

Characterisation of the deubiquitinating enzyme USP20

Simon Caulton

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

September 2016

Abstract

USP20 is a deubiquitinating enzyme that is involved in a number of important cellular pathways, including thyroid metabolism, hypoxic response, seven transmembrane receptor signalling, NF- κ B signalling, centrosome homeostasis and DNA repair. Of recent, it is becoming a major deubiquitinase involved in regulating the DNA-damage response pathway and cell cycle checkpoints.

The protein consists of a zinc finger domain, catalytic domain and two 'domain present in USP' (DUSP) domains; an architecture shared only with its paralogue USP33. There is no structural information on any of the domains of USP20, so crystallisation trials of the domains of USP20 were performed in order to solve their structures by X-ray crystallography. In addition, yeast two-hybrid (Y2H) and *in vitro* assays were used to further characterise known and putative interactors of USP20. Finally, the zinc finger domain and DUSP domains were used in pull down assays to identify USP20-interacting proteins from HEK293 lysate.

Two stable and well-expressing constructs of the zinc finger domain (USP20 1-101 and 1-108) were purified and set up for crystallisation trials. Buffer screens were also performed on the USP20 1-101 construct to increase its stability for crystallisation. Monodisperse, pure protein of any catalytic domain-containing construct of USP20 was unobtainable; only a trigger factor-tagged full length USP20 was purified and active. Two constructs containing the double DUSP domains were produced (USP20 686-914 and 686-894), and both suffered from a low solubility limit. Buffer screening was used to increase its stability, which identified ethylene glycol as a stabilising additive. Due to the nature of commonly used solubility tags, novel tags were

designed that would potentially benefit the crystallisation of the fusion construct. Identified from the PDB and literature searches, the calponin homology domain from human β -spectrin (PDB code 1BKR) and the receiver domain from *Myxococcus xanthus* social motility protein frzS (PDB code 2GKG) were used. Both new tags, as well as MBP were fused to the N-terminus of the DUSP domains (USP20 686-894) to enhance solubility and crystallise the DUSP domains. 2GKG was an effective solubility tag, increased the solubility of the DUSP domains to near that of the MBP fusion. 1BKR, however, was only marginally useful as a solubility tag. In total 97 crystallisation trials were set up for all constructs of USP20, but no crystals containing USP20 protein formed.

Y2H assays were used to investigate the interaction between USP20 domains and B-arrestin-1, TRAF6, RAD17 and PLK1. Of these, only an interaction between USP20's DUSP domains (residues 686-894) and full length PLK1 was observed. Interestingly, further Y2H and ELISA showed a non-canonical, binary interaction between the polo-boxes of PLK1 (residues 367-603) and the DUSP domains. Pull down assays produced a list of possible novel interactors for USP20. These include proteins implicated in processes known, and unknown, to involve USP20. Finally, using ELISA, thermal shift assays and ITC, it was shown that the zinc finger domain of USP20 does not bind to ubiquitin.

Acknowledgments

I would like to thank my supervisor Dr Ingrid Dreveny for giving me the amazing opportunity to be part of the field of structural biology. I am incredibly grateful for Ingrid, and Dr Steve Harper, for teaching me everything I know, and I am incredibly lucky to have had the chance to work with these two people! In addition, Prof Jonas Emsley and Cornelia de Moor have been invaluable thanks to their input and guidance throughout my PhD. I would also like to thank all the people who have been in the lab over the years; their discussion, assistance, friendship and camaraderie have kept me ploughing through, even at times of severe adversity. I would like to thank my collaborators David Heery and Joel Fulton for their expertise with the all-important yeast two-hybrid assays. I am also grateful to Hillary Collins for her guidance with the fluorescent microscopy, which unfortunately never made it into the thesis.

It is also essential that I thank my partner Kayleigh Langham and mother Kathryn Caulton for keeping me alive over the last 4 years. Without them I would not have been able to complete the PhD, and for that I am forever indebted.

Contents

Characterisation of the deubiquitinating enzyme USP20.....	i
Abstract	ii
Acknowledgments	iv
Contents	v
List of Figures.....	ix
List of Tables.....	x
Abbreviations.....	xi
1.....	Introduction: literature
.....	1
1.1 The importance of protein regulation	2
1.2 Ubiquitination	4
1.2.1 A brief history of ubiquitin	4
1.2.2 The cascade to ubiquitination.....	6
1.2.3 The ubiquitin code	10
1.3 Deubiquitination	13
1.3.1 DUB regulation.....	19
1.4 USP20 and USP33.....	22
1.4.1 Thyroid metabolism	26
1.4.2 Hypoxic response	27
1.4.3 Seven-transmembrane receptor signalling.....	29
1.4.4 NF- κ B signalling	31
1.4.5 Centrosome homeostasis	33
1.4.6 DNA damage response.....	34
1.5 Phosphorylation	36
1.5.1 A brief history of phosphorylation.....	36
1.5.2 Kinase structure and function.....	37
1.6 Aims.....	42
1.7 Objectives.....	42
2.....	Introduction: Methods
.....	44
2.1 Protein design and expression.....	45
2.1.1 Protein tags	45
2.2 Protein purification	50
2.2.1 Affinity chromatography.....	50

2.2.2	Size Exclusion Chromatography	51
2.3	Crystallisation and X-ray-diffraction	53
2.3.1	X-ray crystallography history	54
2.3.2	Obtaining crystals.....	55
2.3.3	Cryocooling	58
2.3.4	The crystal and symmetry.....	58
2.3.5	X-rays and diffraction.....	60
2.4	Hybrid assays	65
2.5	ITC	69
2.6	Thermofluor	74
3.....	Materials and methods	76
3.1	Cloning, mutagenesis and splicing by overlap extension PCR	77
3.1.1	PCR	77
3.1.2	Restriction digestion	78
3.1.3	Ligation.....	79
3.1.4	Transformation	80
3.1.5	Cracking and analytical digest.....	80
3.1.6	Site-directed mutagenesis	81
3.1.7	Cloning for catalytic domain-containing constructs.	82
3.1.8	Summary of primers used.....	86
3.2	Protein expression	88
3.3	Protein purification	89
3.3.1	Lysate preparation	89
3.3.2	Nickel affinity chromatography	89
3.3.3	Anion exchange.....	90
3.3.4	Cation exchange.....	91
3.3.5	Size exclusion chromatography	91
3.3.6	Hydrophobic interaction chromatography	92
3.4	SDS-PAGE	92
3.5	Western blot	94
3.6	Analysing protein concentration.....	96
3.7	Crystallisation.....	96
3.8	X-ray diffraction	97
3.9	DUSP domains solubility assays	98
3.9.1	Native PAGE assay.....	98

3.9.2	Cell lysis assay	99
3.9.3	Buffer compositions	100
3.10	Thermofluor	101
3.11	Ubiquitin-AMC assays	103
3.12	Yeast two hybrid	104
3.12.1	Yeast media	104
3.12.2	Genotype	105
3.12.3	Transformation	105
3.12.4	B-galactosidase assay	106
3.13	ELISA	107
3.14	ITC	109
3.15	Far-western blot	110
3.16	Mammalian lysate	110
3.17	Pull down assays	112
4Results: expression, purification and crystallisation	114
4.1	Construct design	115
4.1.1	USP20 bioinformatics	115
4.1.2	Znf-UBP domain design	122
4.1.3	Catalytic domain design	124
4.1.4	DUSP domains	128
4.2	Cloning and mutagenesis	131
4.3	Purification and crystallisation of USP20 domains	134
4.3.1	Znf-UBP	134
4.3.2	Catalytic domains	144
4.3.3	DUSP domains	160
4.4	Thermal shift assays for USP20 domain stabilisation	172
4.5	Novel solubility tag design	178
4.5.1	Identifying PDB proteins	178
4.5.2	Modifying the proteins	183
4.5.3	Tag purifications	189
4.5.4	DUSP domains-solubility tag fusions	193
4.5.5	Comparison of novel tags to other tags	206
4.6	Summary of crystallisation trials	208
5Results: Interaction studies	210

Mapping USP20 protein interactions.....	211
5.1 Investigating USP20 interactions with yeast two hybrid	211
.....	229
5.1.2 <i>In vitro</i> interaction assays	229
5.2 Pull down assays	238
5.2.1 Gene ontology.....	246
5.3 USP20 Znf-UBP and ubiquitin	249
6.....	Discussion
.....	253
6.1 Structural studies of USP20	254
6.2 Interaction of USP20 and PLK1	259
6.3 Binding partners and processes.....	264
6.4 Ubiquitin binding	266
7.....	Further studies
.....	269
8.....	References
.....	271

