening of $^{15}$N-Ub peaks indicative of Znf_A20 dissociation, with the majority of Znf_A20 binding site peaks unaffected by the titration of p62-UBA. There was however a moderate overall increase in line width for the ternary complex spectrum, however this is to be expected given the increase in the mass of the complex from around 15 kDa to around 20 kDa with the addition of the p62-UBA. It is also worth noting that by the titration end the total sample protein concentration equalled 7 mM (~45 mg/ml) resulting in increased viscosity which may contribute to line broadening due to a increased $\tau_c$.

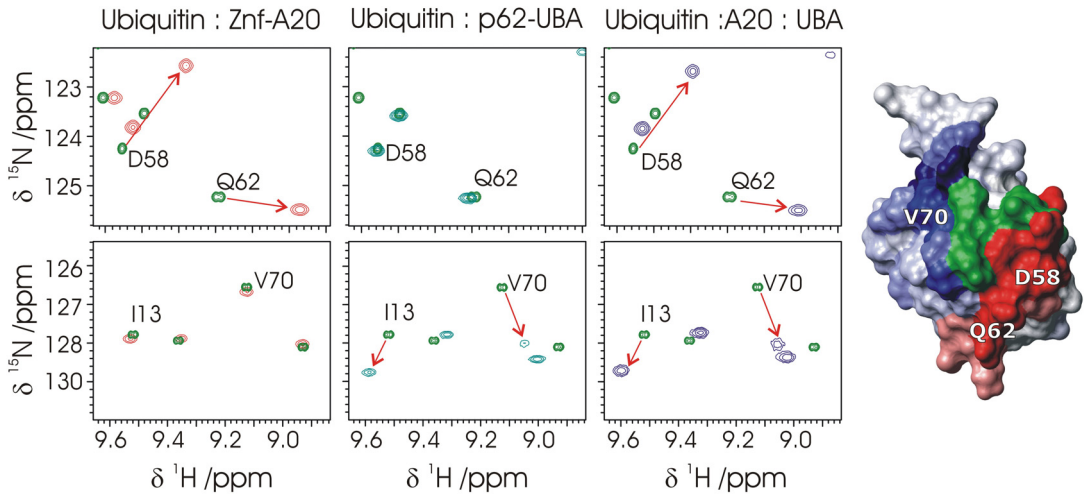



**Figure 9-7: Section of $^{1}$H-$^{15}$N-HSQC spectra containing resonances belonging to Ile13 and Val70 representing the p62-UBA binding patch, and Asp58 and Asn62 for the Znf_A20 binding site. Spectra are shown for free Ub (Green) and with saturating amounts of Znf_A20 (Red), p62-UBA (Blue) or Both (Purple). Large CSPs for binding to the Znf_A20 (Red), p62-UBA (Blue) and Both (Green) are mapped onto the structure Ub.**

The concentration dependent chemical shift changes seen for the p62-UBA domain can be used to determine an apparent $K_d$ for the p62-UBA interaction with Ub in the presence of saturating amounts of Znf_A20 (Figure 9-8). The apparent $K_d$ for the p62-UBA interaction with Ub in the presence of saturating amounts of Znf_A20 equals 800 ± 200 µM. This is within error of that recorded for the binary interaction between Ub and the p62-UBA which determined a $K_d$ of 740 ± 80 µM (Section 7.3). The observed concentration-dependent chemical shift changes can be fitted to calculate the p62-UBA monomer affinity for Ub using a competitive dimer model and the known p62-UBA dimer affinity using the program DynaFit[59,60] as described for the binary interaction between Ub and the p62-UBA. Due to the relatively poor quality of the data and only limited titration points, the fits achieved using DynaFit were poorly defined, however a $K_d$ for the Ub:Znf_A20 complex interaction with the p62-UBA was determined with a value of 40 ± 15 µM within error of the 42 ± 3 µM recorded for the binary complex affinity.

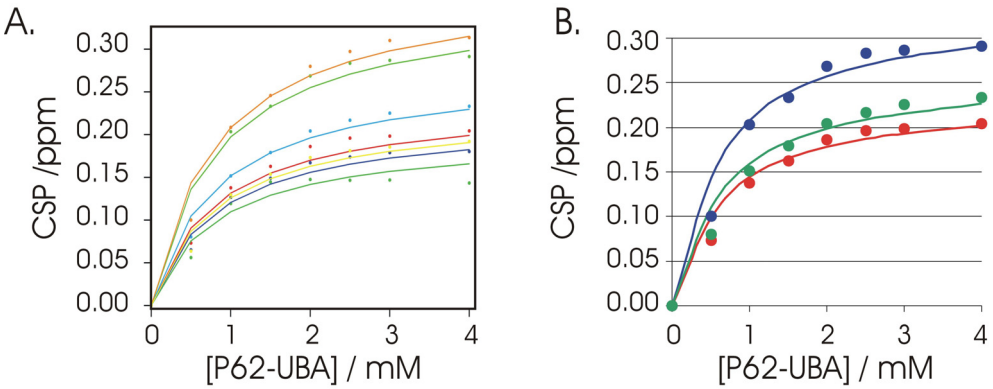



**Figure 9-8: CSP as a function of concentration for the titration of p62-UBA with Ub, residues with large CSP were globally fit to; (A) a one to one binding model in IgorPro™ or (B) a competitive binding model using DynaFit[59,60].**

In order to assess the potential for direct contact between the p62-UBA and the Znf_A20, p62-UBA was titrated up to 2 mM into 1 mM $^{15}$N-Znf_A20 saturated with 2 mM Ub. No CSPs were detected giving a strong indication that no inter-UBD contacts occur as a result of ternary complex formation, the complete lack of measurable CSP may also indicate that there are no measurable changes to the surface of Ub in the Znf_A20 binding patch as this would be expected to cause subtle knock-on effects on the environment of the interface residues of the bound Znf_A20. The only observable difference in $^{1}$H-$^{15}$N-HSQC of Znf_A20 upon the addition of p62-UBA was the same symmetrical line broadening seen in the $^{15}$N-Ub spectrum, consistent with the formation of a ternary complex resulting in an increase in complex mass, as mentioned above this may also be contributed to by an increase in viscosity due to high protein concentration.

#### 9.3.1.3 Mass Spectrometry Examination

In order to confirm that the formation of a ternary complex between the Znf_A20, Ub, and the p62-UBA domain was not an artefact of the high concentrations used in the NMR experiments, native ESI-MS was recorded on a mix of Ub, Znf_A20, and p62-UBA in an attempt to detect mass ions originating from the ternary complex. ESI- MS spectrum were recorded on samples containing Ub, Znf_A20 and p62-UBA mixed in varying ratios with a total protein concentration between 4 and 9 µM. Mass peaks corresponding free Ub, Znf_A20, and p62-UBA were the dominant species in all cases, mass speaks corresponding to the binary

complexes of Ub with Znf_A20 or p62-UBA were also readily detected, mass peaks consistent with the formation of a ternary complex between the three was also visible but with poor signal to noise. This is consistent with the two UBDs having $K_d$ in µM range, giving low populations of binary complex at low µM protein concentration and subsequently low ternary complex. No Mass peaks were seen corresponding to higher order complexes or to binary complexes between the Znf_A20 and the p62-UBA, consistent with the NMR data which shows no interaction between the two UBDs.

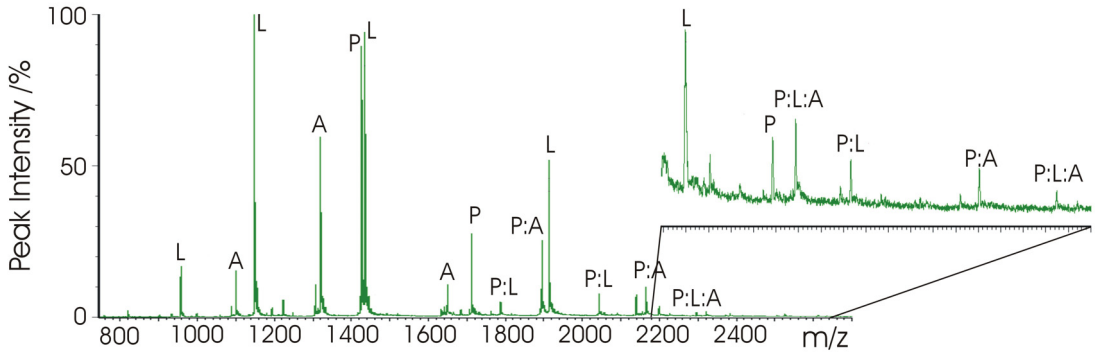



**Figure 9-9: ESI-MS of 1.5 µM Ub (P), 1.5 µM Znf_A20 (A) and 2 µM p62-UBA (L), showing the formation of binary (P:L and P:A) and ternary complex (P:L:A).**

The relative intensities of all species were used to estimate the concentrations of each species from the total protein used. The estimated concentrations of each species were used to estimate the equilibrium constants for the interaction of Znf_A20 and p62-UBA with free Ub and in complex with the other UBD (Figure 9-10 and Table 9-2). Although the estimated $K_d$ for the binding of either UBD to Ub in complex with the other is marginally weaker the intensities of the ternary complex are low which may result in underestimation of the appropriate populations. Regardless of this, the calculated differences in $K_d$ represent only modest changes in binding energy and therefore the estimated equilibrium constants determined from ESI-MS are consistent with the two UBDs binding independent of one another with little or no effect arising from the presence of the other UBD bound to Ub, the measured equilibrium constants are listed in Table 9-2.

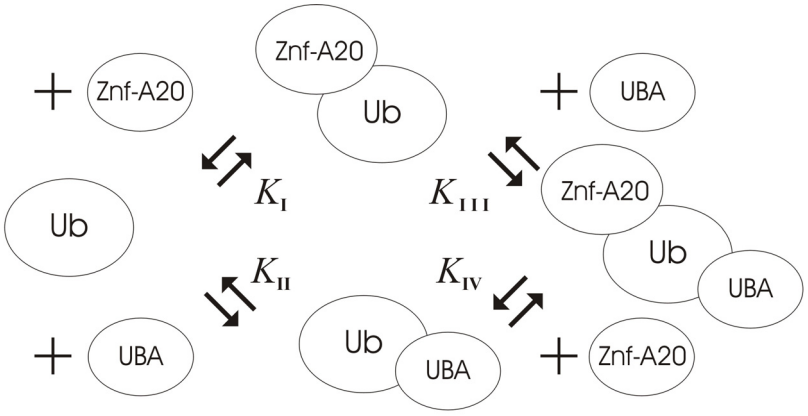



**Figure 9-10: Model of the equilibria associated with ternary complex formation.**

**Table 9-2 Table of equilibrium constants for ternary complex formation measured by ESI-MS.**

| | $K$ (±) / M | $K_d$ (±) /μM |
|---|---|---|
| Znf_A20 Binary ($K_I$) | 250000 (15000) | 4 (2) |
| Znf_A20 Ternary ($K_{IV}$) | 260000 (32000) | 6 (3) |
| p62-UBA Binary ($K_{II}$) | 43500 (1000) | 23 (3) |
| p62-UBA Ternary ($K_{II}$) | 25000 (7000) | 40 (10) |

Given the weak intensity of mass peaks corresponding to the observed ternary complex, MS-MS with collision induced dissociation (CID) was utilised in order to confirm the composition of the observed mass ions. Mass ions corresponding to the proposed ternary complex 9+ ion were isolated using a quadrupole and accelerated through a transfer chamber containing argon gas in order to produce high energy collisions between mass ions and Argon. The acceleration voltage, and therefore the kinetic energy of the collisions with argon, was gradually increased until dissociation of the ternary complex ion was observed (Figure 9-11).

The 9+ ternary complex ion disassociated into the expected component parts, this primarily consisted of dissociation of p62-UBA to leave a complex of Znf_A20 with Ub, although fragments corresponding to free Ub, free Znf_A20, UBA complex with Ub and small quantities of Znf_A20 in complex with p62-UBA were also detected (Figure 9-11). This confirms the expected composition and contact order of this ion. The major fragments correspond to ternary complex splitting to either Znf_A20 in complex with Ub and p62-UBA, or to a lesser extent p62-UBA in

complex with Ub and Znf_A20, this supports the expected complex architecture whereby Znf_A20 and the p62-UBA contact Ub but not each other.

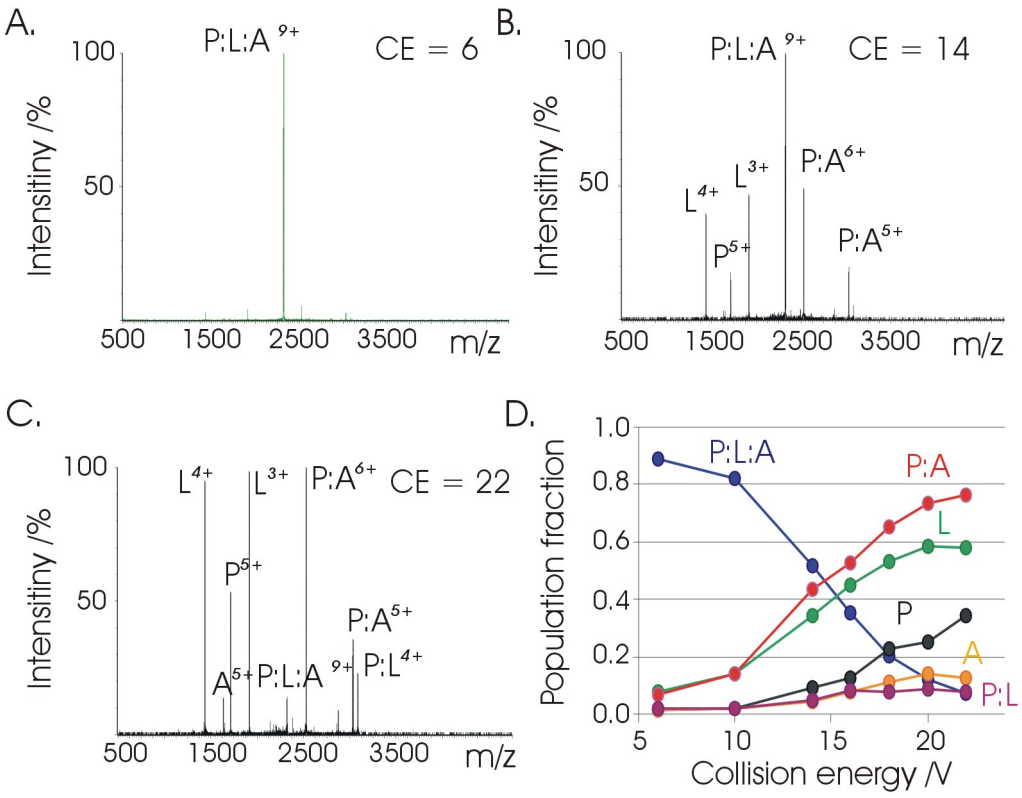



**Figure 9-11: Collision Induced Dissociation of ternary complex 9+ ion using transfer collision energy voltages of (A) 6, (B) 14, and (C) 22 V. (D) Plot of population change with increasing energy. In all cases P = Ub, L = p62-UBA, A = Znf_A20.**

The fragmentation of the ternary complex showed a clear preference for the dissociation of the p62-UBA from the ternary complex, over the dissociation of the Znf_A20 (Figure 9-11). This is consistent with two known aspects of the relative UBD complexes with Ub. The first being the relative affinity; previous reports, and the data provided here, suggest that the Znf_A20 forms a higher affinity complex with Ub than the p62-UBA, which may be reflected in the relative energies required to dissociate the two UBDs in the gas phase by CID. The second key aspect is the gas phase environment and the nature of the two binding sites, the p62-UBA interaction with Ub is stabilised largely by hydrophobic contacts and the hydrophobic effect, in the gas phase the absence of solvent, and therefore loss of the hydrophobic effect, would be expected to cause a reduction in the energy required to dissociate the p62-UBA complex with Ub. On the other hand the Znf_A20

interaction with Ub is mediated by a mix of hydrophobic contacts, hydrogen bonds and potential electrostatic effects, in the gas phase the loss of solvent, and therefore of dielectric constant, results in an accentuation of polar interactions which would be expected to stabilise the Znf_A20 interaction with Ub in the gas phase.

A second interesting observation is that of a small quantity of fragments consisting of Znf_A20 and p62-UBA (Not Shown, population <0.05), this may be the result of small amounts of non-specific association of Znf_A20 or p62-UBA to binary complexes of the other resulting in a ternary complex mass, however there is little evidence for that within the ESI-MS spectrum under the same conditions where no unusual complexes are seen (including no Znf_A20 complexes with p62-UBA). Although it is expected that the populations seen in the ESI-MS spectrum are gas phase representations of the solution populations, the structures may well be radically different to that seen in solution. It is possible that the observation of Znf_A20 fragments with p62-UBA are the result of close proximity in the ternary complex, either in the solution phase and/or the gas phase, or by some long range electrostatic effects in the gas phase bringing the two together.

As observed with the binary complex of Znf_A20 with Ub, CID resulted in the partitioning of the Znf_A20 zinc ion between the Znf_A20 and Ub (Ub-UBA complex) during fragmentation.

#### 9.3.1.4 ITC Titrations

In order to investigate how p62-UBA binding to Ub affects the Ub interaction with Znf_A20, ITC titrations were recorded whereby concentrated Znf_A20 (300 μM) was titrated into either 20 μM Ub or 20 μM Ub in the presence of 400 μM p62-UBA. The free p62-UBA is dimeric in solution, the $K_{dim}$ for the p62-UBA is around 7 μM, and the dilution effects caused by the injection of Znf_A20 in buffer will maintain the free p62-UBA concentration above ~ 300 μM and therefore should not affect the titration. The dissociation of the p62-UBA complex with Ub caused by the dilution may give a small degree of dissociation although the $K_d$ is reported to be around 40-50 μM[61] and the p62-UBA is in 20 fold excess of Ub so any measured dissociation should only contribute a small amount to the overall signal. To account for this an appropriate blank was subtracted from the titration data including buffer injected into 20 μM Ub in the presence of 400 μM p62-UBA.

The titration of Znf_A20 into Ub in the presence of 400 μM p62-UBA gave data which was similar to that recorded in the absence of 400 μM p62-UBA, with a $K_d$ of 8 ± 2.5 μM compared to a $K_d$ of 11.5 ± 1.5 μM in the absence of p62-UBA. This is consistent with little or no change in the affinity of the Znf_A20 for Ub in complex with the p62-UBA as found with the NMR and ESI-MS data, suggesting no significant modulation of one binding site by the other.

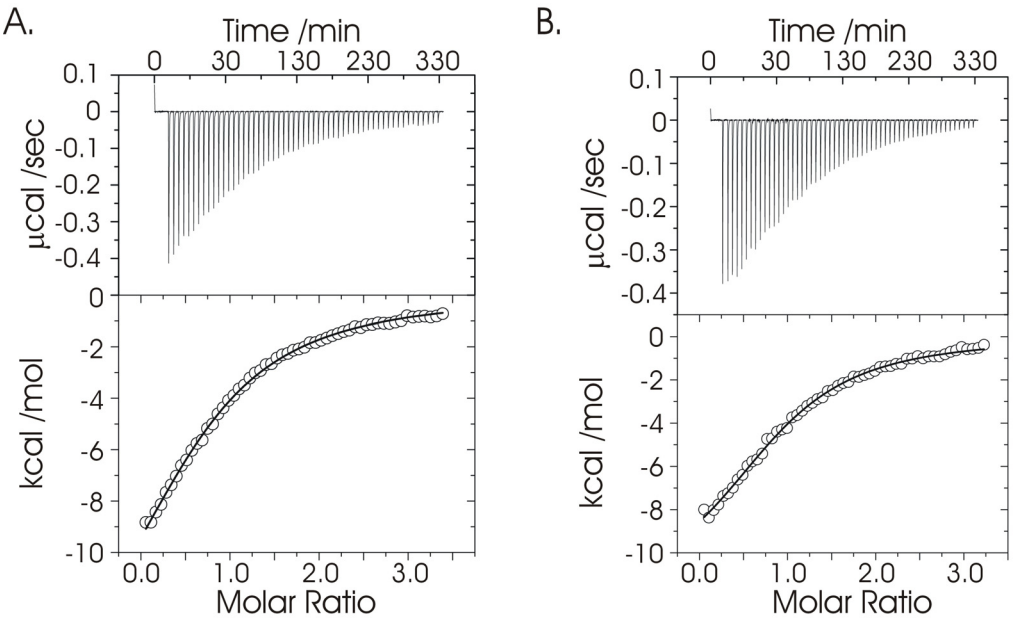



**Figure 9-12: ITC titrations of Znf_A20 into (A) Ub and (B) Ub in the presence of 400 μM p62-UBA.**

The only clear difference between the two ITC titrations is an increase in the recorded noise in the titration in the presence of p62-UBA; this is potentially as a result of viscosity changes due to the relatively high protein concentration or because of complications of overlapping binding and dissociation responses.

### 9.3.2 Biochemical Investigation

The biophysical data presented above has demonstrated that the formation of a ternary complex is possible between the p62-UBA and the Znf_A20 of ZNF216 *in vitro*. In order to gain insight into the possibility of such a complex forming in a cellular environment the interaction was investigated using standard biochemistry approaches by Miss Joanna Strachan in the lab of Dr Robert Layfield.

Immobilised recombinant full length ZNF216 was used to precipitate ubiquitinated proteins from U20S cells. The Cys30Ala/Cys33Ala double-mutant has

been shown to be defective in Ub-binding[113] and was used as a control which was unable to capture ubiquitinated proteins (Figure 9-13). The binding of ubiquitinated proteins were confirmed using SDS-PAGE. The lack of protein binding to the Ub binding impaired mutant serves to confirm that the bound proteins are ubiquitinated proteins rather than those interacting directly with regions of the ZNF216 protein. Western blots were used to detect the presence of p62 which had co-purified with ubiquitinated proteins. Western blotting confirmed that p62 was co-precipitated only with the wild type ZNF216, indicating that a Ub-mediated interaction between the two proteins is possible in this model (Figure 9-13).

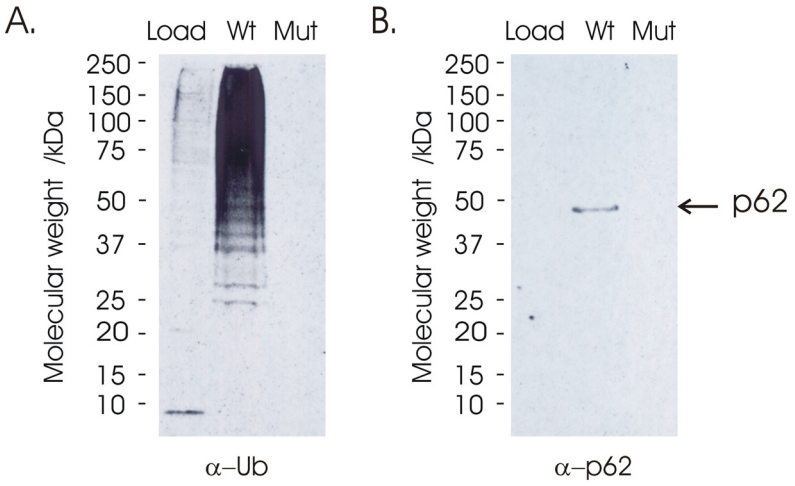



**Figure 9-13: Western blots of pull down experiments of ubiquitinated proteins and p62 from U20S cells using column bound ZNF216 and the C30A/C33A double mutant. Proteins were visualised using either (A) rabbit anti-Ub antibodies or (B) mouse monoclonal anti-p62.**

However it is highly likely that the ubiquitinated proteins isolated by the immobilised ZNF216 contain poly-Ub chains, these pull down assays can not distinguish whether the p62 and ZNF216 are bound to the same Ub moiety or bound to different Ub units on a poly-Ub chain, or different poly-Ub chains attached to the same target protein. However the co-purification of p62 and ZNF216 using potentially natural substrates from a relevant cell line[116,255], combined with the strong functional links between ZNF216 and p62 give weight to the possibility of a biologically relevant ternary complex between p62, ZNF216, and Ub.

### 9.3.3 Potential Implications of Ub Mediated Ternary Complexes

The data presented here adds to the growing understanding of the intricate details of the molecular basis of ubiquitin mediated signalling. The formation of Ub-mediated ternary complexes with UBDs may have many implications and opens up a number of potential mechanisms by which Ub mediated signalling can be interpreted by UBDs.

The formation of a ternary complex between different Ub-receptors containing different UBDs provides a mechanism by which ubiquitinated proteins can be passed from one protein to the next, processes which must occur in many Ub mediated processes including proteasomal shuttling and the ESCRT machinery[2,15,250]. The inability of two UBDs to form a ternary complex is undoubtedly equally as important as those with the ability to form a ternary complex and competitive displacement clearly plays a role in the mechanism of passing Ub from one receptor to another in many cases. However the 'hand-off' mechanism involving Ub mediated ternary complexes has been demonstrated in this study as a viable process and may prove to be a key process in Ub-mediated cellular processes.

The co-localisation of UBDs via Ub may provide a mechanism for the localisation of two proteins which have no other mechanism of association to the same complex machinery, giving rise to potentially functional cooperativity. Alternatively two proteins associated with the same function, such as shuttling to the proteasome in the case of ZNF216 and p62, which can interact with Ub simultaneously, may afford better specificity for a given function. Given the promiscuity of the canonical Ile44 hydrophobic binding patch, the combination of an Ile44 binder with a second UBD, binding to a site with less potential binders may well be key to specificity. Alternatively the binding of a second Ub-receptor to a ubiquitinated protein bound to a proteasomal shuttling protein may alter the rate of shuttling to the proteasome or of proteasomal association and subsequent degradation, providing an extra method for regulation of proteasomal function. This may generate a mechanism for two phase signalling responses or negative feed back in Ub signalling; the overexpression of a Ub-receptor leading to a high concentrations of the binary complex may in turn lead to the formation of a ternary complex which could be recognised by different proteins leading to different responses or down regulation of the original process.

One of the key features attributed to UBD recognition of poly-Ub over different poly-Ub linkages or mono-Ub is the formation of a new surface formed by the joining of two Ub units, similarly the formation of binary or ternary complex between Ub and UBDs may generate novel binding surfaces for recognition by other proteins to form large multi-protein complexes using multiple UBDs.

The suggestion that the conformation of poly-Ub chains could be modulated by the binding of a second UBD to reduce or increase the affinity of binding to a second UBD could prove particularly interesting, providing a potential mechanism for increased specificity for chain linkage types or length. Alternatively a situation may arise whereby the increased expression of a Ub-receptor containing a certain type of UBD may lead to changes in the fate of ubiquitinated protein by reducing the affinity for one Ub-receptor and increasing the affinity of another.

The ability of Ub to accommodate two UBDs may also have implications for the proposed ubiquitin recognition mechanism. The observation that the affinities of the two UBDs for Ub appear unaltered by the binding of the other UBD suggests that the binding of a UBA to Ub did not prevent Ub from adapting to the Znf_A20 or vice versa. This suggests that binding either UBD does not lead to a 'conformationally locked' molecule and it therefore follows that Ub bound to UBDs is still free to sample conformational space.

## 9.4 Conclusions

The vast majority of UBDs bind to Ub at the classical Ile44 centred binding patch, however in recent years the discovery of a number of alternative binding sites on the surface of Ub has led to the suggestion that Ub may not be limited to binary interactions with UBDs and that multiple UBDs could, in theory, bind to Ub simultaneously to form ternary complexes, adding further complexity to the already complicated system of ubiquitin mediated processes[2].

The first comprehensive comparison of the complexes of Ub with the known UBD families has revealed that a relatively small number of Ub mediated UBD ternary complexes is possible, but that all of the examined UBDs appear to be compatible with the formation of at least one ternary complex.