List of figures

Figure	Page	Figure	Page
1.1 Post translational modifications	3	4.25 Small scale catalytic domain-containing construct expressions	155
1.2 Ubiquitin	5	USP20 Δ insert purification	157
1.3 UBA1	7	Purification of USP20 Δ Znf-UBP	158
1.4 UBA1-Ubc4 complex	8	USP20 Δ Znf-UBP Ub-AMC assay	159
1.5 E2-E3 complexes	9	USP20 DUSP domains constructs	160
1.6 Ubiquitin's lysines	10	Purification of USP20 686-914	162
1.7 Polyubiquitin chains	11	Purification of USP20 686-894	164
1.8 The USP domain	14	4.32 Native PAGE additive screen	166
1.9 Ubiquitin specific proteases 1	16	4.33 Stability screening USP20 686-894	168
1.10 Ubiquitin specific proteases 2	17	4.34 Limited proteolysis of USP20 686-914	171
1.11 The catalytic mechanism of USPs	18	4.35 Thermofluor plots	173
1.12 USP7 cryptic catalytic site	21	4.36 Effect of salt concentration on T_m of Znf-UBP domain and DUSP domains	174
1.13 USP20 and USP33	23	4.37 Effect of additives and buffers on T_m of the Znf-UBP domain	176
1.14 The topology of USP20 and USP33	23	4.38 Effect of additives and buffers on T_m of the DUSP domains	177
1.15 Regulation of DIO2	26	4.39 1BKR and 2GKG	181
1.16 CRL2 ^{PHL} complex	27	4.40 1BKR and 2GKG B-factors and electrostatics	182
1.17 VHL-HIF1- α peptide complex	28	4.41 Evaluation of 1BKR SERp clusters	185
1.18 Receptor recycling	30	4.42 Evaluation of 2GKG SERp clusters	185
1.19 Tax and NF- κ B	32	4.43 Upstream and downstream sequences of the solubility tag vectors	188
1.20 USP20 and DNA damage response	35	4.44 Maps of the solubility tag vectors	189
1.21 PLK1 catalytic domain	39	4.45 1BKR and 2GKG gel filtration	191
1.22 PKK1 poloboxes	41	4.46 Crystals of the solubility tags	192
2.1 Fusion structures	49	4.47 USP20 DUSP domains fusion constructs	193
2.2 Diffraction pattern of crystallised 3Clpro	55	4.48 MBP-DUSPs gel filtration	195
2.3 Vapour drop diffusion and the phase diagram	57	4.49 MBP-DUSPs extended purification	197
2.4 Bravais lattices	60	4.50 1BKR-DUSPs purification	199
2.5 Bragg's Law	62	4.51 2GKG-DUSPs purification	201
2.6 Structure factors	63	4.52 Crystals from 2GKG-DUSP crystallisation trials	204
2.7 Phases and amplitudes	64	4.53 Diffraction and assessment of 2GKG-DUSP crystals	205
2.8 Yeast two-hybrid	66	5.1 DUSP domains transcriptional activity	213
2.9 Isothermal titration calorimeter	70	5.2 Review of USP20 domain architecture	214
2.10 Example data from ITC	72	5.3 β -arrestin-1 yeast two-hybrid	216
2.11 Thermofluor	75	5.4 TRAF6 yeast two-hybrid	217
3.1 Principle of splicing by overlap PCR	83	5.5 RAD17 yeast two-hybrid	219
3.2 Thermofluor screening conditions	102	5.6 PLK1 domain architecture	220
4.1 Domains of USP20	116	5.7 PLK1 yeast two-hybrid	220
4.2 Phylogeny of species and inserted domains	117	5.8 Polobox truncation	221
4.3 Multiple species sequence alignment of USP20	119	5.9 PK1-USP20 yeast two-hybrid domain truncation	222
4.4 Alignment of USP20 and USP33	120	5.10 Docking studies of DUSP domains and PLK1 poloboxes	225
4.5 Disorder plots of USP20	121	5.11 PLK1-USP20 yeast two-hybrid mutation assays	226
4.6 Topology of USP20	122	5.12 PLK1-USP20 yeast two hybrid summary	228
4.7 Znf-UBP domains structure and boundaries	123	5.13 Yeast two-hybrid Western blots	229
4.8 Catalytic domain N-terminus	125	5.14 MBP-poloboxes gel filtration and cleavage	231
4.9 Catalytic domain alignments	126	5.15 Poloboxes contamination test	232
4.10 Structural alignments of Box 2/3	127	5.16 DUSPs-MBP-poloboxes pulldown	233
4.11 Sequence alignments	129	5.17 DUSP domains-poloboxes ELISA	235
4.12 Homology models of the DUSP domains of USP20	130	5.18 DUSP domains-poloboxes far-Western	236
4.13 Summary of expression constructs for crystallisation	133	5.19 TF-USP20-PLK1 ELISA	237
4.14 USP20 Znf-UBP domain constructs	134	5.20 USP20 pulldown assays	241
4.15 Purification of Znf-UBP 1-101	136	5.21 Gene ontology clusters 1	247
4.16 Small scale expressions of Znf-UBP domains	139	5.22 Gene ontology clusters 2	248
4.17 Purification of Znf-UBP 6-108	140	5.23 Purified ubiquitin	249
4.18 Nickel column of Znf-UBP 1-108	142	5.24 Znf-UBP and ubiquitin Thermofluor and ELISA	251
4.19 Gel filtration and cleavage of Znf-UBP 1-108	143	5.25 Znf-UBP and ubiquitin ITC	252
4.20 USP20 catalytic domain constructs	144	6.1 Proposed model for USP20-PLK1 interaction	263
4.21 Purification of USP20FL	147	6.2 Ubiquitin-Znf-UBPs interactions	267
4.22 USP20FL Ub-AMC	149	6.3 Binding pockets of Znf-UBPs	268
4.23 Nickel column and anion exchange of TF-USP20FL	151		
4.24 TF-USP20FL gel filtration and Ub-AMC assays	153		

List of tables

Table	Page
2.1 Commonly used fusion tags	47
3.1 PCR protocol	77
3.2 Splicing by overlap PCR primers	83
3.3 Overlap PCR	84
3.4 Cloning of USP20 <i>E. coli</i> expression constructs	86
3.5 Cloning of solubility tag vectors	87
3.6 Cloning of yeast two-hybrid constructs	88
3.7 Expression conditions	92
3.8 Gel filtration columns	93
3.9 SDS-PAGE gel recipes	98
3.10 Native PAGE gel recipes	100
3.11 Screening conditions	104
3.12 Yeast media	116
4.1 Domain boundary predictions	131
4.2 <i>E. coli</i> expression constructs	132
4.3 Solubility tag vectors	160
4.4 Yeast two-hybrid constructs	183
4.5 Summary of catalytic domain-containing constructs	208
4.6 SERp cluster suggestions	209
4.7 Summary of crystallisation trials with His-tagged USP20 constructs	224
4.8 Summary of crystallisation trials with solubility-tagged USP20 constructs	242
5.1 Polobox binding motifs in USP20	243
5.2 Interactors from Znf-UBP domain band 1	244
5.3 Interactors from Znf-UBP domain band 2	245
5.4 Interactors from Znf-UBP domain band 3	247
5.5 Interactors from Znf-UBP domain band 4	47
5.6 Interactors from DUSP domains band 1	77

Abbreviations

1BKR	Calponin homology domain of human β -spectrin	PEG	Polyethylene glycol
2GKG	Receiver domain of frzS	PLK	Pololike kinase
AMP	Adenosine monophosphate	PTM	Post translational modification
APC	Anaphase promoting complex	PVDF	Polyvinylidene fluoride
ATP	Adenosine triphosphate	RING	Really interesting new gene
cAMP	Cyclic adenosine monophosphate	RIPA	Radioimmunoprecipitation assay buffer
COSMIC	Catalogue of somatic mutations in cancer	SD	Synthetic derived
CRL	Cullin RING ligase	SERp	Surface entropy reduction prediction
CV	Column volume	T3	3,5,4'-triiodothyronine
DBD	DNA binding domain	T4	Thyroxine
DDR	DNA damage response	TAD	Transcription activating domain
DTT	Dithiothreitol	TBS	Tris-buffered saline
DUB	Deubiquitinating enzyme	TBST	TBS-tween
DUSP	Domain present in USPs	TCEP	Tris(2-carboxyethyl)phosphine hydrochloride
E1	Ubiquitin activating enzyme	TEV	Tobacco etch virus
E2	Ubiquitin conjugating enzyme	TF	Trigger factor
E3	Ubiquitin ligase	T _m	Melting temperature
EDTA	Ethylenediaminetetraacetic acid	TRAF6	TNF-receptor associated factor 6
GST	Glutathione S transferase	UAS	Upstream activating sequence
HECT	Homologous to E6AP C-terminus	Ub	Ubiquitin
HEK	Human embryonic kidney	Ub-AMC	Ubiquitin-aminomethylcoumarin
HIC	Hydrophobic interaction chromatography	Ubl	Ubiquitin-like
HRP	Horseradish peroxidase	UFD	Ubiquitin fold domain
HTLV	Human T-lymphotropic leukaemia virus	USP	Ubiquitin specific protease
IPTG	Isopropyl β -D-1-thiogalactopyranoside	VHL	von Hippel Lindau
ITC	Isothermal titration calorimetry	WB	Western blot
LTR	Long terminal repeat	WBTB	Western blot transfer buffer
MBP	Maltose binding protein	XDS	X-ray detector software
MEM	Minimum essential medium	YPD	Yeast extract, peptone and dextrose
MW	Molecular weight	Znf-UBP	Zinc finger – ubiquitin binding protein
NMR	Nuclear magnetic resonance		
OD	Optical density		
PB	Polobox		

1 Introduction: literature

1.1 The importance of protein regulation

In eukaryotes there are thousands to tens of thousands of protein coding genes in the genome [1] and an estimated 2-4 million proteins per cubic micrometre of cell [2]. Given this complexity, and the sheer number of cellular proteins, there has to be an incredibly efficient orchestration of these proteins to ensure cellular function. One such form of regulation is post translational modification (PTM). This is an umbrella term for more than 400 discrete modifications that are made to proteins after they are translated. Addition or loss of any moiety at the surface of a protein alters its surface topography, which can have consequences for protein activity and interactions with other proteins. For this reason, they are critical to the proper regulation of nearly all cellular protein functions [3].

PTMs may be in the form of covalent addition of small molecules such as: phosphates or acetyl groups, or polymers, including carbohydrates and proteins. It is not always a single modification per protein; often, they are found in combinations of multiple modifications. In addition, they are often reversible, allowing temporal specificity into the system [3]. Data from SWISS-Prot suggest the most common forms of PTM are N-linked glycosylation, phosphorylation and acetylation [4]. Three types of modification that highlight the diversity of modifications are shown in Figure 1.1.

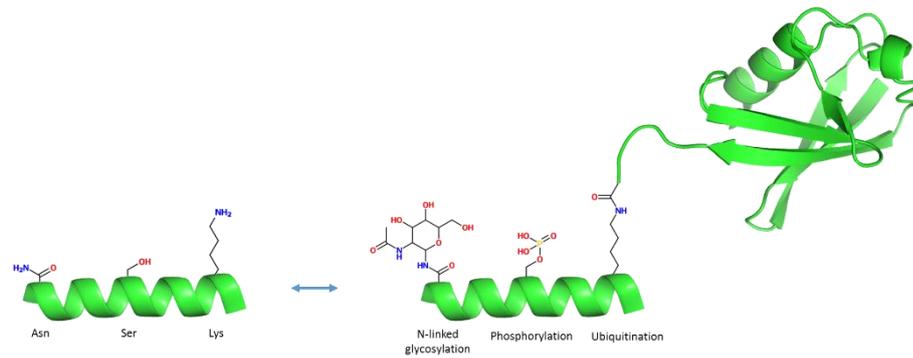


Figure 1.1. Post-translational modifications. Three post translational modifications are shown: N-linked glycosylation on an asparagine residue, phosphorylation on a serine residue and ubiquitination on a lysine residue. The size of the PTM can vary dramatically; being small molecule or macromolecule.

The role of PTMs in regulation of cellular functions make them appealing targets for research. Understanding the PTM code, how this regulation is achieved and the effects of modifications on substrate proteins allows not only understanding of normal physiology, but also permits therapeutic intervention for diseases involving these processes. Phosphorylation is probably one of the most successfully targeted PTM systems to date, with drugs that treat a range of pathologies, including cancer and autoimmune diseases [5, 6]. With the success of this group, drugging other PTM systems is becoming more popular. In particular, the ‘druggability’ of the ubiquitin system is being investigated [7]. Already, bortezomib, a drug that inhibits one of the downstream effectors of ubiquitination – the proteasome – has been shown to be clinically effective for treating multiple myeloma [8]. Other drugs that inhibit proteins that ubiquitinate and deubiquitinate proteins are also being investigated [7, 9, 10]. Because the ubiquitin system is involved in normal and pathological physiology, it is crucial that it’s well characterised on a molecular level.

1.2 Ubiquitination

1.2.1 A brief history of ubiquitin

Ubiquitin was first identified in 1975 by Goldstein *et al.* [11] as a highly conserved 8.5 kDa polypeptide found universally in prokaryotes and eukaryotes, originally termed ubiquitous immunopoeitic polypeptide; subsequently named ubiquitin [12]. However, it was later discovered that the immunopoeitic effect of the protein was due to endotoxin contamination of the protein mixture [13], and that ubiquitin was not present in prokaryote genomes (yeast extract from the bacterial medium was the likely source of the protein) [14]. In 1977, a hint at ubiquitin's role as a post-translational modifier was observed through a unique finding of a protein with two N-termini. This was identified to be ubiquitin covalently-linked to histone 2A [15, 16]. A year later, an article by Ciechanover *et al.* [17] was published, which provided the first insight into the significance of ubiquitin's role in the ubiquitin-proteasome system. It wasn't until 1980 that the APF-1 protein identified in this paper was ascertained to be ubiquitin, and the first model of the system was proposed [18, 19]. These findings commenced the vast field of ubiquitin research, and to acknowledge this work, the 2004 Nobel Prize in Chemistry was awarded to Aaron Ciechanover, Avram Herskho and Irwin Rose [20].

Ubiquitin is a highly conserved, 76 amino-acid, heat-stable protein that functions as a post-translation modification. In all eukaryotes it is encoded by genes that express it as a fusion protein or as a polyubiquitin chain. The human genes UBA52 and RPS27A express a single ubiquitin fused to ribosomal proteins L40 and S27a, respectively [21-24]. The UBB gene expresses a protein of three linear ubiquitin molecules with a C-terminal extension of one cysteine residue. The UBC gene encodes nine ubiquitin molecules with a C-terminal extension of one valine residue. Subsequent processing of all gene products leads to single ubiquitin polypeptides [25, 26].

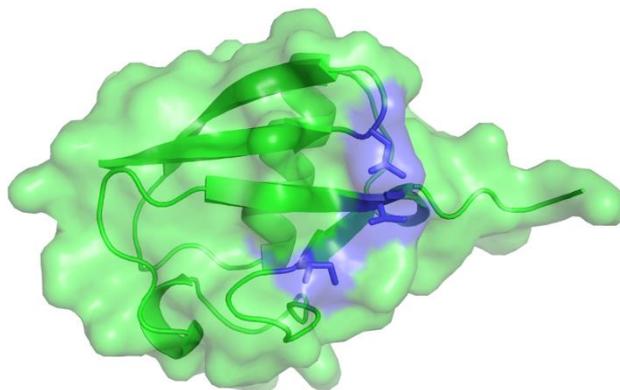


Figure 1.2. Ubiquitin. The β -grasp fold can be seen where a β -sheet wraps around a helix, which is capped by a small 3:10 helix. The C-terminus protrudes from the rest of the folded domain (seen at the right of the image), allowing it to covalently bind to the ubiquitination machinery and substrate proteins. The Van der Waals surface is shown. The hydrophobic patch (including leu8, Ile44 and Val70) is shown in blue. This serves as the site for multiple protein interactions, and is essential for proteasome binding [27] (PDB code: 1UBQ).

Ubiquitin folds into a compact, globular protein comprising a five strand β -sheet, alpha helix and a short 3:10 helix, known as a β -grasp fold [28], shown in Figure 1.2. The C-terminus protrudes from the globular domain to allow binding to substrate proteins. Along with ubiquitin, there are other ubiquitin-like (Ubl) modifiers that also operate as PTMs, such as SUMO and NEDD8. The structure of Ubl proteins are similar to that of ubiquitin and thus contain the ubiquitin superfold that typically follows the same secondary structure pattern: strand-strand-helix-strand-strand-helix-strand [28-34]. Mostly, Ubls differ substantially at their C-terminus, as well as differences in surface topography [35].

1.2.2 The cascade to ubiquitination

Ubiquitination is a highly regulated process performed in a cascade of events by enzymes known as ubiquitin activating enzymes (E1s), ubiquitin conjugating enzymes (E2s) and ubiquitin ligases (E3s). The end result of this cascade is a substrate protein with one or more covalently-linked ubiquitin proteins either on a lysine's epsilon amino group (via an isopeptide bond) or on its N-terminal amino group (via a peptide bond) [36, 37]. Ubiquitination can elicit multiple effects on a substrate protein, including lysosomal or proteasomal degradation, cellular trafficking, structural modification and modulation of protein-protein interactions.

E1s activate ubiquitin by adenylating its C-terminus and covalently-binding to it through a conserved E1 cysteine residue [38, 39]. In humans there are two ubiquitin E1s: Ubl modifier-activating enzyme 1 (UBA1) and UBA6 [40]. There are other UBAs

in the human genome, but they're involved in activating other Ubl proteins, such as SUMO and NEDD8.

Ubiquitin is activated by the E1 enzyme in a multi-step process. First, ubiquitin binds to the E1's adenylation domain by hydrophobic interactions from ubiquitin's β -sheet, polar interactions from its globular domain and through interactions with its C-terminal tail, which protrudes into a cleft formed by the E1. Here, the C-terminal glycine residue is adenylated by the E1's adenylation domain, and is subsequently transferred to the first and second catalytic half domains, where it's covalently linked to the active-site cysteine with a thioester bond. A second ubiquitin molecule binds to the adenylation domain after the transfer of the first, producing the fully active E1-Ub2 complex [41, 42], shown in Figure 1.3.

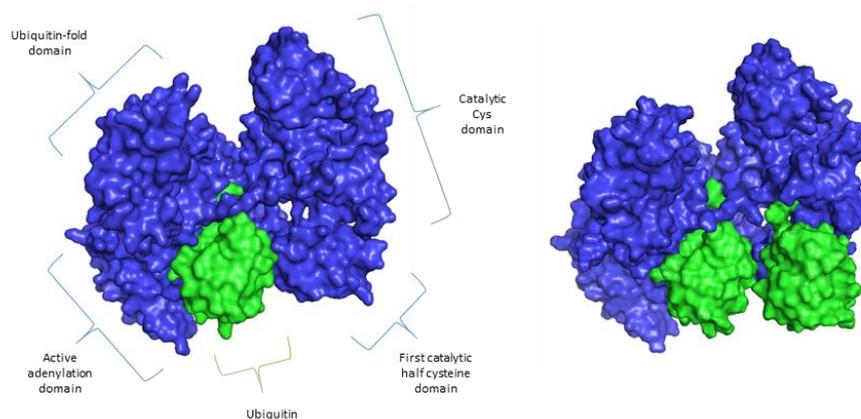


Figure 1.3. UBA1. (Left) UBA1 (blue) with a single non-covalently bound ubiquitin molecule (green). **(Right)** Doubly-loaded UBA1. The second ubiquitin binds to the active-site cysteine residue.

The role of E2s is to receive a Ubl protein from an E1, and covalently bind it through a catalytic cysteine in their ubiquitin conjugation domain. There are at least 38 E2 genes in the human genome [43] and phylogenetic analysis categorises these into 17 subfamilies, 12 of which exclusively conjugate ubiquitin, two conjugate multiple Ubl proteins (including ubiquitin), two only conjugate non-ubiquitin Ubl proteins and one has no active site cysteine [44-46]. E2 selectivity is essential so that they interact with the correct E1-Ubl complex (Figure 1.4) and therefore conjugate the appropriate modifier. Selectivity is largely mediated by the catalytic Cys and ubiquitin fold domains of the E1 [47]. Ubiquitin E2s provide their own contribution to specificity: two E2 lysines interact with negative groove in the E1's UFD, which are not found in other Ubl E2s [41-43].

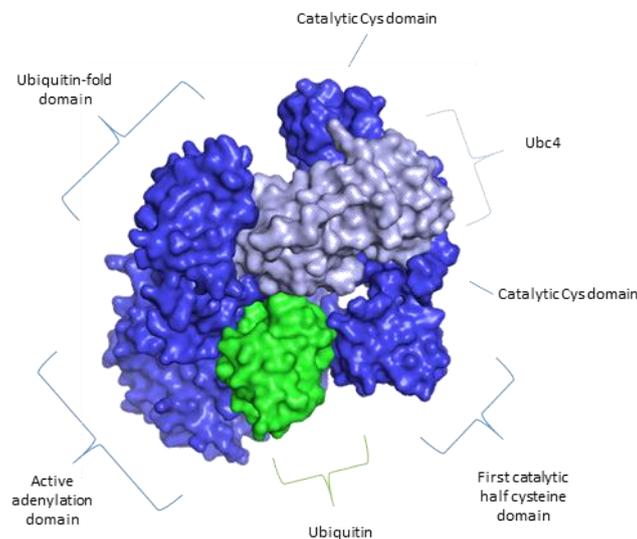


Figure 1.4. UBA1-Ubc4 complex. The E2 enzyme Ubc4 (light blue) binds to the catalytic Cys domain (right side) and the ubiquitin fold domain of UBA1 (dark blue). Ubiquitin is shown in green (PDB code 4II2).

E3s form the largest group in the cascade, which can be broken down into two major subgroups: RING E3s and HECT E3s. Both types of enzymes mediate the transfer of ubiquitin from the E2 to the substrate protein. The major difference between these two groups is that HECT E3s act as a catalytic intermediate, whereas RING E3s do not covalently bond to ubiquitin prior to substrate transfer. HECT E3s generally have an N-terminal, bi-lobed HECT domain and C-terminal protein recruitment domains. One lobe of the HECT domain interacts with the E2 and the other contains a catalytic cysteine that covalently binds ubiquitin. RING E3s contain a RING-type ZnF domain that binds to the E2 [48]. They act as a scaffold to recruit substrates to the E2, and possibly orientate the substrate to a suitable conformation to receive the ubiquitin. Complex structures of Cbl and Ubch7, and E6AP and Ubch7 show that both types of E3 interact with the same residues on the E2 (Figure 1.5) [49, 50].

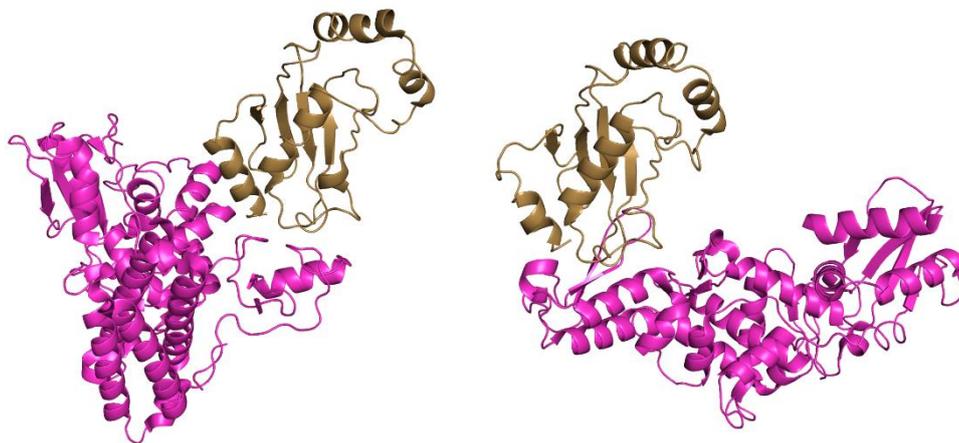


Figure 1.5. E2-E3 complexes. The E2 enzyme Ubch7 binds to the E3 enzymes E6AP (left) and CBL (right). The same region of the Ubch7 (brown) binds to both E3 enzymes (magenta) Made with the crystal structures 1C4Z and 1FBV.

1.2.3 The ubiquitin code

Residues on substrate proteins can either be monoubiquitinated or polyubiquitinated. These describe the addition of a single ubiquitin or a chain of ubiquitin molecules, respectively. Ubiquitin has seven lysine residues (K6, K11, K27, K29, K33, K48 and K63; shown in Figure 1.6) and can be ubiquitinated on any of these residues, as well as its N-terminus, forming the polyubiquitin chains. The chains can be pure-linkage, where every subsequent ubiquitin in the chain is joined to exactly the same lysine residue of the previous ubiquitin, or they can be mixed or branched, where multiple linkage types are present in a single chain [35, 37].