Of the UBD families investigated the Znf_A20 domain appears to be the most diverse in terms of potential ternary complex partners. The Znf_A20 domain binds to a binding patch centred on Asp58 and is the only protein (to date) known to

bind at this binding site. This region of Ub is reported to be relatively invariant to structural changes during binding events at the more promiscuous Ile44 site which may facilitate the formation of ternary complexes between Znf_A20 and an Ile44 binder[159,251].

The formation of a ternary complex between a Znf_A20 and an Ile44 binding UBD has been detected previously in the complex between Ub and Rabex5 where the Znf_A20 domain and MIU domains of two separate Rabex5 molecules are seen to interact with the same Ub molecule in the crystal lattice. The study presented here confirmed the formation of a second of the predicted UBD ternary complexes, using the Znf_A20 domain of ZNF216 and the UBA domain of p62, using a combination of NMR spectroscopy and biophysical approaches. The UBA domain, like the MIU domain, binds to the Ile44 centred binding patch of Ub, however the two domains are structurally different and bind in subtly different ways to the Ile44 hydrophobic patch (Figure 9-14).

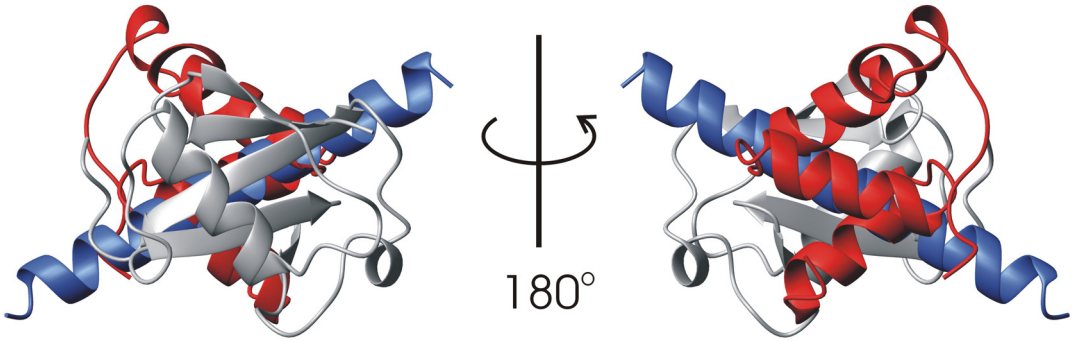



**Figure 9-14: Comparison of the UBA (Red) and MIU (Blue) domains binding to the hydrophobic patch of Ub (Grey).**

The Znf_A20 domain and p62-UBA domain appear to form a ternary complex with no significant effects on the affinity at either binding site, consistent with the prediction that the two domains interact with distinct regions of Ub and are not in close proximity to one another in the ternary complex.

The majority of Ub-receptors have UBDs located at or near to the extreme termini separated by large disordered region[170,171,256], ZNF216 and p62 are no exceptions with NMR and CD data suggesting unstructured regions adjacent to their respective UBDs (Sections 5.3 and 6.5)[170]. Its is considered that such extended

linkers facilitate binding by increasing the diffusional radius of the domain facilitating the so called 'fly-casting' mechanism[172], however these long linkers may also aid the formation of Ub-mediated ternary complexes by reducing steric influence from the rest of the protein during complex formation. It is therefore justifiable to suggest that although the observations made here, using the small isolated domains, would hold true in the full length proteins.

This represents the first study to show a ternary complex formed in solution between two UBDs and Ub. This also represents the first Ub-mediated ternary complex between two distinct proteins with a strong biological link both in terms of function and known binding partners. The formation of ternary complex *in vivo* is still yet to be demonstrated and the potential implications outlined here are merely speculation, however the study here has shown that such complexes are possible and have many potential impacts. As with many aspects of Ub-mediated signalling the *in vitro* experimental observations made here are valuable insights into the molecular processes involved in ubiquitination and Ub recognition which underpin many key cellular proceses[2].

# References


1. Seet,B.T., Dikic,I., Zhou,M.M., and Pawson,T. (2006). Reading protein modifications with interaction domains. *Nat. Rev. Mol. Cell Biol.,* **7,** 473-483.

2. Komander,D. (2009). The emerging complexity of protein ubiquitination. *Biochem. Soc. Trans.,* **37,** 937-953.

3. Bulow,H.E., Tjoe,N., Townley,R.A., Didiano,D., van Kuppevelt,T.H., and Hobert,O. (2008). Extracellular sugar modifications provide instructive and cell-specific information for axon-guidance choices. *Curr. Biol.,* **18,** 1978-1985.

4. Seo,J. and Lee,K.J. (2004). Post-translational modifications and their biological functions: proteomic analysis and systematic approaches. *J. Biochem. Mol. Biol.,* **37,** 35-44.

5. Cohen,P. (2002). The origins of protein phosphorylation. *Nat. Cell Biol.,* **4,** E127-E130.

6. Passmore,L.A. and Barford,D. (2004). Getting into position: the catalytic mechanisms of protein ubiquitylation. *Biochem. J.,* **379,** 513-525.

7. Chen,Z.J. and Sun,L.J. (2009). Nonproteolytic functions of ubiquitin in cell signaling. *Mol. Cell,* **33,** 275-286.

8. Xu,P., Duong,D.M., Seyfried,N.T., Cheng,D., Xie,Y., Robert,J., Rush,J., Hochstrasser,M., Finley,D., and Peng,J. (2009). Quantitative proteomics reveals the function of unconventional ubiquitin chains in proteasomal degradation. *Cell,* **137,** 133-145.

9. Pines,J. and Lindon,C. (2005). Proteolysis: anytime, any place, anywhere? *Nat. Cell Biol.,* **7,** 731-735.

10. Thrower,J.S., Hoffman,L., Rechsteiner,M., and Pickart,C.M. (2000). Recognition of the polyubiquitin proteolytic signal. *EMBO J.,* **19,** 94-102.

11. Finley,D. (2009). Recognition and processing of ubiquitin-protein conjugates by the proteasome. *Annu. Rev. Biochem.,* **78,** 477-513.

12. Iwai,K. and Tokunaga,F. (2009). Linear polyubiquitination: a new regulator of NF-κB activation. *EMBO Rep.,* **10,** 706-713.

13. Hagai,T. and Levy,Y. (2010). Ubiquitin not only serves as a tag but also assists degradation by inducing protein unfolding. *Proc. Natl. Acad. Sci. U. S. A.*, **107**, 2001-2006.

14. Winget,J.M. and Mayor,T. (2010). The diversity of ubiquitin recognition: hot spots and varied specificity. *Mol. Cell,* **38,** 627-635.

15. Harper,J.W. and Schulman,B.A. (2006). Structural complexity in ubiquitin recognition. *Cell,* **124,** 1133-1136.

16. Dikic,I., Wakatsuki,S., and Walters,K.J. (2009). Ubiquitin-binding domains - from structures to functions. *Nat. Rev. Mol. Cell Biol.,* **10,** 659-671.

17. Hurley,J.H., Lee,S., and Prag,G. (2006). Ubiquitin-binding domains. *Biochem. J.,* **399,** 361-372.

18. Hofmann,K. and Bucher,P. (1996). The UBA domain: a sequence motif present in multiple enzyme classes of the ubiquitination pathway. *Trends Biochem. Sci.,* **21,** 172-173.

19. Bomar,M.G., D'Souza,S., Bienko,M., Dikic,I., Walker,G.C., and Zhou,P. (2010). Unconventional ubiquitin recognition by the ubiquitin-binding motif within the Y family DNA polymerases ί and Rev1. *Mol. Cell,* **37,** 408-417.

20. Penengo,L., Mapelli,M., Murachelli,A.G., Confalonieri,S., Magri,L., Musacchio,A., Di Fiore,P.P., Polo,S., and Schneider,T.R. (2006). Crystal structure of the ubiquitin binding domains of rabex-5 reveals two modes of interaction with ubiquitin. *Cell,* **124,** 1183-1195.

21. Lee,S., Tsai,Y.C., Mattera,R., Smith,W.J., Kostelansky,M.S., Weissman,A.M., Bonifacino,J.S., and Hurley,J.H. (2006). Structural basis for ubiquitin recognition and autoubiquitination by Rabex-5. *Nat. Struct. Mol. Biol.,* **13,** 264-271.

22. Trempe,J.F., Brown,N.R., Lowe,E.D., Gordon,C., Campbell,I.D., Noble,M.E., and Endicott,J.A. (2005). Mechanism of Lys48-linked polyubiquitin chain recognition by the Mud1 UBA domain. *EMBO J.,* **24,** 3178-3189.

23. Varadan,R., Assfalg,M., Raasi,S., Pickart,C., and Fushman,D. (2005). Structural determinants for selective recognition of a Lys48-linked polyubiquitin chain by a UBA domain. *Mol. Cell,* **18,** 687-698.

24. Kulathu,Y., Akutsu,M., Bremm,A., Hofmann,K., and Komander,D. (2009). Two-sided ubiquitin binding explains specificity of the TAB2 NZF domain. *Nat. Struct. Mol. Biol.,* **16,** 1328-1330.

25. Komander,D., Reyes-Turcu,F., Licchesi,J.D., Odenwaelder,P., Wilkinson,K.D., and Barford,D. (2009). Molecular discrimination of structurally equivalent Lys 63-linked and linear polyubiquitin chains. *EMBO Rep.,* **10,** 466-473.

26. Rahighi,S., Ikeda,F., Kawasaki,M., Akutsu,M., Suzuki,N., Kato,R., Kensche,T., Uejima,T., Bloor,S., Komander,D., Randow,F., Wakatsuki,S., and Dikic,I. (2009). Specific recognition of linear ubiquitin chains by NEMO is important for NF-κB activation. *Cell,* **136,** 1098-1109.

27. Chen,Z.J. (2005). Ubiquitin signalling in the NF-κB pathway. *Nat. Cell Biol.,* **7,** 758-765.

28. Dikic,I. and Dotsch,V. (2009). Ubiquitin linkages make a difference. *Nat. Struct. Mol. Biol.,* **16,** 1209-1210.

29. Xia,Z.P., Sun,L., Chen,X., Pineda,G., Jiang,X., Adhikari,A., Zeng,W., and Chen,Z.J. (2009). Direct activation of protein kinases by unanchored polyubiquitin chains. *Nature,* **461,** 114-119.

30. Enesa,K., Zakkar,M., Chaudhury,H., Luong,l.A., Rawlinson,L., Mason,J.C., Haskard,D.O., Dean,J.L., and Evans,P.C. (2008). NF-κB suppression by the deubiquitinating enzyme Cezanne: a novel negative feedback loop in pro-inflammatory signaling. *J. Biol. Chem.,* **283,** 7036-7045.

31. Jin,W., Chang,M., Paul,E.M., Babu,G., Lee,A.J., Reiley,W., Wright,A., Zhang,M., You,J., and Sun,S.C. (2008). Deubiquitinating enzyme CYLD negatively regulates RANK signaling and osteoclastogenesis in mice. *J. Clin. Invest,* **118,** 1858-1866.

32. Wertz,I.E., O'Rourke,K.M., Zhou,H., Eby,M., Aravind,L., Seshagiri,S., Wu,P., Wiesmann,C., Baker,R., Boone,D.L., Ma,A., Koonin,E.V., and Dixit,V.M. (2004). De-ubiquitination and ubiquitin ligase domains of A20 downregulate NF-κB signalling. *Nature,* **430,** 694-699.

33. Skaug,B., Jiang,X., and Chen,Z.J. (2009). The role of ubiquitin in NF-κB regulatory pathways. *Annu. Rev. Biochem.,* **78,** 769-796.

34. Platt, G. W. Protein Engineering Studies of the B-Hairpin of Ubiquitin. 2003. University of Nottingham.
Ref Type: Thesis/Dissertation

35. Chen, P. Probing the Dynamics and Folding of the p62-UBA and NBR1 PB1 Domains. 2009.
Ref Type: Thesis/Dissertation

36. Gallagher, T. R. The Structure and Dynamics of the p62-UBA Domain. 2006. University of Nottingham.
Ref Type: Thesis/Dissertation

37. Vogelstein,B. and Gillespie,D. (1979). Preparative and analytical purification of DNA from agarose. *Proc. Natl. Acad. Sci. U. S. A,* **76,** 615-619.

38. Engebrecht,J., Heilig,J.S., and Brent,R. (2001). Preparation of bacterial plasmid DNA. *Curr. Protoc. Immunol.,* **Chapter 10.**

39. Siegel,L. and Monty,K. (1966). Determination of molecular weights and frictional ratios of proteins in impure systems by use of gel filtration and density gradient centrifugation. Application to crude preparations of sulfite and hydroxylamine reductases. *Biochim. Biophys. Acta,* **112,** 346-362.

40. Gill,S.C. and von Hippel,P.H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.,* **182,** 319-326.

41. Greenfield,N.J. (2006). Using circular dichroism spectra to estimate protein secondary structure. *Nat. Protoc.,* **1,** 2876-2890.

42. Greenfield,N.J. (2006). Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nat. Protoc.,* **1,** 2527-2535.

43. Heinz,U., Kiefer,M., Tholey,A., and Adolph,H.W. (2005). On the competition for available zinc. *J. Biol. Chem.,* **280,** 3197-3207.

44. Greenfield,N.J. (2006). Analysis of the kinetics of folding of proteins and peptides using circular dichroism. *Nat. Protoc.,* **1,** 2891-2899.

45. Veros,C.T. and Oldham,N.J. (2007). Quantitative determination of lysozyme-ligand binding in the solution and gas phases by electrospray ionisation mass spectrometry. *Rapid Commun. Mass Spectrom.,* **21,** 3505-3510.

46. Kaltashov,I.A. and Mohimen,A. (2005). Estimates of protein surface areas in solution by electrospray ionization mass spectrometry. *Anal. Chem.,* **77,** 5370-5379.

47. Rosu,F., De Pauw,E., and Gabelica,V. (2008). Electrospray mass spectrometry to study drug-nucleic acids interactions. *Biochimie,* **90,** 1074-1087.