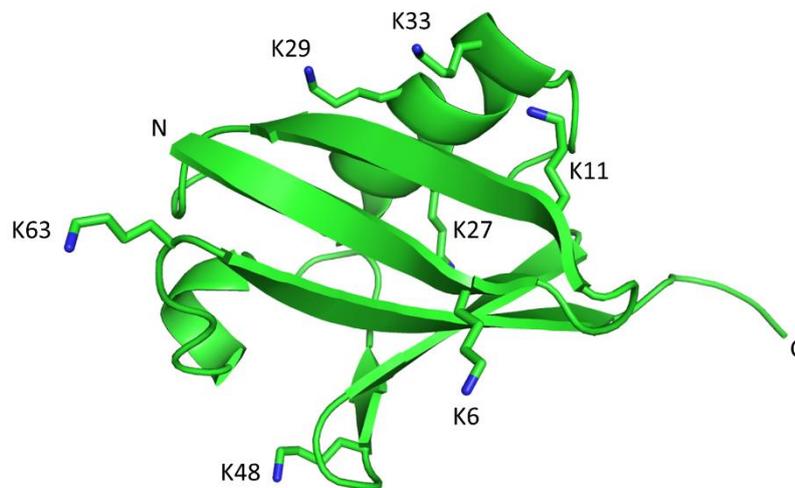


Figure 1.6. Ubiquitin's lysines. The seven lysines (side chains only) of ubiquitin are shown. K63 and the N-terminus can be seen on the opposite face of the protein to the C-terminus. K48 is closer to the C-terminus, around half way along the protein. These differences in locality are the cause of the structural differences of polyubiquitin chains produced from a specific linkage (PDB code: 1UBQ).

Single linkage chains are the best characterised forms of polyubiquitin; the physiological role of mixed and branched chains remain elusive. In particular, K48- and K63-linked chains have been best characterised of all. The linkage type of the polyubiquitin chain determines the chain's surface topography. Generally, the chains form two types of structure: globular or linear with some modulations. The globular chain is exemplified by K48-linked chains where each ubiquitin in the chain tends to pack against another forming a closed structure. K63-linked chains tend to form linear structures, where the C-terminus of each ubiquitin is bound to the opposite side of the proximal ubiquitin molecule. Structures of K48- and K63-linked ubiquitin are shown in Figure 1.7. Chain length and surface topography are what provide functional specificity by allowing specific recognition of effector proteins [37].

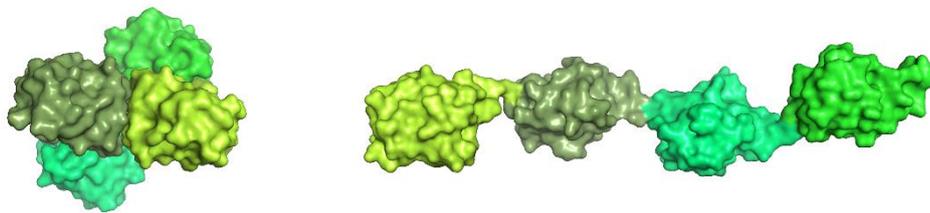


Figure 1.7. Polyubiquitin chains. K48-linked chains (left; PDB code: 2O6V) generally form a more closed structure than more linear chains such as K63-linked polyubiquitin (right; PDB code:2JF5).

Substrate modification with K48-linked chains is associated with proteasomal degradation [51]. The proteasome is a proteinacious mega-structure that acts as one of the cells major routes of protein degradation [52]. This huge, multi-subunit protease recognises K48-linked chains through its S5a/Rpn10 and Rpn13 subunits

[53, 54]. It then digests the substrate protein via its catalytic core components. In addition to binding of closed conformation K48-linked chains, most linear chains (except K63) also bind to the proteasome and lead to substrate degradation. K63-linked chains bypass proteasomal degradation by binding to the effector complex, ESCRT0 [55]. Ubiquitination of substrate proteins with K63-linked chains are associated with transport to the endocytic pathways, lysosomal degradation, DNA repair and gene transcription [56-60]. The roles of other chain types remain unclear; K11-linked chains may be associated with endoplasmic reticulum associated degradation [61], K6-linked chains may have a role in DNA repair and cell division [62, 63], and N-terminus-linked chains may act as a source of ubiquitin that can be freed and used in other polyubiquitin chains. The roles of mixed and branched chains are less clear.

Because of the importance of substrate ubiquitination, proper control of the system is essential. It is, therefore, required that cellular enzymes are able to cleave ubiquitin from a substrate to reverse these cellular processes.

1.3 Deubiquitination

Deubiquitinating enzymes primarily (DUBs) play the antagonistic role in the ubiquitin system. They contain a catalytic domain that is capable of cleaving the isopeptide bond between the ubiquitin C-terminus and a lysine residue's epsilon amino group, or the peptide bond between the ubiquitin C-terminus and an N-terminal amino group [64].

One of the primary roles for DUBs is the removal of ubiquitin modifications from substrate proteins in order to reverse the cellular fate of the ubiquitinated protein. In addition, DUBs have other cellular roles. As ubiquitin is always expressed as a fusion protein, DUBs are required to produce free ubiquitin by cleavage of ubiquitin monomers from the chains of linear ubiquitin, and removal of the non-ubiquitin C-terminal extensions on these proteins. During the ubiquitination process, the C-terminus of ubiquitin can become accidentally modified by small nucleophilic molecules. DUBs can remove these molecules, restoring the proper C-terminus of ubiquitin. They also renew the ubiquitin pool by cleaving polyubiquitin chains that have been previously removed from substrate proteins or the proteasome following substrate degradation [65].

DUBs can be divided into five major groups that differ with respect to their catalytic fold. Four groups are cysteine proteases: ubiquitin C-terminal hydrolases, ubiquitin specific proteases (USPs) ovarian tumour proteases and Machado-Joseph disease proteases. One group, the JAB1/MPN/Mov34 metalloenzymes, are metalloproteases [66].

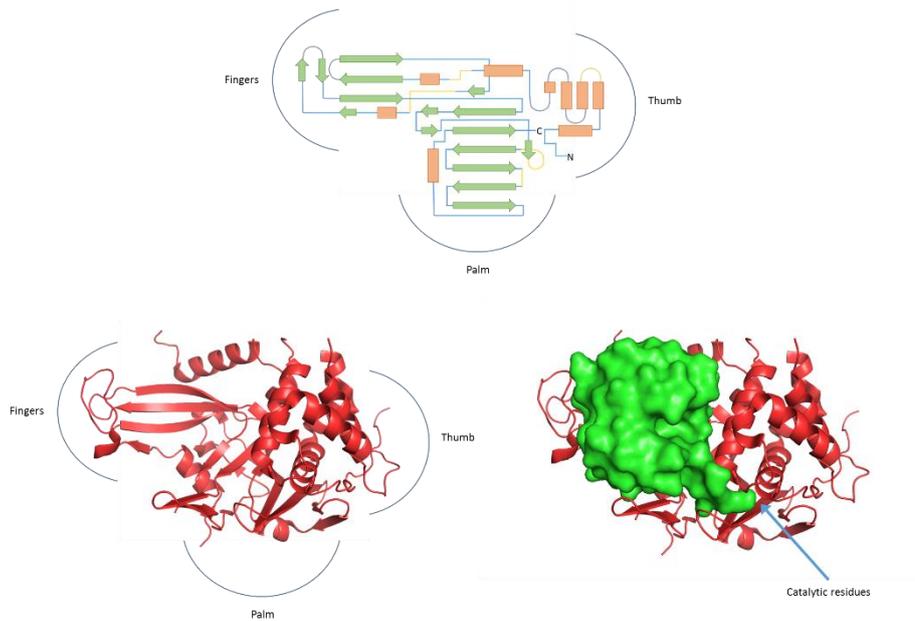


Figure 1.8. The USP domain. The topology of core fold of the USP domains is shown in the upper image. Helices are red rectangles, β -strands are green arrows. The yellow loops indicate common insertion points where additional loops or domains may be found. The crystal structure of the catalytic domain of USP21 is shown in the lower images (PDB code 2Y5B). Lower left shows the crystal structure without ubiquitin visible. The fingers can be seen on the left of the structure, the thumb as a bundle of helices on the right, and the palm is a β -sheet cradling the thumb centre-rear. Lower right shows the same domain with ubiquitin bound (diubiquitin is present in the structure, but the second ubiquitin has been hidden for clarity). The tail of ubiquitin inserts into the cleft formed by the palm and thumb subdomains. The catalytic cysteine and histidine residues are located here (blue arrow) that cleave ubiquitin from its substrate (or additional ubiquitin).

In humans, there are over 90 DUBs, over 50 of which can be found in the USP group. The USP domain fold is highly conserved, and has been likened to fingers, palm and thumb (Figure 1.8) [67]. Ubiquitin binds to the domain and its C-terminus protrudes through a cleft formed by the palm and thumb regions. Here, the catalytic cysteine and histidine residues of the catalytic triad are located. The histidine is polarised, typically by an aspartate residue, which raises its pKa. The close locality of the polarised histidine and cysteine lowers the pKa of the cysteine. This allows it to

perform a nucleophilic attack on the isopeptide bond between ubiquitin's C-terminal glycine and the substrate residue [64]. The full mechanism of the catalytic reaction is shown in Figure 1.9.

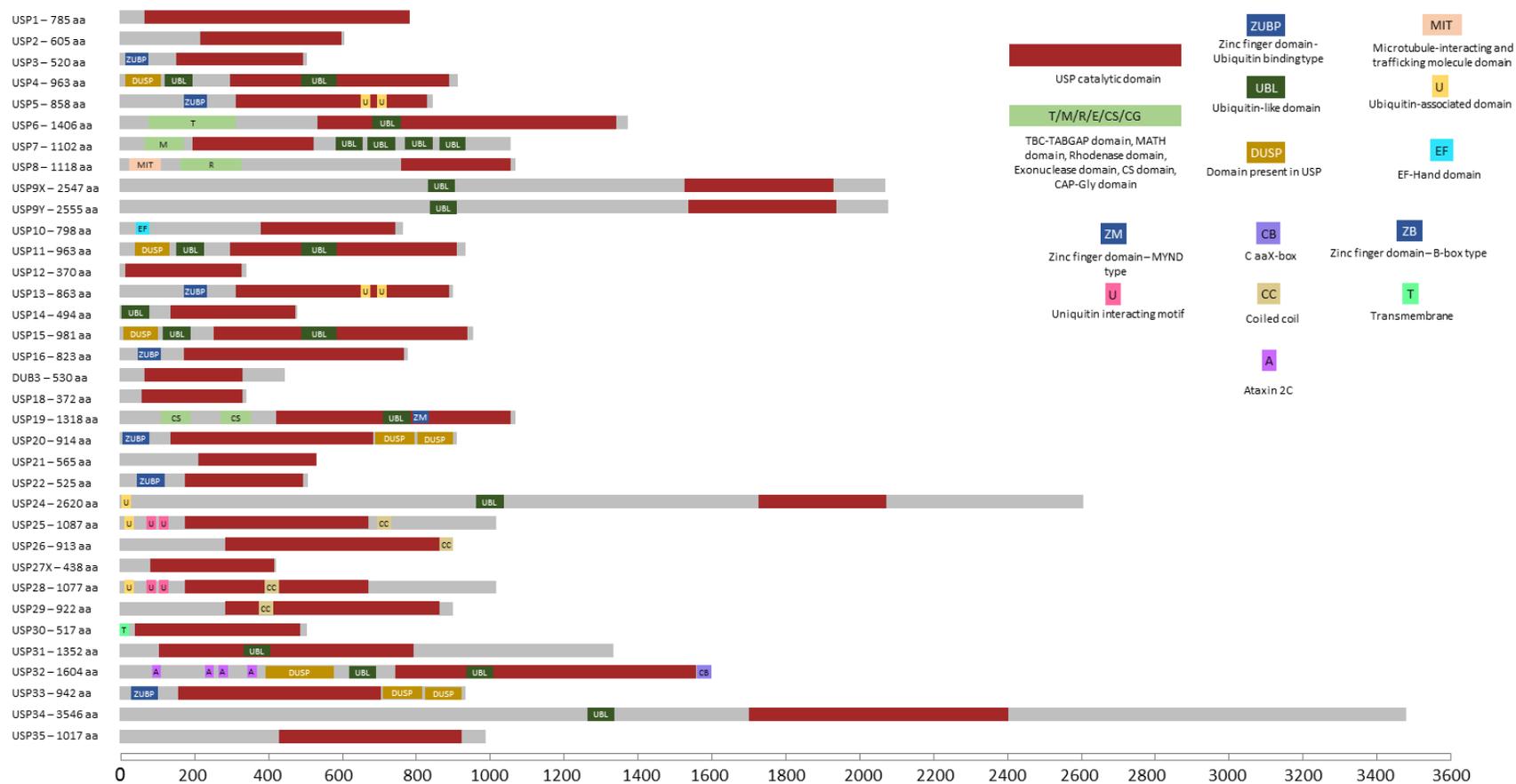


Figure 1.9. Ubiquitin specific proteases 1. Cartoons of USPs are shown. The group is diverse with varying lengths and many accessory domains. A legend is given to show what domains are present.

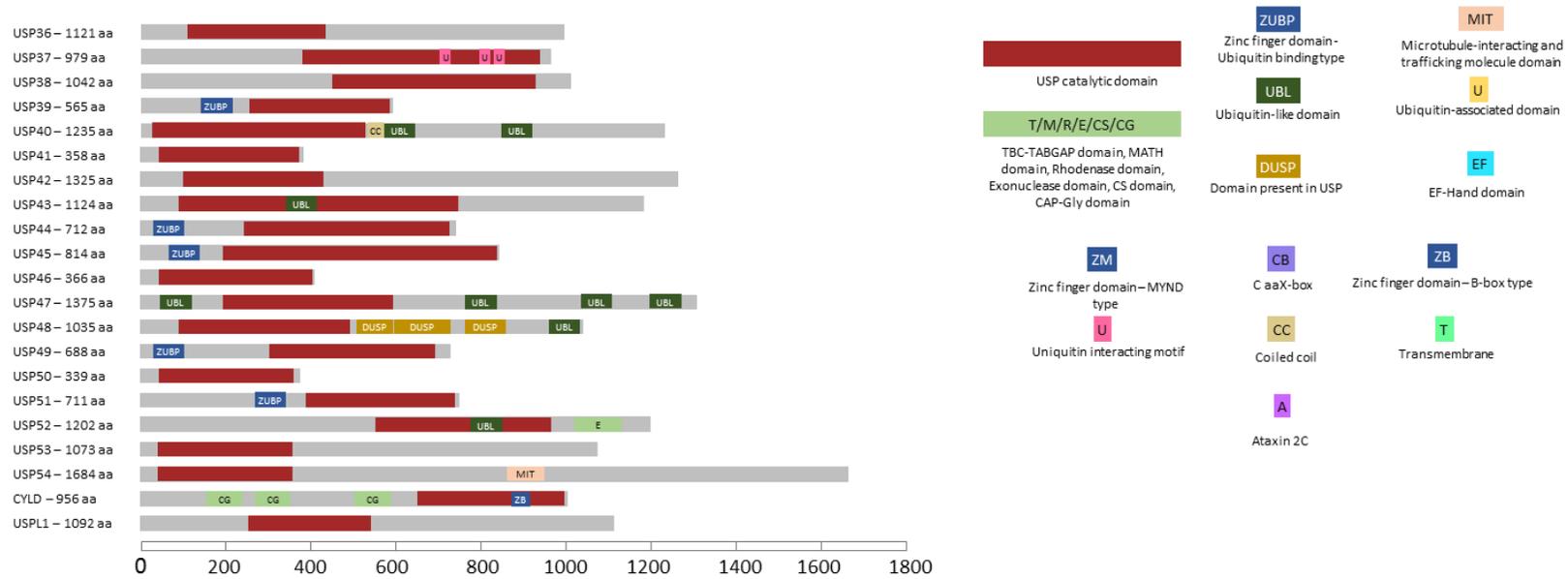


Figure 1.10. Ubiquitin specific proteases 2. Cartoons of USPs are shown. The group is diverse with varying lengths and many accessory domains. A legend is given to show what domains are present.

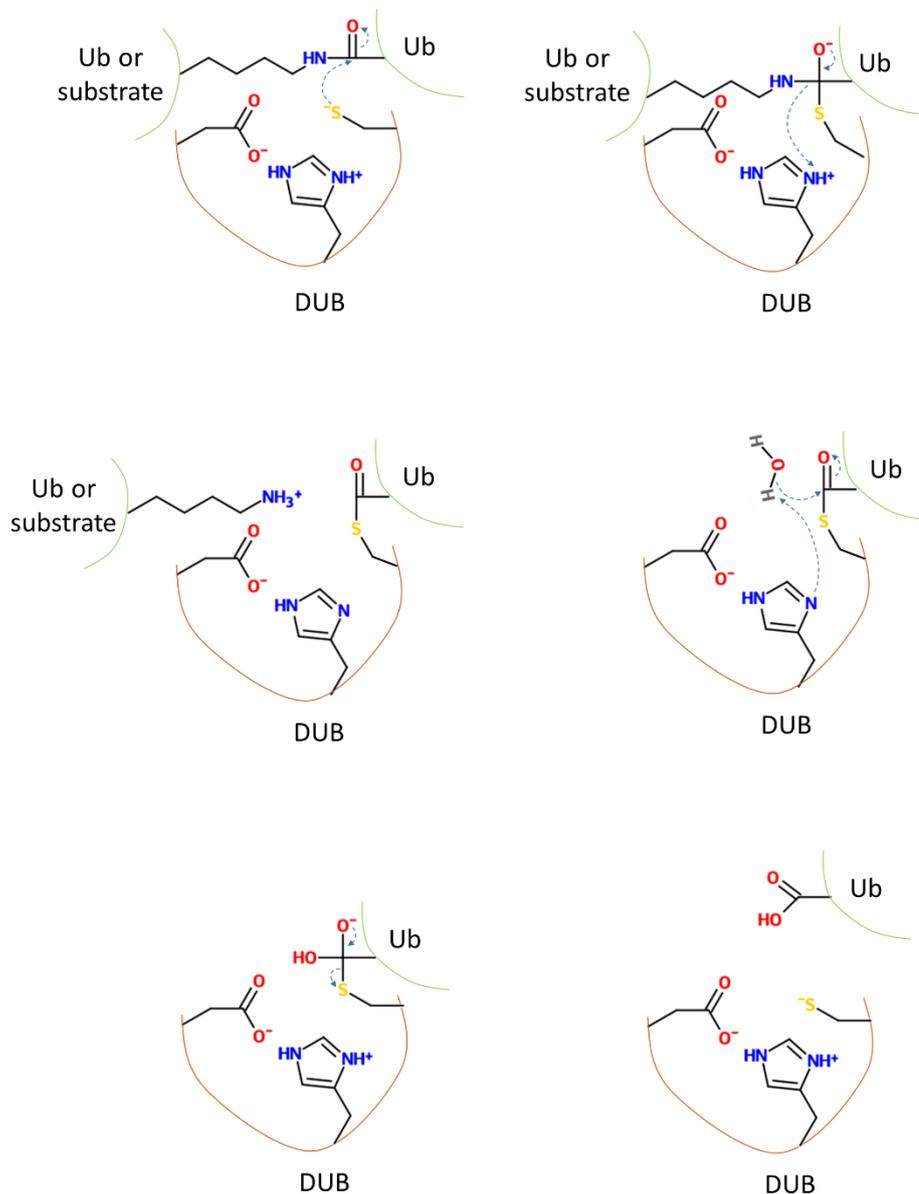


Figure 1.11. Catalytic mechanism of USPs. (A) With ubiquitin bound, the deprotonated catalytic Cys performs a nucleophilic attack on the terminal carbonyl carbon of the ubiquitin molecule. **(B)** A tetrahedral intermediate is formed, where the oxygen is stabilised by the oxyanion hole. Here, Asp or Asn stabilises the oxyanion with hydrogen bonds through water molecules (not shown in diagram). **(C)** The tetrahedral intermediate collapses, releasing the Ub/Substrate lysine and forming an acyl intermediate linking the ubiquitin and the DUB. **(D + E)** A water molecule leads to the hydrolysis of the acyl intermediate. The water molecule's attack on the acyl carbon produces a negatively charged intermediate, which then collapses, releasing the ubiquitin molecule and reinitialising the catalytic site of the DUB, shown in **(F)** [64, 68, 69]. Diagram adapted from [68].

1.3.1 DUB regulation

DUBs are regulated by multiple mechanisms to ensure proper homeostatic deubiquitination. One of these is appropriate temporospatial regulation, which can be either pre- or post-translation. Various DUBs are only expressed when required, such as USP1, which is associated with DNA repair. Its mRNA transcript levels are cell cycle dependent; only achieving high levels in S-phase when the protein is required [70]. Localisation can also be affected by splice variants. For example, isoform 3 of USP33 has only small modifications to its primary sequence, but this causes its isoform-specific localisation to the golgi apparatus [71]. Also, the addition of certain exons from the USP25 and USP28 genes causes tissue specific expression of the proteins [72].

Protein recruitment and localisation are essential mechanisms for ensuring a DUB interacts with the correct regulatory proteins and substrates, and that these processes occur in the appropriate organelle or cellular compartment. DUBs are modular proteins that utilise the presence of additional domains to recruit substrates or be recruited by scaffolding proteins or regulatory enzymes. The accessory domain repertoire of DUBs is highly diverse, allowing DUB substrate diversity. As the catalytic domains of DUBs primarily recognise ubiquitin itself (not including insertions into the loops of the core fold), specificity in substrate recruitment is thought to be mainly achieved by these accessory domains. It is also important to note that DUBs utilise differences in the structures of Ubl proteins, particularly at their C-terminus, in order

to ensure they act only as deubiquitinases (and do not remove other Ubl proteins). Only few examples of cross-reactive DUBs have been observed [35, 73]. Many DUBs are found within single cellular compartments only, and some are found at specific protein structures, such as POH1, UCH37 and USP14, which are bound to the proteasome [74, 75].

DUBs can be found in active and inactive forms in the cell. Inactive forms can be induced into an active conformation by ubiquitin binding, substrate binding, self-interactions, binding to other proteins or PTM. Structures for the catalytic domain have been obtained for USP7 with and without ubiquitin. These structures show that the native catalytic domain has a cryptic catalytic site; its histidine and cysteine residues are nearly 10Å away from each other (Figure 1.10). However, upon binding of ubiquitin, the loops surrounding the cleft change conformation and close up onto the ubiquitin C-terminus. The new conformation aligns the catalytic His and Cys residues, thus creating an active enzyme only upon substrate binding [67]. Additionally, USP7 has five C-terminal Ubl domains that interact with the catalytic domain to ensure active conformation [76].