48. Hernandez,H., Dziembowski,A., Taverner,T., Seraphin,B., and Robinson,C.V. (2006). Subunit architecture of multimeric complexes isolated directly from cells. *EMBO Rep.,* **7,** 605-610.

49. Scott,D.J., Harding,S.E., and Rowe,A.J. (2005). *Ultracentrifugation: Techniques and Methods.* Royal Society of Chemistry.

50. le Maire,M., Arnou,B., Olesen,C., Georgin,D., Ebel,C., and Moller,J.V. (2008). Gel chromatography and analytical ultracentrifugation to determine the extent of detergent binding and aggregation, and Stokes radius of membrane proteins using sarcoplasmic reticulum $Ca^{2+}$-ATPase as an example. *Nat. Protoc.,* **3,** 1782-1795.

51. Markley,J.L., Bax,A., Arata,Y., Hilbers,C.W., Kaptein,R., Sykes,B.D., Wright,P.E., and Wuthrich,K. (1998). Recommendations for the presentation of NMR structures of proteins and nucleic acids. *J. Mol. Biol.,* **280,** 933-952.

52. Vranken,W.F., Boucher,W., Stevens,T.J., Fogh,R.H., Pajon,A., Llinas,P., Ulrich,E.L., Markley,J.L., Ionides,J., and Laue,E.D. (2005). The CCPN data

model for NMR spectroscopy: Development of a software pipeline. *Proteins-Structure Function and Bioinformatics,* **59,** 687-696.

53. Piotto,M., Saudek,V., and Sklenar,V. (1992). Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR,* **2,** 661-665.

54. Hwang,T.L. and Shaka,A.J. (1995). Water Suppression That Works. Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients. *Journal of Magnetic Resonance, Series A,* **112,** 275-279.

55. Bax,A. and Davis,D.G. (1985). MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. *Journal of Magnetic Resonance,* **65,** 355-360.

56. Otting,G. and Wuthrich,K. (1989). Extended Heteronuclear Editing of 2D $H^1$-NMR Spectra of Isotope-Labeled Proteins, Using the X(ω-1, ω-2) Double Half Filter. *Journal of Magnetic Resonance,* **85,** 586-594.

57. Schleucher,J., Schwendinger,M., Sattler,M., Schmidt,P., Schedletzky,O., Glaser,S.J., Sorensen,O.W., and Griesinger,C. (1994). A general enhancement scheme in heteronuclear multidimensional NMR employing pulsed field gradients. *J. Biomol. NMR,* **4,** 301-306.

58. Rance,M., Louis,J.M., and Palmer,A.G. (1999). Sensitivity Improvement of Transverse Relaxation-Optimized Spectroscopy. *Journal of Magnetic Resonance,* **136,** 92-101.

59. Kuzmic,P. (2009). DynaFit--a software package for enzymology. *Methods Enzymol.,* **467,** 247-280.

60. Kuzmic,P. (1996). Program DYNAFIT for the analysis of enzyme kinetic data: application to HIV proteinase. *Anal. Biochem.,* **237,** 260-273.

61. Long,J., Garner,T.P., Pandya,M.J., Craven,C.J., Chen,P., Shaw,B., Williamson,M.P., Layfield,R., and Searle,M.S. (2010). Dimerisation of the UBA Domain of p62 Inhibits Ubiquitin Binding and Regulates NF-κB Signalling. *J. Mol. Biol.,* **396,** 178-194.

62. Williamson,M.P. (2003). Many residues in cytochrome c populate alternative states under equilibrium conditions. *Proteins-Structure Function and Genetics,* **53,** 731-739.

63. Andersen,N.H., Neidigh,J.W., Harris,S.M., Lee,G.M., Liu,Z., and Tong,H. (1997). Extracting Information from the Temperature Gradients of Polypeptide NH Chemical Shifts. 1. The Importance of Conformational Averaging. *Journal of the American Chemical Society,* **119,** 8547-8561.

64. Koradi,R., Billeter,M., and Wuthrich,K. (1996). MOLMOL: A program for display and analysis of macromolecular structures. *Journal of Molecular Graphics,* **14,** 51-55.

65. Cierpicki,T. and Otlewski,J. (2001). Amide proton temperature coefficients as hydrogen bond indicators in proteins. *Journal of Biomolecular Nmr,* **21,** 249-261.

66. Kay,L.E., Torchia,D.A., and Bax,A. (1989). Backbone dynamics of proteins as studied by $^{15}N$ inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry,* **28,** 8972-8979.

67. Wu,Y., Shih,S.C., and Goto,N.K. (2007). Probing the structure of the Ff bacteriophage major coat protein transmembrane helix dimer by solution NMR. *Biochim. Biophys. Acta,* **1768,** 3206-3215.

68. Liang,B., Bushweller,J.H., and Tamm,L.K. (2006). Site-directed parallel spin-labeling and paramagnetic relaxation enhancement in structure determination of membrane proteins by solution NMR spectroscopy. *J. Am. Chem. Soc.,* **128,** 4389-4397.

69. Ottiger,M., Delaglio,F., and Bax,A. (1998). Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J. Magn Reson.,* **131,** 373-378.

70. Sass,H.J., Musco,G., Stahl,S.J., Wingfield,P.T., and Grzesiek,S. (2001). An easy way to include weak alignment constraints into NMR structure calculations. *J. Biomol. NMR,* **21,** 275-280.

71. Meiler,J., Blomberg,N., Nilges,M., and Griesinger,C. (2000). A new approach for applying residual dipolar couplings as restraints in structure elucidation. *J. Biomol. NMR,* **16,** 245-252.

72. Dominguez,C., Boelens,R., and Bonvin,A.M. (2003). HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.,* **125,** 1731-1737.

73. de Vries,S.J., van Dijk,M., and Bonvin,A.M. (2010). The HADDOCK web server for data-driven biomolecular docking. *Nat. Protoc.,* **5,** 883-897.

74. Marion,D., Driscoll,P.C., Kay,L.E., Wingfield,P.T., Bax,A., Gronenborn,A.M., and Clore,G.M. (1989). Overcoming the overlap problem in the assignment of $^{1}H$ NMR spectra of larger proteins by use of three-dimensional heteronuclear $^{1}H$-$^{15}N$ Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1 β. *Biochemistry,* **28,** 6150-6156.

75. Zuiderweg,E.R. and Fesik,S.W. (1989). Heteronuclear three-dimensional NMR spectroscopy of the inflammatory protein C5a. *Biochemistry,* **28,** 2387-2391.

76. Kay,L.E., Mitsuhiko,I., Rolf,T., and Bax,A. (1990). Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *Journal of Magnetic Resonance (1969),* **89,** 496-514.

77. Grzesiek,S. and Bax,A. (1992). Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. *Journal of Magnetic Resonance (1969),* **96,** 432-440.

78. Muhandiram,D.R. and Kay,L.E. (1994). Gradient-Enhanced Triple-Resonance Three-Dimensional NMR Experiments with Improved Sensitivity. *Journal of Magnetic Resonance, Series B,* **103,** 203-216.

79. Clubb,R.T., Thanabal,V., and Wagner,G. (1992). A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue $^{1}$HN, $^{15}$N, and $^{13}$C' chemical shifts in $^{15}$N-$^{13}$C-labelled proteins. *Journal of Magnetic Resonance,* **97,** 213-217.

80. Grzesiek,S. and Bax,A. (1992). Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *Journal of the American Chemical Society,* **114,** 6291-6293.

81. Kay,L.E., Xu,G.Y., Singer,A.U., Muhandiram,D.R., and Formankay,J.D. (1993). A Gradient-Enhanced HCCH TOCSY Experiment for Recording Side-Chain H-$^{1}$ and C-$^{13}$ Correlations in $H_2O$ Samples of Proteins. *Journal of Magnetic Resonance Series B,* **101,** 333-337.

82. Higman, V. A. Protein NMR - A Practical Guide. Higman, V. A. http://www.protein-nmr.org.uk/assignment_theory.html . 2010. 30-5-2010.
Ref Type: Electronic Citation

83. Berjanskii,M.V., Neal,S., and Wishart,D.S. (2006). PREDITOR: a web server for predicting protein torsion angle restraints. *Nucleic Acids Res.,* **34,** W63-W69.

84. Shen,Y., Delaglio,F., Cornilescu,G., and Bax,A. (2009). TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR,* **44,** 213-223.

85. The UniProt Consortium (2009). The Universal Protein Resource (UniProt) 2009. *Nucleic Acids Res.,* **37,** D169-D174.

86. Thompson,J.D., Gibson,T.J., Plewniak,F., Jeanmougin,F., and Higgins,D.G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research,* **25,** 4876-4882.

87. Schneider,T.D. and Stephens,R.M. (1990). Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.,* **18,** 6097-6100.

88. Crooks,G.E., Hon,G., Chandonia,J.M., and Brenner,S.E. (2004). WebLogo: a sequence logo generator. *Genome Res.,* **14,** 1188-1190.

89. Schultz,J., Milpetz,F., Bork,P., and Ponting,C.P. (1998). SMART, a simple modular architecture research tool: identification of signaling domains. *Proc. Natl. Acad. Sci. U. S. A,* **95,** 5857-5864.

90. Sigrist,C.J., Cerutti,L., de Castro,E., Langendijk-Genevaux,P.S., Bulliard,V., Bairoch,A., and Hulo,N. (2010). PROSITE, a protein domain database for functional characterization and annotation. *Nucleic Acids Res.,* **38,** D161-D166.

91. Gould,C.M., Diella,F., Via,A., Puntervoll,P., Gemund,C., Chabanis-Davidson,S., Michael,S., Sayadi,A., Bryne,J.C., Chica,C., Seiler,M., Davey,N.E., Haslam,N., Weatheritt,R.J., Budd,A., Hughes,T., Pas,J., Rychlewski,L., Trave,G., Aasland,R., Helmer-Citterich,M., Linding,R., and Gibson,T.J. (2010). ELM: the status of the 2010 eukaryotic linear motif resource. *Nucleic Acids Res.,* **38,** D167-D180.

92. Peng,K., Radivojac,P., Vucetic,S., Dunker,A.K., and Obradovic,Z. (2006). Length-dependent prediction of protein intrinsic disorder. *BMC. Bioinformatics.,* **7,** 208.

93. Dosztanyi,Z., Csizmok,V., Tompa,P., and Simon,I. (2005). IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics.,* **21,** 3433-3434.

94. Cheng,J., Randall,A.Z., Sweredoski,M.J., and Baldi,P. (2005). SCRATCH: a protein structure and structural feature prediction server. *Nucleic Acids Res.,* **33,** W72-W76.

95. Ward,J.J., Sodhi,J.S., McGuffin,L.J., Buxton,B.F., and Jones,D.T. (2004). Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol.,* **337,** 635-645.

96. Guex,N., Peitsch,M.C., and Schwede,T. (2009). Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PDBViewer: a historical perspective. *Electrophoresis,* **30 Suppl 1,** S162-S173.

97. Arnold,K., Bordoli,L., Kopp,J., and Schwede,T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics.,* **22,** 195-201.

98. Wishart,D.S., Arndt,D., Berjanskii,M., Tang,P., Zhou,J., and Lin,G. (2008). CS23D: a web server for rapid protein structure generation using NMR chemical shifts and sequence data. *Nucleic Acids Res.,* **36,** W496-W502.

99. Bomar,M.G., Pai,M.T., Tzeng,S.R., Li,S.S., and Zhou,P. (2007). Structure of the ubiquitin-binding zinc finger domain of human DNA Y-polymerase η. *EMBO Rep.,* **8,** 247-251.

100. Nilges,M., Gronenborn,A.M., Brunger,A.T., and Clore,G.M. (1988). Determination of three-dimensional structures of proteins by simulated annealing with interproton distance restraints. Application to crambin, potato carboxypeptidase inhibitor and barley serine proteinase inhibitor 2. *Protein Eng,* **2,** 27-38.

101. Nilges,M., Clore,G.M., and Gronenborn,A.M. (1988). Determination of three-dimensional structures of proteins from interproton distance data by

dynamical simulated annealing from a random array of atoms. Circumventing problems associated with folding. *FEBS Lett.,* **239,** 129-136.

102. Dosset,P., Hus,J.C., Marion,D., and Blackledge,M. (2001). A novel interactive tool for rigid-body modeling of multi-domain macromolecules using residual dipolar couplings. *J. Biomol. NMR,* **20,** 223-231.

103. Bhattacharya,A., Tejero,R., and Montelione,G.T. (2007). Evaluating protein structures determined by structural genomics consortia. *Proteins,* **66,** 778-795.

104. Huang,Y.J., Powers,R., and Montelione,G.T. (2005). Protein NMR recall, precision, and F-measure scores (RPF scores): structure quality assessment measures based on information retrieval statistics. *J. Am. Chem. Soc.,* **127,** 1665-1674.

105. Laskowski,R.A., Moss,D.S., and Thornton,J.M. (1993). Main-chain bond lengths and bond angles in protein structures. *J. Mol. Biol.,* **231,** 1049-1067.

106. Lovell,S.C., Davis,I.W., Arendall,W.B., III, de Bakker,P.I., Word,J.M., Prisant,M.G., Richardson,J.S., and Richardson,D.C. (2003). Structure validation by Cα geometry: φ,ψ and Cβ deviation. *Proteins,* **50,** 437-450.

107. Luthy,R., Bowie,J.U., and Eisenberg,D. (1992). Assessment of protein models with three-dimensional profiles. *Nature,* **356,** 83-85.

108. Sippl,M.J. (1993). Recognition of errors in three-dimensional structures of proteins. *Proteins,* **17,** 355-362.

109. Neal,S., Nip,A.M., Zhang,H., and Wishart,D.S. (2003). Rapid and accurate calculation of protein $^{1}H$, $^{13}C$ and $^{15}N$ chemical shifts. *J. Biomol. NMR,* **26,** 215-240.

110. Zweckstetter,M., Hummer,G., and Bax,A. (2004). Prediction of charge-induced molecular alignment of biomolecules dissolved in dilute liquid-crystalline phases. *Biophys. J.,* **86,** 3444-3460.

111. Sandri,M. (2008). Signaling in muscle atrophy and hypertrophy. *Physiology (Bethesda. ),* **23,** 160-170.

112. Murton,A.J., Constantin,D., and Greenhaff,P.L. (2008). The involvement of the ubiquitin proteasome system in human skeletal muscle remodelling and atrophy. *Biochim. Biophys. Acta,* **1782,** 730-743.

113. Hishiya,A., Iemura,S., Natsume,T., Takayama,S., Ikeda,K., and Watanabe,K. (2006). A novel ubiquitin-binding protein ZNF216 functioning in muscle atrophy. *EMBO J.,* **25,** 554-564.

114. Lecker,S.H., Solomon,V., Mitch,W.E., and Goldberg,A.L. (1999). Muscle protein breakdown and the critical role of the ubiquitin-proteasome pathway in normal and disease states. *J. Nutr.,* **129,** 227S-237S.

115. Scott,D.A., Greinwald,J.H., Jr., Marietta,J.R., Drury,S., Swiderski,R.E., Vinas,A., DeAngelis,M.M., Carmi,R., Ramesh,A., Kraft,M.L., Elbedour,K., Skworak,A.B., Friedman,R.A., Srikumari Srisailapathy,C.R., Verhoeven,K., Van Gamp,G., Lovett,M., Deininger,P.L., Batzer,M.A., Morton,C.C., Keats,B.J., Smith,R.J., and Sheffield,V.C. (1998). Identification and mutation analysis of a cochlear-expressed, zinc finger protein gene at the DFNB7/11 and dn hearing-loss loci on human chromosome 9q and mouse chromosome 19. *Gene,* **215,** 461-469.

116. Hishiya,A., Ikeda,K., and Watanabe,K. (2005). A RANKL-inducible gene ZNF216 in osteoclast differentiation. *J. Recept. Signal. Transduct. Res.,* **25,** 199-216.

117. Huang,J., Teng,L., Li,L., Liu,T., Li,L., Chen,D., Xu,L.G., Zhai,Z., and Shu,H.B. (2004). ZNF216 Is an A20-like and IκB kinase γ-interacting inhibitor of NFκB activation. *J. Biol. Chem.,* **279,** 16847-16853.

118. Goodsell, D. S. Molecule of the Month: Zinc Fingers. www.rcsb.org/pdb/static.do?p=education_discussion/molecule_of_the_month/pdb87_1.html . 2007. RSCPDB. 4-6-2010.
Ref Type: Electronic Citation

119. Laity,J.H., Lee,B.M., and Wright,P.E. (2001). Zinc finger proteins: new insights into structural and functional diversity. *Curr. Opin. Struct. Biol.,* **11,** 39-46.

120. Vij,S. and Tyagi,A.K. (2006). Genome-wide analysis of the stress associated protein (SAP) gene family containing A20/AN1 zinc-finger(s) in rice and their phylogenetic relationship with Arabidopsis. *Mol. Genet. Genomics,* **276,** 565-575.

121. Duan,W., Sun,B., Li,T.W., Tan,B.J., Lee,M.K., and Teo,T.S. (2000). Cloning and characterization of AWP1, a novel protein that associates with serine/threonine kinase PRK1 in vivo. *Gene,* **256,** 113-121.

122. Mattera,R., Tsai,Y.C., Weissman,A.M., and Bonifacino,J.S. (2006). The Rab5 guanine nucleotide exchange factor Rabex-5 binds ubiquitin (Ub) and functions as a Ub ligase through an atypical Ub-interacting motif and a zinc finger domain. *J. Biol. Chem.,* **281,** 6874-6883.

123. Fenner,B.J., Scannell,M., and Prehn,J.H. (2009). Identification of polyubiquitin binding proteins involved in NF-κB signaling using protein arrays. *Biochim. Biophys. Acta,* **1794,** 1010-1016.

124. Diatchenko,L., Romanov,S., Malinina,I., Clarke,J., Tchivilev,I., Li,X., and Makarov,S.S. (2005). Identification of novel mediators of NF-κB through genome-wide survey of monocyte adherence-induced genes. *J. Leukoc. Biol.,* **78,** 1366-1377.

125. De Valck,D., Heyninck,K., Van Criekinge,W., Contreras,R., Beyaert,R., and Fiers,W. (1996). A20, an inhibitor of cell death, self-associates by its zinc finger domain. *FEBS Lett.,* **384,** 61-64.

126. Kanneganti,V. and Gupta,A.K. (2008). Overexpression of OsiSAP8, a member of stress associated protein (SAP) gene family of rice confers tolerance to salt, drought and cold stress in transgenic tobacco and rice. *Plant Mol. Biol.,* **66,** 445-462.

127. Grishin,N.V. (2001). Treble clef finger--a functionally diverse zinc-binding structural motif. *Nucleic Acids Res.,* **29,** 1703-1714.

128. Eyles,S.J. and Kaltashov,I.A. (2004). Methods to study protein dynamics and folding by mass spectrometry. *Methods,* **34,** 88-99.

129. Krezel,A. and Maret,W. (2007). Different redox states of metallothionein/thionein in biological tissue. *Biochem. J.,* **402,** 551-558.

130. Wallimann,P., Kennedy,R.J., Miller,J.S., Shalongo,W., and Kemp,D.S. (2003). Dual wavelength parametric test of two-state models for circular dichroism spectra of helical polypeptides: anomalous dichroic properties of alanine-rich peptides. *J. Am. Chem. Soc.,* **125,** 1203-1220.

131. Wuthrich,K., Wider,G., Wagner,G., and Braun,W. (1982). Sequential resonance assignments as a basis for determination of spatial protein structures by high resolution proton nuclear magnetic resonance. *J. Mol. Biol.,* **155,** 311-319.

132. Wuthrich,K. (1989). Protein structure determination in solution by nuclear magnetic resonance spectroscopy. *Science,* **243,** 45-50.

133. Schwieters,C.D., Kuszewski,J.J., Tjandra,N., and Clore,G.M. (2003). The Xplor-NIH NMR molecular structure determination package. *J. Magn Reson.,* **160,** 65-73.

134. Bernado,P., Blanchard,L., Timmins,P., Marion,D., Ruigrok,R.W., and Blackledge,M. (2005). A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. *Proc. Natl. Acad. Sci. U. S. A,* **102,** 17002-17007.

135. Bouvignies,G., Bernado,P., and Blackledge,M. (2005). Protein backbone dynamics from N-HN dipolar couplings in partially aligned systems: a comparison of motional models in the presence of structural noise. *J. Magn Reson.,* **173,** 328-338.

136. Biological Magnetic Resonance Data Bank. www.bmrb.wisc.edu/ . 2010. The University of Wisconsin.
Ref Type: Electronic Citation

137. Schwarzinger,S., Kroon,G.J., Foss,T.R., Wright,P.E., and Dyson,H.J. (2000). Random coil chemical shifts in acidic 8 M urea: implementation of random coil shift data in NMRView. *J. Biomol. NMR,* **18,** 43-48.

138. Kornhaber,G.J., Snyder,D., Moseley,H.N., and Montelione,G.T. (2006). Identification of zinc-ligated cysteine residues based on $^{13}C\alpha$ and $^{13}C\beta$ chemical shift data. *J. Biomol. NMR,* **34,** 259-269.

139. Garcia,D.L.T., Huertas,M.L., and Carrasco,B. (2000). Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophys. J.,* **78,** 719-730.

140. Daniel,J.M., Friess,S.D., Rajagopalan,S., Wendt,S., and Zenobi,R. (2002). Quantitative determination of noncovalent binding interactions using soft ionization mass spectrometry. *International Journal of Mass Spectrometry,* **216,** 1-27.

141. Wishart,D.S., Sykes,B.D., and Richards,F.M. (1992). The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry,* **31,** 1647-1651.

142. Wishart,D.S. and Sykes,B.D. (1994). The $^{13}C$ chemical-shift index: a simple method for the identification of protein secondary structure using $^{13}C$ chemical-shift data. *J. Biomol. NMR,* **4,** 171-180.

143. Berjanskii,M. and Wishart,D.S. (2006). NMR: prediction of protein flexibility. *Nat. Protoc.,* **1,** 683-688.

144. Berjanskii,M.V. and Wishart,D.S. (2007). The RCI server: rapid and accurate calculation of protein flexibility using chemical shifts. *Nucleic Acids Res.,* **35,** W531-W537.

145. Jacobson,N.E. (2007). *NMR Spectroscopy Explained.* John Wiley & sons, Inc, Hoboken.

146. Schubert,M., Labudde,D., Oschkinat,H., and Schmieder,P. (2002). A software tool for the prediction of Xaa-Pro peptide bond conformations in proteins based on $^{13}C$ chemical shift statistics. *J. Biomol. NMR,* **24,** 149-154.

147. Cheng,H.N. and Bovey,F.A. (1977). Cis-trans equilibrium and kinetic studies of acetyl-L-proline and glycyl-L-proline. *Biopolymers,* **16,** 1465-1472.

148. Raasi,S., Varadan,R., Fushman,D., and Pickart,C.M. (2005). Diverse polyubiquitin interaction properties of ubiquitin-associated domains. *Nat. Struct. Mol. Biol.,* **12,** 708-714.

149. Huang,J., Wang,M.M., Jiang,Y., Bao,Y.M., Huang,X., Sun,H., Xu,D.Q., Lan,H.X., and Zhang,H.S. (2008). Expression analysis of rice A20/AN1-type zinc finger genes and characterization of ZFP177 that contributes to temperature stress tolerance. *Gene,* **420,** 135-144.

150. Tang,E.D., Wang,C.Y., Xiong,Y., and Guan,K.L. (2003). A role for NF-κB essential modifier/IκB kinase-gamma (NEMO/IKKgamma) ubiquitination in the activation of the IκB kinase complex by tumor necrosis factor-α. *J. Biol. Chem.,* **278,** 37297-37305.

151. O'Donnell,M.A., Legarda-Addison,D., Skountzos,P., Yeh,W.C., and Ting,A.T. (2007). Ubiquitination of RIP1 regulates an NF-κB-independent cell-death switch in TNF signaling. *Curr. Biol.,* **17,** 418-424.

152. Zhang,D., Raasi,S., and Fushman,D. (2008). Affinity makes the difference: nonselective interaction of the UBA domain of Ubiquilin-1 with monomeric ubiquitin and polyubiquitin chains. *J. Mol. Biol.,* **377,** 162-180.

153. Zweckstetter,M. (2008). NMR: prediction of molecular alignment from structure using the PALES software. *Nat. Protoc.,* **3,** 679-690.

154. Wang,X., Bansal,S., Jiang,M., and Prestegard,J.H. (2008). RDC-assisted modeling of symmetric protein homo-oligomers. *Protein Sci.,* **17,** 899-907.

155. Mason,P.E., Neilson,G.W., Enderby,J.E., Saboungi,M.L., Dempsey,C.E., MacKerell,A.D., Jr., and Brady,J.W. (2004). The structure of aqueous guanidinium chloride solutions. *J. Am. Chem. Soc.,* **126,** 11462-11470.

156. Jennings,W.B., Farrell,B.M., and Malone,J.F. (2001). Attractive intramolecular edge-to-face aromatic interactions in flexible organic molecules. *Acc. Chem. Res.,* **34,** 885-894.

157. Takayama,Y., Castaneda,C.A., Chimenti,M., Garcia-Moreno,B., and Iwahara,J. (2008). Direct evidence for deprotonation of a lysine side chain buried in the hydrophobic core of a protein. *J. Am. Chem. Soc.,* **130,** 6714-6715.

158. Toth,G., Watts,C.R., Murphy,R.F., and Lovas,S. (2001). Significance of aromatic-backbone amide interactions in protein structure. *Proteins,* **43,** 373-381.

159. Perica,T. and Chothia,C. (2010). Ubiquitin-molecular mechanisms for recognition of different structures. *Curr. Opin. Struct. Biol.*. **20**, 367-376

160. Loewenthal,R., Sancho,J., Reinikainen,T., and Fersht,A.R. (1993). Long-range surface charge-charge interactions in proteins. Comparison of experimental results with calculations from a theoretical method. *J. Mol. Biol.,* **232,** 574-583.

161. Grimsley,G.R., Shaw,K.L., Fee,L.R., Alston,R.W., Huyghues-Despointes,B.M., Thurlkill,R.L., Scholtz,J.M., and Pace,C.N. (1999). Increasing protein stability by altering long-range coulombic interactions. *Protein Sci.,* **8,** 1843-1849.

162. Raschke,T.M. (2006). Water structure and interactions with protein surfaces. *Curr. Opin. Struct. Biol.,* **16,** 152-159.

163. Ben Saad,R., Zouari,N., Ben Ramdhan,W., Azaza,J., Meynard,D., Guiderdoni,E., and Hassairi,A. (2010). Improved drought and salt stress tolerance in transgenic tobacco overexpressing a novel A20/AN1 zinc-finger "AlSAP" gene isolated from the halophyte grass *Aeluropus littoralis*. *Plant Mol. Biol.,* **72,** 171-190.

164. Kang,Y., Vossler,R.A., Diaz-Martinez,L.A., Winter,N.S., Clarke,D.J., and Walters,K.J. (2006). UBL/UBA ubiquitin receptor proteins bind a common tetraubiquitin chain. *J. Mol. Biol.,* **356,** 1027-1035.

165. Satou,Y. and Satoh,N. (1997). Posterior end mark 2 (pem-2), pem-4, pem-5, and pem-6: maternal genes with localized mRNA in the ascidian embryo. *Dev. Biol.,* **192,** 467-481.

166. Dosztanyi,Z., Csizmok,V., Tompa,P., and Simon,I. (2005). The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *J. Mol. Biol.,* **347,** 827-839.

167. Vaynberg,J. and Qin,J. (2006). Weak protein-protein interactions as probed by NMR spectroscopy. *Trends Biotechnol.,* **24,** 22-27.

168. Andersen,N.H., Liu,Z., and Prickett,K.S. (1996). Efforts toward deriving the CD spectrum of a $3^{10}$ helix in aqueous medium. *FEBS Lett.,* **399,** 47-52.