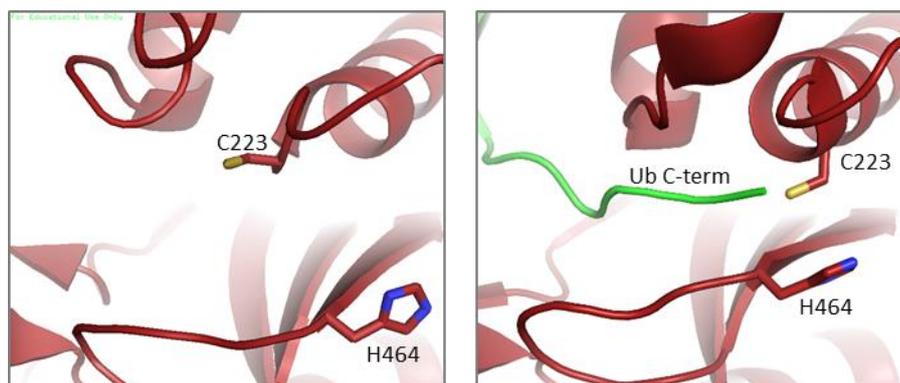


Figure 1.12. USP7 cryptic catalytic site. The catalytic Cys and His residues of USP7 (red) are remote from one another when ubiquitin is not present (left; PDB code 4M5X). When ubiquitin (green) binds, a conformational change aligns the residues to allow catalytic cleavage of ubiquitin (Right; PDB code 5JTJ).

PTM-mediated activation has been observed in multiple DUBs including A20 [77], DUBA [78] and USP37 [79], where phosphorylation of these enzymes leads to their activation. Phosphorylation of USP7S leads to its stabilisation and continuation of substrate deubiquitination [80]. Alternatively, phosphorylation of CYLD by IKK-epsilon lowers the activity of the DUB [81]. In some cases phosphorylation indirectly activates DUBs by providing a binding site for activating proteins, as in the case of USP1 and UAF1 [82].

1.4 USP20 and USP33

USP20 and USP33 are part of the USP subgroup of DUBs. They form a distinct clade in the phylogeny as they both contain a dissevered catalytic domain flanked N-terminally by a 'ubiquitin binding protein type zinc finger' (Znf-UBP) domain and C-terminally by two tandem 'domain present in USP' (DUSP) domains [83] (Figure 1.13). It should be noted, however, that although the zinc finger domains of USP20 and USP33 are classified as Znf-UBPs, the USP33 domain has been shown not to bind to ubiquitin [84]. HDAC6 and USP5 Znf-UBPs, which do bind ubiquitin, have a pocket where the C-terminus of ubiquitin inserts and binds [85, 86]. Allen and Bycroft [84] concluded that the lack of interaction of ubiquitin and USP33 was because an arginine residue in HDAC6 and USP5 was replaced with glutamic acid in USP33. USP20 also has a glutamic acid residue here, so it is likely that USP20 does not bind to ubiquitin; however, no structure of its Znf-UBP has been solved and the interaction has not been tested.

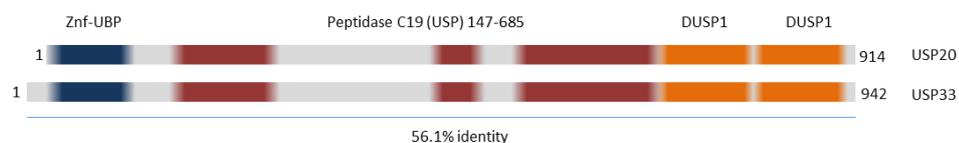


Figure 1.13. USP20 and USP33. A schematic of USP20 (top) and USP33 (bottom) are shown. The domain architectures are similar; an N-terminal Znf-UBP domain (blue), a USP domain (red) and two C-terminal DUSP domains (yellow). Grey regions indicate non-domain forming sequences of the two proteins. Overall, they show 56.1% sequence identity.

A study by Ye *et al.* [87] identified multiple common sites in the USP fold where the catalytic domain has inserts that include disordered loops and folded domains (Figure 1.14). The catalytic domains of USP20 and USP33 are interspersed with two disordered inserts: one between boxes two and three and one between boxes three and four. There are no conserved domains predicted within them and the exact function of these inserts in USP20 and USP33 are unknown, although they likely mediate interactions with other proteins.

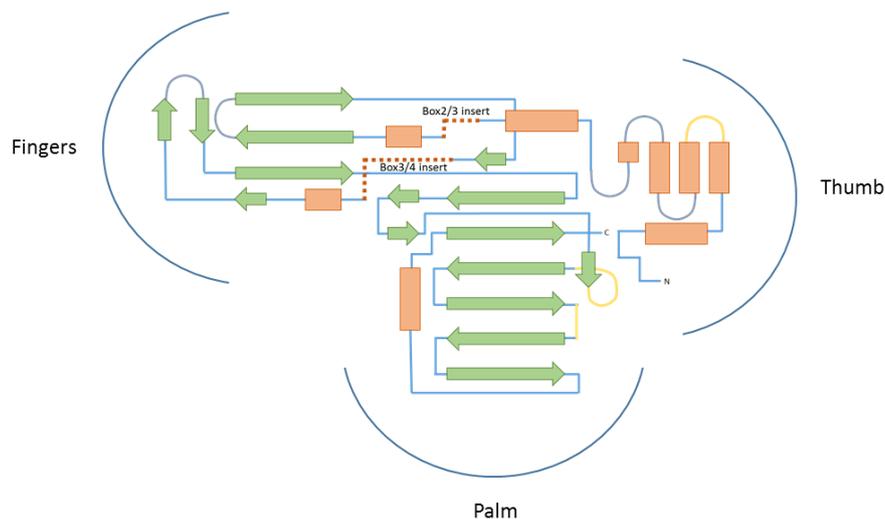


Figure 1.14. The topology of USP20 and USP33 catalytic domains. Both USP20 and USP33 show inserts between boxes 2 and 3, and 3 and 4 (red dotted lines). USP20 has a 182-residue insert between boxes 2 and 3, and a 52 residue insert between boxes 3 and 4. USP33 has a 172 residue insert between boxes 2 and 3 and a 52 residue insert between boxes 3 and 4 [87].

Also, little is known about the exact function of the conserved domains flanking the catalytic domain and only few interactions have been mapped within these two proteins. Only an NMR structure of the Znf-UBP domain of USP33 has been determined from both of these proteins [84], leaving much to question about the other domains such as: whether the tandem DUSP domains function as individual units or form compact globular structures, and the entire architecture of the native full length protein.

The two proteins share 56% identity, which is a possible reason for their overlap in function. Indeed upon knock down of either enzyme, compensation by the other is observed (especially upon USP33 knock down). In many cases, the proteins share interactions with each other, introducing functional redundancy such as: pVHL [83, 88, 89]. In some cases their interactions are discrete such as: Robo1 (USP33-specific) and HIF1- α (USP20-specific) [90, 91]. It is likely that the lower homology regions of the two proteins (between Znf-UBP and catalytic domain, and the catalytic insert) allow their individual functions. Together, these USPs have roles in thyroid metabolism, hypoxic response, seven transmembrane receptor signalling, NF- κ B signalling, centrosome homeostasis and DNA repair.

Together, these roles of USP20 and USP33 appear quite disjointed and its difficult to give a specific overall role for USP20 (or USP33). Indeed, no one has yet identified a unifying role for either enzyme. It is possible, in some cases, to identify roles for proteins through phenotypes of knock-down mice. Using the International Mouse Phenotype Consortium [92], USP20 deficient mice show a phenotype of increased natural killer (NK) cell number in addition to skeletal and corneal defects. USP33

deficient mice have decreased body weight through decreased lean mass and fat, increased bilirubin levels, increased mean corpuscular haemoglobin and thrombocytosis. Again, this does little to suggest what the specific role of either enzyme is, but shows that they do indeed have diverse cellular functions as they present with different phenotypes.

However, some links can be hypothesised with the observations. Increased NK cells identified as the USP20 phenotype links well with DDR [93]. It has been seen multiple times that DNA damage sensing proteins also activate the innate immune system, which includes NK cells. USP20 is phosphorylated in response to DNA damage and may play a role in activating the innate immune response. Additionally, the β -adrenergic signalling is also involved in the innate response, and, in particular, the β 2-AR has roles in most cell types of the innate immune system. In NK cells, the β 2-AR increases the cytotoxicity during social disruption [94]. Additionally, hypoxia and hypoxia inducible factor (HIF) expression also activate the innate immune response possibly through nuclear factor- κ B (NF- κ B) [95]. Indeed, USP20 has been linked multiple times to NF- κ B [96, 97], as has the β 2-AR, and its signalling is well known to activate the innate immune system [98]. Whether all of the identified roles for USP20 would ultimately affect innate immunity or specifically natural killer cells remains to be concluded. Further investigation of USP20 may unify all of its currently known functions and highlight its importance in the innate immune system. A more in depth view of USP20's known role in the literature is given below.

1.4.1 Thyroid metabolism

USP20 and USP33 regulate thyroid metabolism by deubiquitinating iodothyronine deiodinase type 2 (DIO2) [89]. DIO2 is a membrane-bound homodimer that converts inactive thyroxine (T4) into 3,5,4'-triiodothyronine (T3). It is selectively expressed in brain, thyroid, brown adipose and muscle tissues [99-101]. T3 binds to thyroid hormone receptors that in turn bind to thyroid response elements in DNA [102, 103], activating genes involved in development and homeostasis.

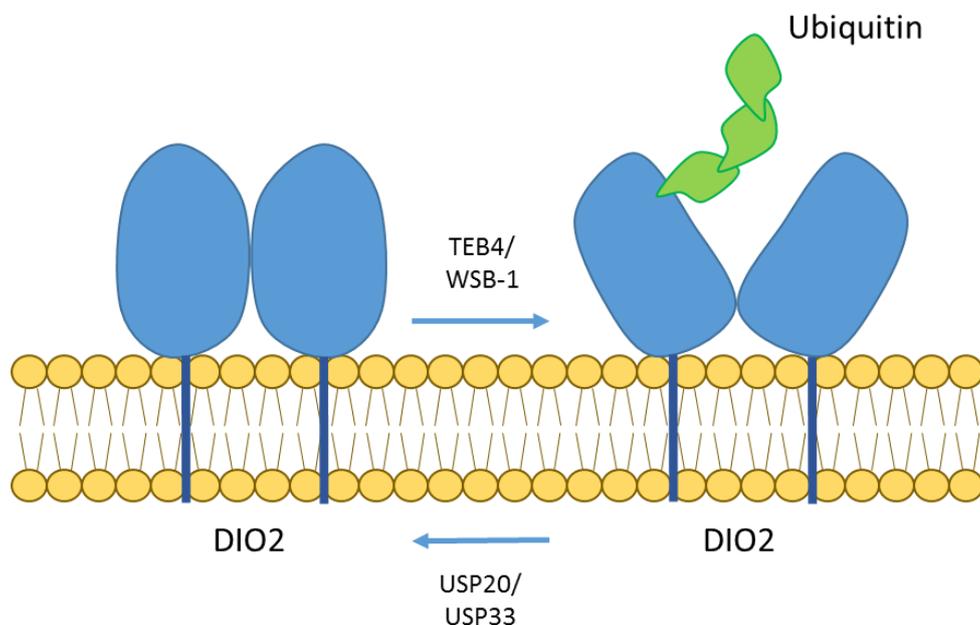


Figure 1.15. Regulation of DIO2. The membrane bound DIO2 (blue) is prevented from forming proper dimers when ubiquitinated. USP20 and USP33 reverse this, stabilising the protein and allowing catalytic function.