169. Ahmed,M.A., Bamm,V.V., Harauz,G., and Ladizhansky,V. (2007). The BG21 isoform of Golli myelin basic protein is intrinsically disordered with a highly flexible amino-terminal domain. *Biochemistry,* **46,** 9700-9712.

170. Najat,D., Garner,T., Hagen,T., Shaw,B., Sheppard,P.W., Falchetti,A., Marini,F., Brandi,M.L., Long,J.E., Cavey,J.R., Searle,M.S., and Layfield,R. (2009). Characterization of a non-UBA domain missense mutation of sequestosome 1 (SQSTM1) in Paget's disease of bone. *J. Bone Miner. Res.,* **24,** 632-642.

171. Walters,K.J., Lech,P.J., Goh,A.M., Wang,Q., and Howley,P.M. (2003). DNA-repair protein hHR23a alters its protein structure upon binding proteasomal subunit S5a. *Proc. Natl. Acad. Sci. U. S. A,* **100,** 12694-12699.

172. Huang,Y. and Liu,Z. (2009). Kinetic advantage of intrinsically disordered proteins in coupled folding-binding process: a critical assessment of the "fly-casting" mechanism. *J. Mol. Biol.,* **393,** 1143-1159.

173. Moscat,J. and Diaz-Meco,M.T. (2009). p62 at the crossroads of autophagy, apoptosis, and cancer. *Cell,* **137,** 1001-1004.

174. Wilson,M.I., Gill,D.J., Perisic,O., Quinn,M.T., and Williams,R.L. (2003). PB1 domain-mediated heterodimerization in NADPH oxidase and signaling complexes of atypical protein kinase C with Par6 and p62. *Mol. Cell,* **12,** 39-50.

175. Nakamura,K., Kimple,A.J., Siderovski,D.P., and Johnson,G.L. (2010). PB1 domain interaction of p62/sequestosome 1 and MEKK3 regulates NF-κB activation. *J. Biol. Chem.,* **285,** 2077-2089.

176. Lange,S., Xiang,F., Yakovenko,A., Vihola,A., Hackman,P., Rostkova,E., Kristensen,J., Brandmeier,B., Franzen,G., Hedberg,B., Gunnarsson,L.G., Hughes,S.M., Marchand,S., Sejersen,T., Richard,I., Edstrom,L., Ehler,E.,

Udd,B., and Gautel,M. (2005). The kinase domain of titin controls muscle gene expression and protein turnover. *Science,* **308,** 1599-1603.

177. Lamark,T., Perander,M., Outzen,H., Kristiansen,K., Overvatn,A., Michaelsen,E., Bjorkoy,G., and Johansen,T. (2003). Interaction codes within the family of mammalian Phox and Bem1p domain-containing proteins. *J. Biol. Chem.,* **278,** 34568-34581.

178. Sanz,L., Sanchez,P., Lallena,M.J., Diaz-Meco,M.T., and Moscat,J. (1999). The interaction of p62 with RIP links the atypical PKCs to NF-κB activation. *EMBO J.,* **18,** 3044-3053.

179. Seibenhener,M.L., Babu,J.R., Geetha,T., Wong,H.C., Krishna,N.R., and Wooten,M.W. (2004). Sequestosome 1/p62 is a polyubiquitin chain binding protein involved in ubiquitin proteasome degradation. *Mol. Cell Biol.,* **24,** 8055-8068.

180. Long,J., Gallagher,T.R., Cavey,J.R., Sheppard,P.W., Ralston,S.H., Layfield,R., and Searle,M.S. (2008). Ubiquitin recognition by the ubiquitin-associated domain of p62 involves a novel conformational switch. *J. Biol. Chem.,* **283,** 5427-5440.

181. Sanz,L., Diaz-Meco,M.T., Nakano,H., and Moscat,J. (2000). The atypical PKC-interacting protein p62 channels NF-κB activation by the IL-1-TRAF6 pathway. *EMBO J.,* **19,** 1576-1586.

182. Feng,Y.F. and Longmore,G.D. (2005). The LIM protein Ajuba influences interleukin-1-induced NF-κ B activation by affecting the assembly and activity of the protein kinase C ζ/p62/TRAF6 signaling complex. *Molecular and Cellular Biology,* **25,** 4010-4022.

183. Lau,A., Wang,X.J., Zhao,F., Villeneuve,N.F., Wu,T., Jiang,T., Sun,Z., White,E., and Zhang,D.D. (2010). A noncanonical mechanism of Nrf2 activation by autophagy deficiency: direct interaction between Keap1 and p62. *Mol. Cell Biol.,* **30,** 3275-3285.

184. Saito,A., Kawai,K., Takayama,H., Sudo,T., and Osada,H. (2008). Improvement of photoaffinity SPR imaging platform and determination of the binding site of p62/SQSTM1 to p38 MAP kinase. *Chem. Asian J.,* **3,** 1607-1612.

185. Garcia-Alai,M.M., Gallo,M., Salame,M., Wetzler,D.E., McBride,A.A., Paci,M., Cicero,D.O., and Prat-Gay,G. (2006). Molecular basis for phosphorylation-dependent, PEST-mediated protein turnover. *Structure,* **14,** 309-319.

186. Jin,Z., Li,Y., Pitti,R., Lawrence,D., Pham,V.C., Lill,J.R., and Ashkenazi,A. (2009). Cullin3-based polyubiquitination and p62-dependent aggregation of caspase-8 mediate extrinsic apoptosis signaling. *Cell,* **137,** 721-735.

187. Yu,H.B., Kielczewska,A., Rozek,A., Takenaka,S., Li,Y., Thorson,L., Hancock,R.E., Guarna,M.M., North,J.R., Foster,L.J., Donini,O., and

Finlay,B.B. (2009). Sequestosome-1/p62 is the key intracellular target of innate defense regulator peptide. *J. Biol. Chem.,* **284,** 36007-36011.

188. Feng,Y.F., Zhao,H.B., Luderer,H.F., Epple,H., Faccio,R., Ross,F.P., Teitelbaum,S.L., and Longmore,G.D. (2007). The LIM protein, LIMD1, regulates AP-1 activation through an interaction with TRAF6 to influence osteoclast development. *Journal of Biological Chemistry,* **282,** 39-48.

189. Clausen,T.H., Lamark,T., Isakson,P., Finley,K., Larsen,K.B., Brech,A., Overvatn,A., Stenmark,H., Bjorkoy,G., Simonsen,A., and Johansen,T. (2010). p62/SQSTM1 and ALFY interact to facilitate the formation of p62 bodies/ALIS and their degradation by autophagy. *Autophagy.,* **6,** 330-344.

190. Wooten,M.W., Geetha,T., Seibenhener,M.L., Babu,J.R., Diaz-Meco,M.T., and Moscat,J. (2005). The p62 scaffold regulates nerve growth factor-induced NF-κB activation by influencing TRAF6 polyubiquitination. *J. Biol. Chem.,* **280,** 35625-35629.

191. Pankiv,S., Clausen,T.H., Lamark,T., Brech,A., Bruun,J.A., Outzen,H., Overvatn,A., Bjorkoy,G., and Johansen,T. (2007). p62/SQSTM1 Binds Directly to Atg8/LC3 to Facilitate Degradation of Ubiquitinated Protein Aggregates by Autophagy. *J. Biol. Chem.,* **282,** 24131-24145.

192. Moscat,J., Diaz-Meco,M.T., and Wooten,M.W. (2007). Signal integration and diversification through the p62 scaffold protein. *Trends Biochem. Sci.,* **32,** 95-100.

193. Babu,J.R., Geetha,T., and Wooten,M.W. (2005). Sequestosome 1/p62 shuttles polyubiquitinated tau for proteasomal degradation. *J. Neurochem.,* **94,** 192-203.

194. Ichimura,Y. and Komatsu,M. (2010). Selective degradation of p62 by autophagy. *Semin. Immunopathol.*. **32**, 431-436.

195. Lee,S.J., Pfluger,P.T., Kim,J.Y., Nogueiras,R., Duran,A., Pages,G., Pouyssegur,J., Tschop,M.H., Diaz-Meco,M.T., and Moscat,J. (2010). A functional role for the p62-ERK1 axis in the control of energy homeostasis and adipogenesis. *EMBO Rep.,* **11,** 226-232.

196. Geetha,T., Seibenhener,M.L., Chen,L., Madura,K., and Wooten,M.W. (2008). p62 serves as a shuttling factor for TrkA interaction with the proteasome. *Biochem. Biophys. Res. Commun.,* **374,** 33-37.

197. Moreno,L., Frazziano,G., Cogolludo,A., Cobeno,L., Tamargo,J., and Perez-Vizcaino,F. (2007). Role of PKCζ and its adaptor protein p62 in KV channel modulation in pulmonary arteries. *Mol. Pharmacol.*.

198. Mizushima,N., Levine,B., Cuervo,A.M., and Klionsky,D.J. (2008). Autophagy fights disease through cellular self-digestion. *Nature,* **451,** 1069-1075.

199. Lamark,T., Kirkin,V., Dikic,I., and Johansen,T. (2009). NBR1 and p62 as cargo receptors for selective autophagy of ubiquitinated targets. *Cell Cycle,* **8,** 1986-1990.

200. Kirkin,V., McEwan,D.G., Novak,I., and Dikic,I. (2009). A role for ubiquitin in selective autophagy. *Mol. Cell,* **34,** 259-269.

201. Seibenhener,M.L., Geetha,T., and Wooten,M.W. (2007). Sequestosome 1/p62 - More than just a scaffold. *Febs Letters,* **581,** 175-179.

202. Cavey,J.R., Ralston,S.H., Sheppard,P.W., Ciani,B., Gallagher,T.R., Long,J.E., Searle,M.S., and Layfield,R. (2006). Loss of ubiquitin binding is a unifying mechanism by which mutations of SQSTM1 cause Paget's disease of bone. *Calcif. Tissue Int.,* **78,** 271-277.

203. Roodman,G.D. (2010). Insights into the pathogenesis of Paget's disease. *Ann. N. Y. Acad. Sci.,* **1192,** 176-180.

204. Goode,A. and Layfield,R. (2010). Recent advances in understanding the molecular basis of Paget disease of bone. *J. Clin. Pathol.,* **63,** 199-203.

205. Wuyts,W., Van Wesenbeeck,L., Morales-Piga,A., Ralston,S., Hocking,L., Vanhoenacker,F., Westhovens,R., Verbruggen,L., Anderson,D., Hughes,A., and Van Hul,W. (2001). Evaluation of the role of RANK and OPG genes in Paget's disease of bone. *Bone,* **28,** 104-107.

206. Middleton-Hardie,C., Zhu,Q., Cundy,H., Lin,J.M., Callon,K., Tong,P.C., Xu,J., Grey,A., Cornish,J., and Naot,D. (2006). Deletion of aspartate 182 in OPG causes juvenile Paget's disease by impairing both protein secretion and binding to RANKL. *J. Bone Miner. Res.,* **21,** 438-445.

207. Visconti,M.R., Langston,A.L., Alonso,N., Goodman,K., Selby,P.L., Fraser,W.D., and Ralston,S.H. (2010). Mutations of SQSTM1 are associated with severity and clinical outcome in Paget's disease of bone. *J. Bone Miner. Res.*.

208. Layfield,R., Ciani,B., Ralston,S.H., Hocking,L.J., Sheppard,P.W., Searle,M.S., and Cavey,J.R. (2004). Structural and functional studies of mutations affecting the UBA domain of SQSTM1 (p62) which cause Paget's disease of bone. *Biochem. Soc. Trans.,* **32,** 728-730.

209. Ciani,B., Layfield,R., Cavey,J.R., Sheppard,P.W., and Searle,M.S. (2003). Structure of the ubiquitin-associated domain of p62 (SQSTM1) and implications for mutations that cause Paget's disease of bone. *J. Biol. Chem.,* **278,** 37409-37412.

210. Evans,C.L., Long,J.E., Gallagher,T.R., Hirst,J.D., and Searle,M.S. (2008). Conformation and dynamics of the three-helix bundle UBA domain of p62 from experiment and simulation. *Proteins,* **71,** 227-240.

211. Heinen,C., Garner,T.P., Long,J., Bottcher,C., Ralston,S.H., Cavey,J.R., Searle,M.S., Layfield,R., and Dantuma,N.P. (2010). Mutant p62/SQSTM1

UBA domains linked to Paget's disease of bone differ in their abilities to function as stabilization signals. *FEBS Lett.,* **584,** 1585-1590.

212. Myszka,D.G. and Chaiken,I.M. (1994). Design and characterization of an intramolecular antiparallel coiled coil peptide. *Biochemistry,* **33,** 2363-2372.

213. Lau,S.Y., Taneja,A.K., and Hodges,R.S. (1984). Synthesis of a model protein of defined secondary and quaternary structure. Effect of chain length on the stabilization and formation of two-stranded alpha-helical coiled-coils. *J. Biol. Chem.,* **259,** 13253-13261.

214. Dey,S., Pal,A., Chakrabarti,P., and Janin,J. (2010). The subunit interfaces of weakly associated homodimeric proteins. *J. Mol. Biol.,* **398,** 146-160.

215. Deschamps,M.L., Pilka,E.S., Potts,J.R., Campbell,I.D., and Boyd,J. (2005). Probing protein-peptide binding surfaces using charged stable free radicals and transverse paramagnetic relaxation enhancement (PRE). *J. Biomol. NMR,* **31,** 155-160.

216. Jahnke,W. (2002). Spin labels as a tool to identify and characterize protein-ligand interactions by NMR spectroscopy. *Chembiochem.,* **3,** 167-173.

217. Varadan,R., Walker,O., Pickart,C., and Fushman,D. (2002). Structural properties of polyubiquitin chains in solution. *Journal of Molecular Biology,* **324,** 637-647.

218. Fawzi,N.L., Doucleff,M., Suh,J.Y., and Clore,G.M. (2010). Mechanistic details of a protein-protein association pathway revealed by paramagnetic relaxation enhancement titration measurements. *Proc. Natl. Acad. Sci. U. S. A*. **107**, 1379-1384

219. Cierpicki,T. and Otlewski,J. (2001). Amide proton temperature coefficients as hydrogen bond indicators in proteins. *Journal of Biomolecular Nmr,* **21,** 249-261.

220. Cierpicki,T., Zhukov,R., Byrd,A.R., and Otlewski,J. (2002). Hydrogen bonds in human ubiquitin reflects in temperature coefficients of amide protons. *Journal of Magn Resin,* **157,** 178-180.

221. Kumar,A., Srivastava,S., and Hosur,R., V (2007). NMR Characterization of the Energy Landscape of SUMO-1 in the Native-state Ensemble. *Journal of Molecular Biology,* **367,** 1480-1493.

222. Tunnicliffe,R.B., Waby,J.L., Williams,R.J., and Williamson,M.P. (2005). An experimental investigation of conformational fluctuations in proteins G and L. *Structure,* **13,** 1677-1684.

223. Eswar,N. and Ramakrishnan,C. (2000). Deterministic features of side-chain main-chain hydrogen bonds in globular protein structures. *Protein Eng,* **13,** 227-238.

224. Baxter,N.J., Hosszu,L.L., Waltho,J.P., and Williamson,M.P. (1998). Characterisation of low free-energy excited states of folded proteins. *J. Mol. Biol.,* **284,** 1625-1639.

225. Sandhu,K.S. and Dash,D. (2006). Conformational flexibility may explain multiple cellular roles of PEST motifs. *Proteins-Structure Function and Bioinformatics,* **63,** 727-732.

226. Singh,G.P., Ganapathi,M., Sandhu,K.S., and Dash,D. (2006). Intrinsic unstructuredness and abundance of PEST motifs in eukaryotic proteomes. *Proteins-Structure Function and Bioinformatics,* **62,** 309-315.

227. Tompa,P., Prilusky,J., Silman,I., and Sussman,J.L. (2008). Structural disorder serves as a weak signal for intracellular protein degradation. *Proteins,* **71,** 903-909.

228. Cavey,J.R., Ralston,S.H., Hocking,L.J., Sheppard,P.W., Ciani,B., Searle,M.S., and Layfield,R. (2005). Loss of ubiquitin-binding associated with Paget's disease of bone p62 (SQSTM1) mutations. *J. Bone Miner. Res.,* **20,** 619-624.

229. Layfield,R. and Hocking,L.J. (2004). SQSTM1 and Paget's disease of bone. *Calcif. Tissue Int.,* **75,** 347-357.

230. Michou,L., Collet,C., Laplanche,J.L., Orcel,P., and Cornelis,F. (2006). Genetics of Paget's disease of bone. *Joint Bone Spine,* **73,** 243-248.

231. Gota,C., Okabe,K., Funatsu,T., Harada,Y., and Uchiyama,S. (2009). Hydrophilic fluorescent nanogel thermometer for intracellular thermometry. *J. Am. Chem. Soc.,* **131,** 2766-2767.

232. Bayrer,J.R., Zhang,W., and Weiss,M.A. (2005). Dimerization of doublesex is mediated by a cryptic ubiquitin-associated domain fold: implications for sex-specific gene regulation. *J. Biol. Chem.,* **280,** 32989-32996.

233. Yang,Y., Zhang,W., Bayrer,J.R., and Weiss,M.A. (2008). Doublesex and the regulation of sexual dimorphism in Drosophila melanogaster: structure, function, and mutagenesis of a female-specific domain. *J. Biol. Chem.,* **283,** 7280-7292.

234. Zhou,Z.R., Gao,H.C., Zhou,C.J., Chang,Y.G., Hong,J., Song,A.X., Lin,D.H., and Hu,H.Y. (2008). Differential ubiquitin binding of the UBA domains from human c-Cbl and Cbl-b: NMR structural and biochemical insights. *Protein Sci.,* **17,** 1805-1814.

235. Kozlov,G., Peschard,P., Zimmerman,B., Lin,T., Moldoveanu,T., Mansur-Azzam,N., Gehring,K., and Park,M. (2007). Structural basis for UBA-mediated dimerization of c-Cbl ubiquitin ligase. *J. Biol. Chem.,* **282,** 27547-27555.

236. Peschard,P., Kozlov,G., Lin,T., Mirza,I.A., Berghuis,A.M., Lipkowitz,S., Park,M., and Gehring,K. (2007). Structural basis for ubiquitin-mediated

dimerization and activation of the ubiquitin protein ligase Cbl-b. *Mol. Cell,* **27,** 474-485.

237. Sims,J.J., Haririnia,A., Dickinson,B.C., Fushman,D., and Cohen,R.E. (2009). Avid interactions underlie the Lys63-linked polyubiquitin binding specificities observed for UBA domains. *Nat. Struct. Mol. Biol.,* **16,** 883-889.

238. Kozlov,G., Nguyen,L., Lin,T., De Crescenzo,G., Park,M., and Gehring,K. (2007). Structural basis of ubiquitin recognition by the ubiquitin-associated (UBA) domain of the ubiquitin ligase EDD. *J. Biol. Chem.,* **282,** 35787-35795.

239. Swanson,K.A., Hicke,L., and Radhakrishnan,I. (2006). Structural basis for monoubiquitin recognition by the Ede1 UBA domain. *J. Mol. Biol.,* **358,** 713-724.

240. Saio,T., Yokochi,M., and Inagaki,F. (2009). The NMR structure of the p62 PB1 domain, a key protein in autophagy and NF-κB signaling pathway. *J. Biomol. NMR,* **45,** 335-341.

241. Moscat,J., Diaz-Meco,M.T., and Wooten,M.W. (2009). Of the atypical PKCs, Par-4 and p62: recent understandings of the biology and pathology of a PB1-dominated complex. *Cell Death. Differ.,* **16,** 1426-1437.

242. Saio,T., Yokochi,M., Kumeta,H., and Inagaki,F. (2010). PCS-based structure determination of protein-protein complexes. *J. Biomol. NMR,* **46,** 271-280.

243. Pedersen,S., Bloch,P.L., Reeh,S., and Neidhardt,F.C. (1978). Patterns of protein synthesis in *E. coli*: a catalog of the amount of 140 individual proteins at different growth rates. *Cell,* **14,** 179-190.

244. Carlson,N. and Rechsteiner,M. (1987). Microinjection of ubiquitin: intracellular distribution and metabolism in HeLa cells maintained under normal physiological conditions. *J. Cell Biol.,* **104,** 537-546.

245. Pielak,G.J., Li,C., Miklos,A.C., Schlesinger,A.P., Slade,K.M., Wang,G.F., and Zigoneanu,I.G. (2009). Protein nuclear magnetic resonance under physiological conditions. *Biochemistry,* **48,** 226-234.

246. Li,C., Charlton,L.M., Lakkavaram,A., Seagle,C., Wang,G., Young,G.B., Macdonald,J.M., and Pielak,G.J. (2008). Differential dynamical effects of macromolecular crowding on an intrinsically disordered protein and a globular protein: implications for in-cell NMR spectroscopy. *J. Am. Chem. Soc.,* **130,** 6310-6311.

247. Hu,Q., Shen,W., Huang,H., Liu,J., Zhang,J., Huang,X., Wu,J., and Shi,Y. (2007). Insight into the binding properties of MEKK3 PB1 to MEK5 PB1 from its solution structure. *Biochemistry,* **46,** 13478-13489.

248. Rea,S.L., Walsh,J.P., Ward,L., Yip,K., Ward,B.K., Kent,G.N., Steer,J.H., Xu,J., and Ratajczak,T. (2006). A novel mutation (K378X) in the

sequestosome 1 gene associated with increased NF-κB signaling and Paget's disease of bone with a severe phenotype. *J. Bone Miner. Res.,* **21,** 1136-1145.

249. Rea,S.L., Walsh,J.P., Ward,L., Magno,A.L., Ward,B.K., Shaw,B., Layfield,R., Kent,G.N., Xu,J., and Ratajczak,T. (2009). Sequestosome 1 mutations in Paget's disease of bone in Australia: prevalence, genotype/phenotype correlation, and a novel non-UBA domain mutation (P364S) associated with increased NF-κB signaling without loss of ubiquitin binding. *J. Bone Miner. Res.,* **24,** 1216-1223.

250. Raiborg,C. and Stenmark,H. (2009). The ESCRT machinery in endosomal sorting of ubiquitylated membrane proteins. *Nature,* **458,** 445-452.

251. Lange,O.F., Lakomek,N.A., Fares,C., Schroder,G.F., Walter,K.F., Becker,S., Meiler,J., Grubmuller,H., Griesinger,C., and de Groot,B.L. (2008). Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science,* **320,** 1471-1475.

252. Wlodarski,T. and Zagrovic,B. (2009). Conformational selection and induced fit mechanism underlie specificity in noncovalent interactions with ubiquitin. *Proc. Natl. Acad. Sci. U. S. A*. **106**, 19346-19351.

253. Loladze,V.V. and Makhatadze,G.I. (2002). Removal of surface charge-charge interactions from ubiquitin leaves the protein folded and very stable. *Protein Sci.,* **11,** 174-177.

254. Zhang,N., Wang,Q., Ehlinger,A., Randles,L., Lary,J.W., Kang,Y., Haririnia,A., Storaska,A.J., Cole,J.L., Fushman,D., and Walters,K.J. (2009). Structure of the s5a:k48-linked diubiquitin complex and its interactions with rpn13. *Mol. Cell,* **35,** 280-290.

255. Layfield,R., Ciani,B., Ralston,S.H., Hocking,L.J., Sheppard,P.W., Searle,M.S., and Cavey,J.R. (2004). Structural and functional studies of mutations affecting the UBA domain of SQSTM1 (p62) which cause Paget's disease of bone. *Biochemical Society Transactions,* **32,** 728-730.

256. Lamark,T. and Johansen,T. (2010). Autophagy: links with the proteasome. *Curr. Opin. Cell Biol.,* **22,** 192-198.

257. Carlson,J.C., Kanter,A., Thuduppathy,G.R., Cody,V., Pineda,P.E., McIvor,R.S., and Wagner,C.R. (2003). Designing protein dimerizers: the importance of ligand conformational equilibria. *J. Am. Chem. Soc.,* **125,** 1501-1507.

258. Lin,C.T., Moore,P.A., Auberry,D.L., Landorf,E.V., Peppler,T., Victry,K.D., Collart,F.R., and Kery,V. (2006). Automated purification of recombinant proteins: combining high-throughput with high yield. *Protein Expr. Purif.,* **47,** 16-24.

259. Kumar,T.R., Prakash,V., and Muralidhar,K. (1990). Studies on buffalo pituitary lutropin (LH): physicochemical and immunological properties. *J. Protein Chem.,* **9,** 727-733.

260. Colbert,J.D., Plechanovova,A., and Watts,C. (2009). Glycosylation directs targeting and activation of cystatin f from intracellular and extracellular sources. *Traffic.,* **10,** 425-437.

261. Hiraishi,H., Yokoi,F., and Kumon,A. (1999). Bovine liver phosphoamidase as a protein histidine/lysine phosphatase. *J. Biochem.,* **126,** 368-374.

262. Wilkinson,K.D., Cox,M.J., O'Connor,L.B., and Shapira,R. (1986). Structure and activities of a variant ubiquitin sequence from bakers' yeast. *Biochemistry,* **25,** 4999-5004.

# Appendix A : Gene sequences

p62/SQSTM1 (*Homo sapien*)

ATGGCGTCGCTCACCGTGAAGGCCTACCTTCTGGGCAAGGAGGACGCGGCGCGCGAGATTCGCCGCTT
CAGCTTCTGCTGCAGCCCCGAGCCTGAGGCGGAAGCCGAGGCTGCGGCGGGTCCGGGACCCTGCGAGC
GGCTGCTGAGCCGGGTGGCCGCCCTGTTCCCCGCGCTGCGGCCTGGCGGCTTCCAGGCGCACTACCGC
GATGAGGACGGGGACTTGGTTGCCTTTTCCAGTGACGAGGAATTGACAATGGCCATGTCCTACGTGAA
GGATGACATCTTCCGAATCTACATTAAAGAGAAAAAAGAGTGCCGGCGGGACCACCGCCCACCGTGTG
CTCAGGAGGCGCCCCGCAACATGGTGCACCCCAATGTGATCTGCGATGGCTGCAATGGGCCTGTGGTA
GGAACCCGCTACAAGTGCAGCGTCTGCCCAGACTACGACTTGTGTAGCGTCTGCGAGGGAAAGGGCTT
GCACCGGGGGCACACCAAGCTCGCATTCCCCAGCCCCTTCGGGCACCTGTCTGAGGGCTTCTCGCACA
GCCGCTGGCTCCGGAAGGTGAAACACGGACACTTCGGGTGGCCAGGATGGGAAATGGGTCCACCAGGA
AACTGGAGCCCACGTCCTCCTCGTGCAGGGGAGGCCCGCCCTGGCCCCACGGCAGAATCAGCTTCTGG
TCCATCGGAGGATCCGAGTGTGAATTTCCTGAAGAACGTTGGGGAGAGTGTGGCAGCTGCCCTTAGCC
CTCTGGGCATTGAAGTTGATATCGATGTGGAGCACGGAGGGAAAAGAAGCCGCCTGACCCCCGTCTCT
CCAGAGAGTTCCAGCACAGAGGAGAAGAGCAGCTCACAGCCAAGCAGCTGCTGCTCTGATCCCAGCAA
GCCGGGTGGGAATGTTGAGGGCGCCACGCAGTCTCTGGCGGAGCAGATGAGAAAGATCGCCTTGGAGT
CCGAGGGGCGCCCTGAGGAACAGATGGAGTCGGATAACTGTTCAGGAGGAGATGATGACTGGACCCAT
CTGTCTTCAAAAGAAGTGGACCCGTCTACAGGTGAACTCCAGTCCCTACAGATGCCAGAATCCGAAGG
GCCAAGCTCTCTGGACCCCTCCCAGGAGGGACCCACAGGGCTGAAGGAAGCTGCCTTGTACCCACATC
TCCCGCCAGAGGCTGACCCGCGGCTGATTGAGTCCCTCTCCCAGATGCTGTCCATGGGCTTCTCTGAT
GAAGGCGGCTGGCTCACCAGGCTCCTGCAGACCAAGAACTATGACATCGGAGCGGCTCTGGACACCAT
CCAGTATTCAAAGCATCCCCCGCCGTTGTGA