During the conversion of T4 to T3, DIO2 ubiquitination occurs, preventing further catalytic activity. Ubiquitination leads to DIO2's degradation and prevents its dimerisation (Figure 1.15). Complete dimerisation is essential for proper catalytic function [104]. USP20 and USP33 (residues 585-786) bind to the C-terminus of DIO2 (residues 166-233) and deubiquitinate it. DIO2 is required for the cold shock

response by brown adipose tissue and during this physiological process, USP33 levels (not USP20) are increased; suggesting that USP33 has a role in T3 mediated thermal homeostasis [89]. Increased DIO2 levels are observed in multiple diseases including brain tumours, neuroblastoma, follicular thyroid carcinoma, Graves' disease and McCune Alright syndrome [105-107]. No studies have yet been performed to identify whether USP20 or USP33 have roles in these diseases, although the COSMIC database shows mutations in this region for USP20 in thyroid cancer and USP33 in medulloblastoma [108, 109].

1.4.2 Hypoxic response

USP20 and USP33 play a role in the von Hippel Lindau tumour suppressor protein (pVHL) and hypoxia inducible factor 1 (HIF1) hypoxic response pathway. pVHL is an adaptor module for E3 ligase complex CRL2 (CRL2^{VHL}; Figure 1.16) [110-113].

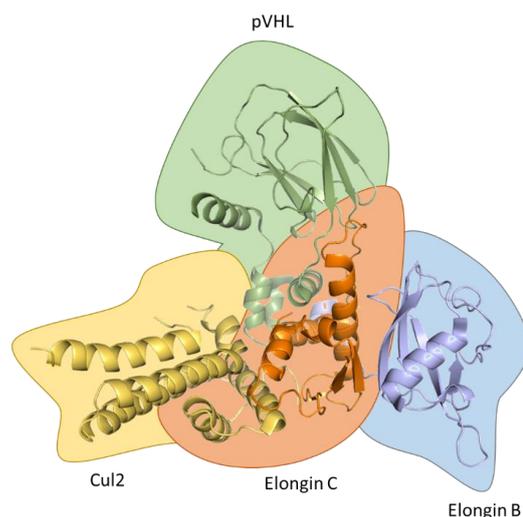


Figure 1.16. CRL2^{VHL} complex. The crystal structure of part of the CRL2^{VHL} complex displays a tripod structure (PDB code 4WQO). The top of VHL is the β -domain; the location where HIF1- α binds. Rbx1, the E3 ligase binds to Cul2 and ubiquitinates HIF1- α .

pVHL is formed from two domains: a primarily α -helical domain (α domain) and one primarily β -strand domain (β domain), in which the latter binds the E3 substrate HIF1- α . HIF1- α is a basic helix-loop-helix Per-Arnt-Sim transcription factor essential to the cellular hypoxic response [114-119]. During normoxia, HIF1- α is hydroxylated on two proline residues, which leads to binding of pVHL (Figure 1.17), ubiquitination and rapid proteasomal degradation of HIF1- α [120-130]. Lysines 532, 538 and 547 have been proposed as possible residues for ubiquitination [131, 132].

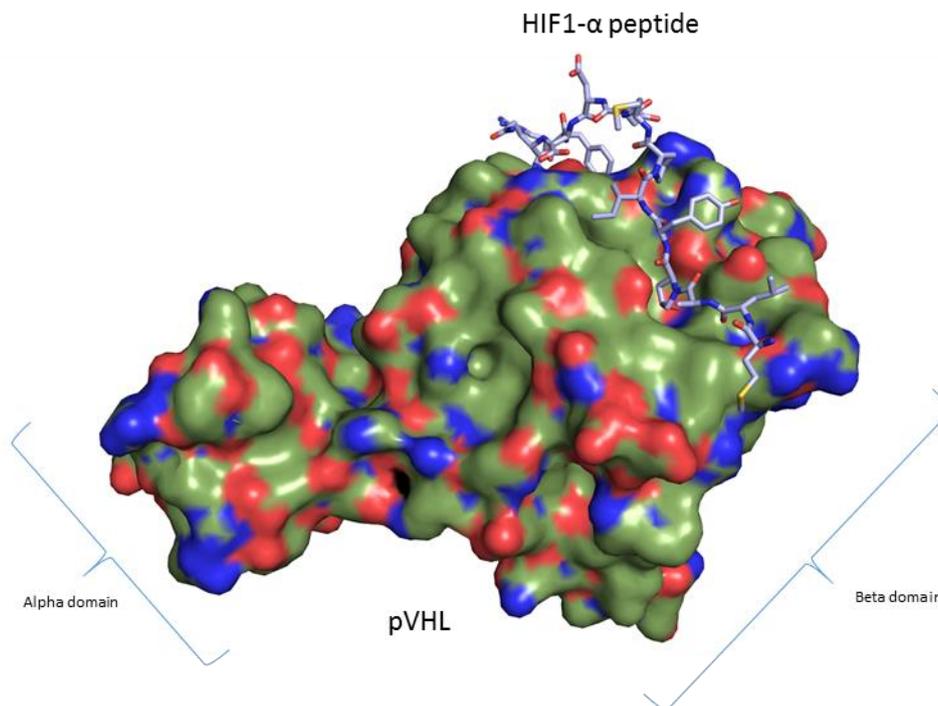


Figure 1.17. VHL-HIF1- α peptide complex. A crystal structure of a VHL-HIF1- peptide complex (PDB code 1LM8) shows the hydroxyproline residue of the peptide (light blue) binding into a pocket of VHL (green).

USP20 binds and deubiquitinates HIF1- α , which stabilises the HIF1 complex and upregulates the transcription of hypoxic response proteins [91]. This interaction is specific to USP20 only; no interaction was identified with USP33. The interaction was mapped to residues 269-390, which is between boxes 2 and 3 of the catalytic domain (in the large 182 residue insertion) [87]. Alignments of the amino acid sequence in this region only shows 21% identity between USP20 and USP33 [133].

pVHL negatively regulates USP20's effect on HIF1- α . Both USP20 and USP33 bind to pVHL and are ubiquitinated by the CRL2^{VHL} complex, leading to their degradation [83, 134]. The role of these DUBs in the hypoxic response may have implications in familial and sporadic cancer. Mutations in the VHL gene are associated with a familial cancer syndrome known as VHL disease; characterised by a massively increased risk of multiple forms of cancer [135]. This gene is also commonly mutated in sporadic renal cell carcinomas [136]. *In-vitro* analysis shows that pVHL's ability to interact with USP20 and USP33 is perturbed with naturally occurring mutations in its β -domain [83, 134]. This means that USP20 and USP33 could have a role in cancer physiology as aberrant stabilisation, although as yet untested, could lead to increased levels of HIF1- α .

1.4.3 Seven-transmembrane receptor signalling

USP20 and USP33 regulate the β_2 adrenergic receptor (β_2 AR) signalling pathway [88]. The β_2 AR is a seven-transmembrane, G-protein coupled receptor primarily expressed in pulmonary tissue [137]. Activation of the receptor leads to cell signalling through

effector proteins that include protein kinase A (PKA) and mitogen activated protein kinase (MAPK) [138]. Regulation is achieved by a process called desensitisation; classified as a reduction in cAMP signalling following ligand interaction, which is caused by receptor-level or downstream processes [139-141]. As a counter to desensitisation, upon stimulation, the receptor may only be temporarily internalised, dephosphorylated, and returned to the outer membrane, leading to resensitisation [142]. For some receptors, however, internalisation is not temporary: ubiquitination of the receptor can signal for its lysosomal degradation, preventing the receptor's return to the membrane [143] (Figure 1.18). The agonist-induced phosphorylation of the receptor allows β -arrestin to bind, which recruits the E3 ligase MDM2. This leads to the transient ubiquitination of β -arrestin and the internalisation of the receptor [144, 145]. NEDD4, another E3 ligase, is also recruited by β -arrestin. It ubiquitinates β_2 AR, signalling for its destruction by fate of endosome [146, 147]. Lysines in the third intracellular loop and C-terminus of β_2 AR are most commonly ubiquitinated when signalling for lysosomal signalling [148].

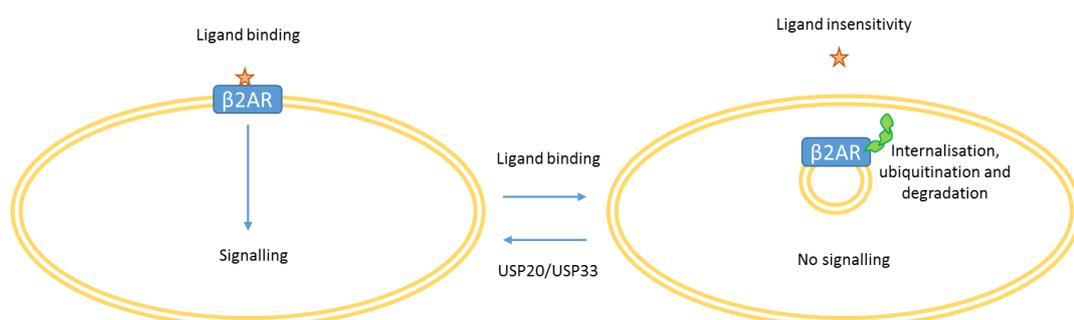


Figure 1.18. Receptor recycling. Upon ligand binding, the receptor is internalised and may become ubiquitinated. This leads to its degradation unless its fate is saved by USP20 or USP33, restoring the cell's sensitivity to ligand.

USP20 and USP33 bind to the β_2 AR, deubiquitinate it and prevent its ligand-induced lysosomal degradation. The deubiquitination causes its recycling from late endosomal and lysosomal compartments, resensitising the cell to ligand [88]. In addition, USP33 also deubiquitinates β -arrestin-2; one of the adapters for NEDD4 recruitment [149]. Berthouze *et al.* [88] suggest that when β -arrestin binds, NEDD4 and USP33 are dynamically exchanged, allowing NEDD4 to ubiquitinate the receptor, and USP33 to deubiquitinate β -arrestin. They also propose that modulation of the deubiquitinating enzymes could aid in the reduction of side effects seen from β_2 AR targeting drugs used to treat asthma.

Interestingly, this pathway has also been linked to that of pVHL. Egl-9 family hypoxia-inducible factor 3 can hydroxylate P382 and P395 of β_2 AR during normal physiological conditions [150]. This leads to pVHL's binding to the hydroxylated prolines, ubiquitination and proteasomal degradation. The exact role of β_2 AR in hypoxia is still unclear.