ZNF216 (*Rattus norvegicus*)

ATGGCTCAGGAGACTAACCAGACCCCAGGGCCCATGCTGTGTAGTACTGGATGTGGCTTTTATGGGAA
TCCTAGGACAAATGGAATGTGTTCTGTTTGCTACAAAGAACATCTTCAGAGACAGCAGAATAGTGGCA
GAATGAGCCCAATGGGGACAGCTAGTGGTTCCAACAGTCCTACCTCAGACTCTGCGTCTGTACAAAGA
GCAGATGCTACTTTAAACAACTGTGAAGGTGCTGCTGGCAGCACATCTGAAAAATCAAGAAATGTGCC
TGTGGCTGCCTTGCCTGTAACTCAACAAATGACAGAAATGAGCATTTCAAGAGAGGACAAAATAACCT
CCCCGAAAACAGAGGTGTCAGAGCCAGTTGTCACTCAGCCCAGTCCATCAGTTTCTCAGCCCAGTTCT
TCTCAAAGTGAAGAAAAAGCTCCTGAGTTGCCCAAACCAAAGAAGAACAGATGTTTTATGTGTAGAAA
GAAAGTTGGCCTTACAGGGTTTGACTGCCGATGTGGAAATTTGTTTTGTGGACTTCACCGTTACTCTG
ACAAGCACAACTGTCCTTATGATTACAAAGCAGAAGCTGCAGCAAAAATCAGAAAAGAAAATCCAGTT
GTTGTGGCTGAAAAAATCCAGAGAATATAA

Ubiquitin (*Saccharomyces cerevisiae*)

ATGCAGATCTTCGTCAAGACGTTAACCGGTAAAACCATAACTCTAGAAGTTGAATCTTCCGATACCAT
CGACAACGTTAAGTCGAAAATTCAAGACAAGGAAGGCATTCCACCTGATCAACAAAGATTAATCTTTG
CCGGTAAGCAGCTCGAGGACGGTAGAACGCTGTCTGATTACAACATTCAGAAGGAGTCGACCTTACAT
CTTGTCTTAAGACTAAGAGGTGGTTGA

# Appendix B : Primer Sequences

**Ubiquitin K48C**

'Forward': 5'-GTTCCAACAGTCCTACCTAAGACTCTGCGTCTGTA-3'

'Backward': 5'-TACAGACGCAGAGTCTTAGGTAGGACTGTTGGAAC-3'

**ZNF216 S61X (Stop)**

'Forward': 5'-GTTCCAACAGTCCTACCTAAGACTCTGCGTCTGTA-3'

'Backward': 5'-TACAGACGCAGAGTCTTAGGTAGGACTGTTGGAAC-3'

**Znf_A20 F22Y**

'Forward': 5'-GTACTGGATGTGGCTATTATGGGAATCCTAG-3'

'Backward': 5'-CTAGGATTCCCATAATAGCCACATCCAGTAC-3'

**Znf_A20 V34K**

'Forward':

5'-CCTAGGACAAATGGAATGTGTTCTAAGTGCTACAAAGAACATCTTCAGAG-3'

'Backward':

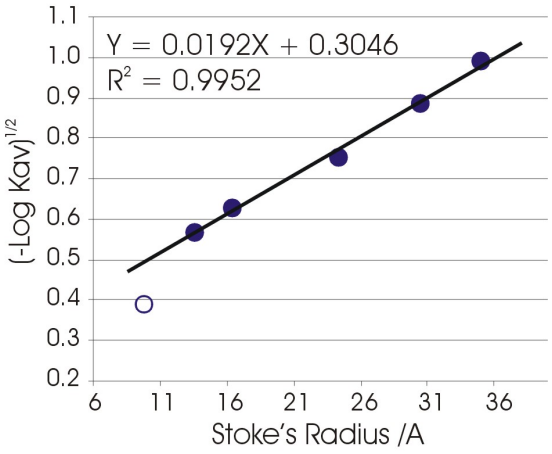
5'-CTCTGAAGATGTTCTTTGTAGCACTTAGAACACATTCCATTTGTCCTAGG-3'

# Appendix C : Calculation of Stoke's radius

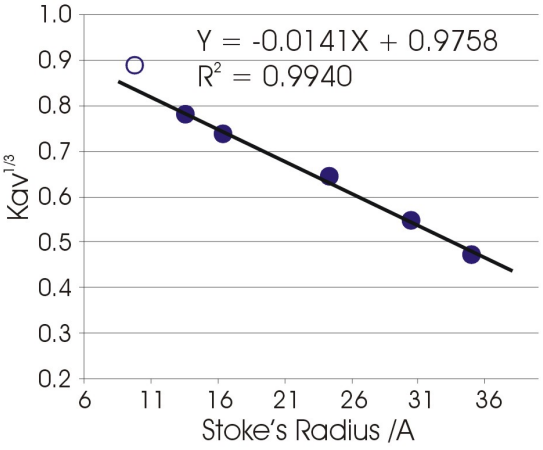
A.

| Protein | Mwt /kDa | Stoke's Radius /Å |
|---|---|---|
| Aprotinin | 6.6 | 9.83[257] |
| Ribonuclease A | 13.7 | 16.4[258] |
| Carbonic anhydrase | 29 | 24.3[259,260] |
| Ovalbumin | 43 | 30.5[258] |
| Conalbumin | 75 | 40.4[260] |
| Ubiquitin | 8.5 | 13.6[261,262] |

B.



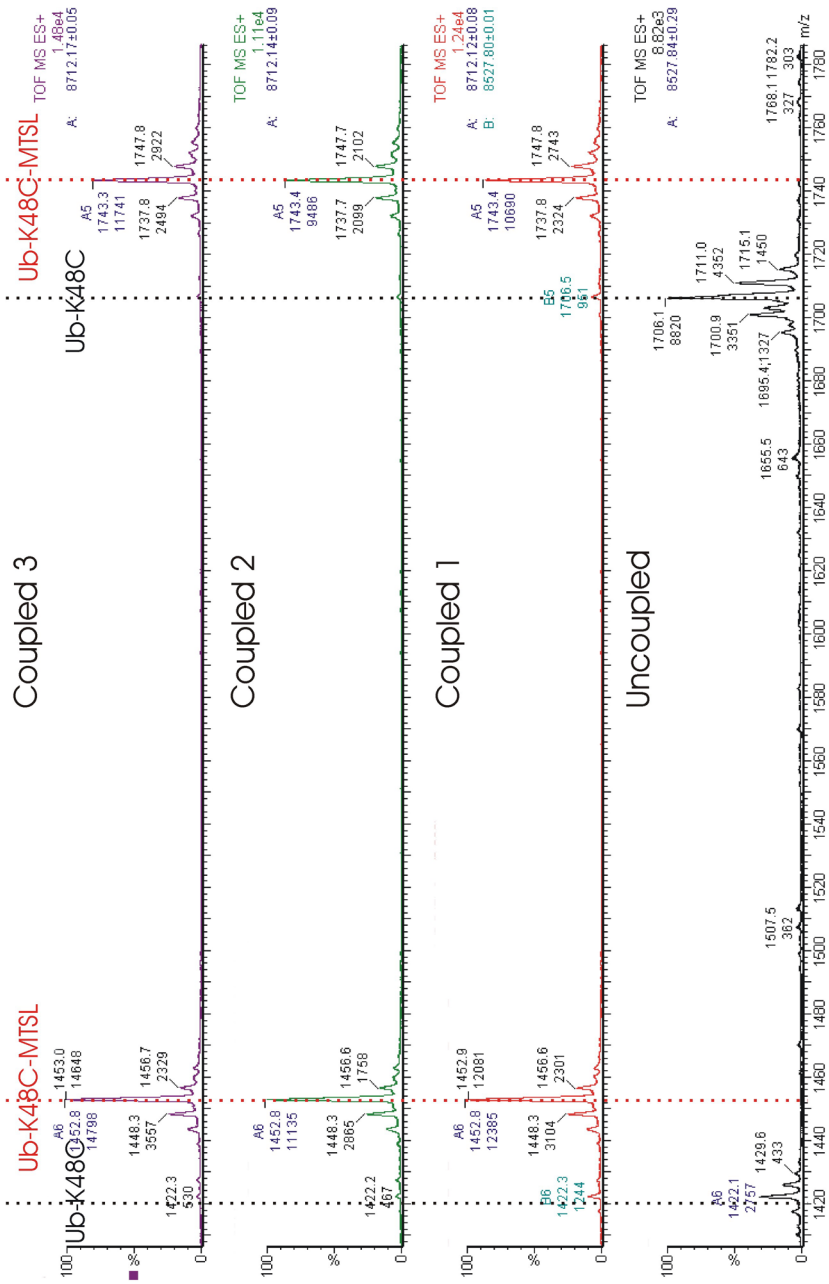

C.




Stoke's radius was calculated from a set of known calibration proteins (A) used to construct linear plots of Stoke's radius against √-(Log Kav) (B) or $^{3}\sqrt{}$Kav, the calibration protein Aprotinin (open circle) deviated from linearity and was excluded from the fit.

# Appendix D : ESI-MS of Ub-K48C-MTSL Coupling

ESI-MS of MTSL coupled K48C-Ub showing over 90% incorporation



# Appendix E : Dynafit Input Script

Input script describing a competitive equilibrium between dimerisation and binding.

```
Model for UBA domerisation

;_____________________________________________________________________
_____

[task]

   data  = equilibria
   task  = fit
   model = fixed ?

[mechanism]

   P + L <==> PL        :  K1   dissoc
   L + L <==> LL        :  K2   dissoc

[constants]

   K1 = 0.05 ?
   K2 =  0.007

[concentrations]

   P = 1

[responses]

   PL = 0.22 ?

[equilibria]

   variable     L
   file         ./examples/enzymology/DNA+promoter/data/Wt44.txt

[output]

   directory    ./examples/enzymology/DNA+promoter/output/Wt/44


[end]
```

# Publications

Publications released/prepared during the duration of the PhD of T. Garner, Publications relevant to the work presented here are marked with a '*'

1. G. D. Balkwill, T. P. Garner, H. E. L. Williams, and M. S. Searle, Folding topology of a bimolecular DNA quadruplex containing a stable mini-hairpin motif within the diagonal loop, *J.Mol.Biol*. **385**:1600 (2009).
2. G. D. Balkwill, T. P. Garner, and M. S. Searle, Folding of single-stranded DNA quadruplexes containing an autonomously stable mini-hairpin loop, *Mol.Biosyst*. **5**:542 (2009).
3. *D. Najat, T. P. Garner, T. Hagen, B. Shaw, P. W. Sheppard, A. Falchetti, F. Marini, M. L. Brandi, J. E. Long, J. R. Cavey, M. S. Searle, and R. Layfield, Characterization of a non-UBA domain missense mutation of sequestosome 1 (SQSTM1) in Paget's disease of bone, *J.Bone Miner.Res*. **24**:632 (2009).
4. T. P. Garner, H. E. L. Williams, K. Gluszyk, S. Roe, N. J. Oldham, M. F. G. Stevens, J. E. Moses and M. S. Searle. Selectivity of small molecule ligands for parallel and anti-parallel DNA G-quadruplex structures. *Org. Biomol. Chem*. **20**:4194 (2009).
5. G. D. Balkwil, K. Derecka, T. P. Garner, C. Hodgman, A. P. F. Flint and M. S. Searle, Repression of translation of human estrogen receptorα by G-quadruplex formation. *Biochemistry*. **48**:11487 (2009).
6. *J. Long, T. P. Garner, M. Pandya, P. Chen, B. Shaw, M. P. Williamson, C. J. Craven, R. Layfield and M. S. Searle, Dimerisation of the UBA domain of p62 inhibits ubiquitin binding and regulates NF-kB signalling, *J.Mol.Biol*. **396**:178 (2010).
7. S. Roe, D. J. Ritson, T. P. Garner, M. S. Searle and J. E. Moses, Tuneable DNA-based asymmetric catalysis using a G-quadruplex supramolecular assembly. *Chem. Comm.* **46**:4309 (2010).
8. *C. Heinen, T. P. Garner, J. Long, C. Böttcher, J. R. Cavey, M. S. Searle, R. Layfield and N. P. Dantuma, Mutant p62/SQSTM1 UBA domains linked to Paget's disease of bone differ in their abilities to function as stabilization signals. *FEBS lett.* **584**: 1585 (2010).
9. K. Derecka, G. D. Balkwil, T. P. Garner, A. P. F. Flint and M. S. Searle, A unique G-quadruplex motif located within the promoter sequence of the bovine estrogen receptorα gene (ESR1). *Biochemistry*. **49**:7625 (2010).
10. J. Linder, T. P. Garner, H. E. L. Williams, M. S. Searle and C. J. Moody, Telomestatin: formal total synthesis and the potent stabilizing activity of its seco-derivatives for the human intramolecular G-quadruplex. *JACS* 133, 1044 (2011).
11. *T. P. Garner, J. Strachan, E. C. Shedden, J. Long, J. Cavey, R. Layfield and M. S. Searle, ZNF216 and P62 receptors linked to NF-κB signal transduction are co-localised through formation of a ubiquitin-mediated ternary complex .(Submitted to Nature Structural and Molecular Biology June 2010)
12. M. G. Hummersone, T. P. Garner, W. D. Wilson, C. Hutton, I. Hutchinson, H. E. L. Williams, M. S. Searle and M. F. G. Stevens, On and off target effects of telomere-uncapping G-quadruplex-selective ligands based on a pentacyclic acridinium salt. (In Prep)
13. *T. P. Garner, R. Layfield, J. Long, and M. S. Searle, Impact of p62/SQSTM1 UBA domains mutations linked to Paget's disease of bone on ubiquitin recognition. (Submited to Biochemistry, in revision march 2011)