1.4.4 NF- κ B signalling

USP20 and USP33 are both involved in deubiquitinating proteins in NF- κ B signalling. These include the human T-lymphotropic leukaemia virus type-1 (HTLV-1) Tax protein and tumour necrosis factor receptor-associated factor-6 (TRAF6) [151]. HTLV-1 infection can lead to the development of acute T-cell leukaemia and HTLV-1 associated myelopathy, primarily mediated by Tax; a 272 amino acid protein that

directly facilitates the malignant transformation of T-cells through signalling pathways such as NF- κ B, CREB, SRF and AP-1 [152-155].

Tax is both ubiquitinated and sumoylated on its C-terminal residues in a mutually exclusive manner [156, 157]. Sumo is a UBL and is used for PTM in a similar manner to that of ubiquitin [158]. Sumoylation of Tax ensures its retention in the nucleus, where it effects its transcriptional activity [157, 159, 160]. Tax ubiquitination causes its nuclear export and cytoplasmic retention, where it activates NF- κ B [157, 159-166] (Figure 1.19).

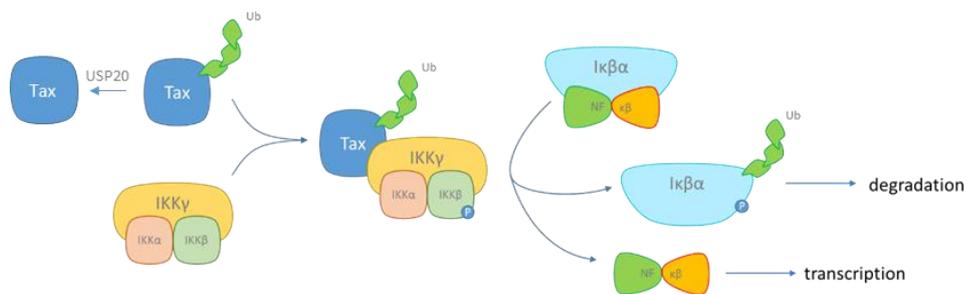


Figure 1.19. Tax and NF- κ B. Tax constitutively activates the IKK complex by binding IKK and causing the phosphorylation of IKK β . As consequence, I κ B α is phosphorylated, ubiquitinated and subsequently degraded; releasing NF- κ B and allowing gene transcription.

USP20 and USP33 are negative regulators of NF- κ B signalling because they can deubiquitinate TRAF6 [151]. TRAF6 is an E3 ligase that through activation of key cellular proteins, activates NF- κ B activation [167-169]. As TRAF6 ubiquitination is required for signal transduction, deubiquitination by USP20 and USP33 serve to regulate these proteins in a non-degradative manner. The interaction of USP20 and

TRAF6 is dependent on first binding to β -arrestin-2 [97]. Additionally, USP20 (not USP33) can bind and deubiquitinate Tax, again negatively regulating this protein's oncogenic signalling. Yasunaga *et al.* [151] showed that increased expression of USP20 reduced proliferation of T-cell leukaemia cells *in vitro* and downregulation of USP20 was observed in acute T-cell leukaemia caused by chromosomal deletion of 9q34 [170]. Together these studies show that USP20 and USP33 are integral to regulation of cellular signalling for proliferation, and may have important roles in acute T-cell leukaemia.

1.4.5 Centrosome homeostasis

USP20 and USP33 have been shown to bind and deubiquitinate the centriolar protein CP110 [171]. Additionally, NEURL4, another a centriolar protein, is pulled down by both USP20 [171, 172] and USP33 [171]. CP110 is a distal centriolar capping protein involved in determining the length of centrioles [173] and suppressing the conversion of centrioles into cilia [174]. In addition, it is involved in ensuring proper duplication of centrioles during S-phase, where it's most highly expressed [175, 176]. CP110 is ubiquitinated by the CRL1^{CyclinF} complex [177, 178]. NEURL4 is a daughter-specific centriolar protein that interacts with CP110 and promotes its ubiquitination, preventing overamplification of CP110 [179].

USP20 and USP33 bind to CP110, where USP33 is confirmed to deubiquitinate it [171]. This reverses the fate of CP110 following CRL1^{CyclinF}-mediated ubiquitination,

positively regulating the duplication of centrioles. USP20 and USP33 do not affect NEURL4 levels, suggesting that their substrate is specifically CP110 in the centriole. Interestingly, Li *et al.* [171] determine that USP33 is the primary protein mediating control of these centriolar proteins because knockdown of USP33 has the most prominent affect on CP110 levels. This is most likely an incorrect conclusion, as (1) USP20 seems to marginally pull down more of CP110 and NEURL4 in pull down assays. (2) Berthouze *et al.* [88] showed that USP20 knock-down massively upregulates the levels of USP33, whereas USP33 knock-down is only modestly compensated by USP20. This would explain the observations in this paper even if USP20 had an equal role to USP33 in this system. (3) NEURL4 was pulled down with USP20 in a proteomic screen for DUB interacting proteins [172]. It is, therefore, conceivable that USP20 does have a prominent role in centrosome maintenance.

Overexpression of CP110 and USP33 has been observed in pancreatic cancer tissues and cell lines [171, 180, 181], suggesting that they may be involved in pathogenesis of cancer, particularly in those with centriolar defects. Therefore, in this setting, USP20 and USP33 may act as oncogenes, and be adequate targets for therapy.

1.4.6 DNA damage response

A few recent reports have linked USP20 with DNA damage response (DDR) (Figure 1.20). It has been found to deubiquitinate CLASPIN and RAD17. Also, USP20 was identified in a large proteomic screen to bind to PLK1, but this is as yet unvalidated (Although it has been shown that USP20 depletion increased levels of the kinase)

[182]. CLASPIN and RAD17 are essential for the activation of the Chk1 cell cycle checkpoint. PLK1 inactivates Chk1 in dividing cells.

USP20 binds to RAD17 and deubiquitinates it, preventing its degradation. The interaction is mediated chiefly by its Znf-UBP domain, but also by its catalytic domain [182]. Upon UV-induced DNA damage, USP20 is phosphorylated on residues T170, T232, S305 and S662 by ATR. This causes the E3 enzyme HERC2 to dissociate from USP20, allowing it to deubiquitinate and rescue CLASPIN [183, 184].

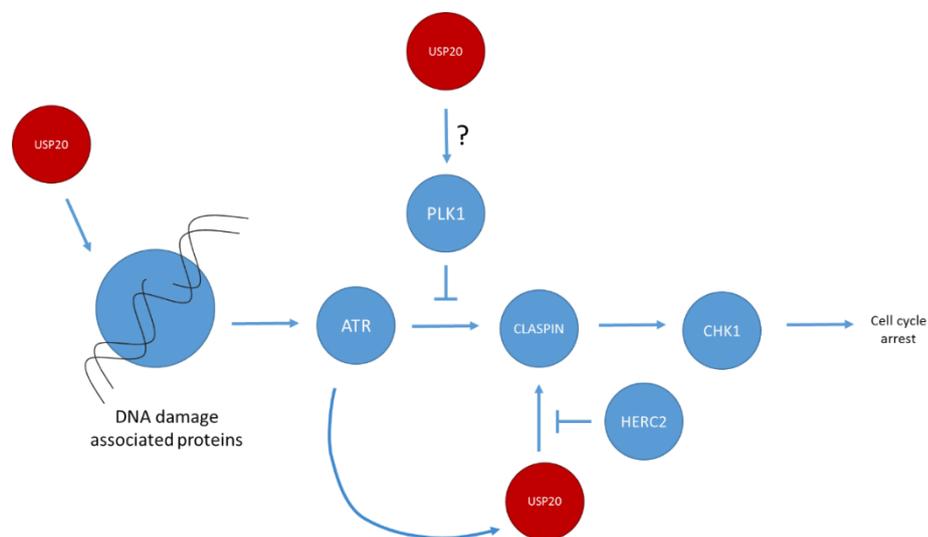


Figure 1.20 . USP20 and DNA damage response. USP20 (red) acts in multiple places in the response to DNA damage. It stabilises RAD17, which is involved in sensing of the insult, stabilises CLASPIN after phosphorylation by ATR, and possibly has a role in stabilisation of PLK1. Interestingly, these roles seem to oppose each other as stabilising CLASPIN leads to cell cycle arrest, and stabilising PLK1 would lead to cell cycle transition. It is likely that some form of regulation is at play to ensure that USP20 is promoting the correct cellular pathway.

These reports show that USP20 is highly integrated in DNA damage response, and has roles in genomic integrity, possibly through its affect on homologous recombination [182]. Through cell line and mouse models USP20 was observed to be an effective tumour suppressor through its role in DDR [182-184].

1.5 Phosphorylation

1.5.1 A brief history of phosphorylation

Although the presence of phosphoproteins was known in the early 1900s, primarily by chemical analysis of egg-yolk [185], the actual processes of phosphorylation and dephosphorylation were not elucidated until much later. The first evidence for the phosphorylation system was observed during research into glycogen metabolism [186-188]. It started in the mid-1930's when Carl and Gerti Cori identified an AMP-dependent glycogen phosphorylase (phosphorylase b). This enzyme was inactive in the absence of AMP. However, Kiessling [189] found another form of the protein in a constitutively active state, independent of AMP concentration (phosphorylase a). After initially discounting Kiessling, Cori and Green [190] then confirmed this by crystallising the active form of the phosphorylase. Their hypothesis was that active phosphorylase a has a covalently-linked AMP prosthetic group. However, problems ensued when they tried to confirm release of this group upon conversion of phosphorylase a to phosphorylase b.

This remained unsolved for the next decade until, in the mid-1950's, Fischer and Krebs [191-193], and Wosilait and Sutherland [194] identified that this conversion included phosphorylation and dephosphorylation events. Fischer and Krebs found that the conversion of the inactive to active form was Mg-ATP dependent and required another enzyme that transferred a phosphate group from the ATP to a serine residue on phosphorylase b. This additional enzyme was termed

phosphorylase kinase, one of the first kinases discovered. However, these groups were marginally beaten to the identification of the first kinase by Burnett and Kennedy [195]. They identified an enzyme that catalysed the phosphorylation of casein. To commend the works of these scientists, Cori and Cori were awarded the Nobel Prize in Physiology or Medicine in 1947 and Fischer and Krebs were awarded the Nobel Prize in Physiology or Medicine in 1992 [196].

1.5.2 Kinase structure and function

Kinases are the enzymes responsible for catalysing the phosphorylation of substrate molecules. In cellular physiology, protein kinases are responsible for post-translational modification of proteins by phosphorylation. Here, the γ phosphate group (PO_3^{2-}) from ATP is typically attached to a hydroxyl group of a serine, threonine or tyrosine residue [197, 198]. The effect of this depends on what substrate protein is being phosphorylated, and what residue of this protein is phosphorylated. The modification may activate or inactivate an enzyme, or promote or preclude protein-protein interactions. This is achieved through conformational change and/or alteration of surface topography and charge distributions. For example, phosphorylation shifts the structural equilibrium of glycogen phosphorylase to an active form by movement of a loop out of the active site in an allosteric manner [199-201].

There are over 500 kinases in the human genome; collectively termed the kinome [202]. Structurally, they all contain a similar fold comprising two lobes: an N-terminal lobe containing a beta sheet and a single helix, and a C-terminal lobe that is mostly alpha helical (Figure 1.11). The cleft formed between these lobes provides the ATP and metal ion binding site. The ATP molecule makes hydrophobic interactions with residues in this cleft and hydrogen bonds to residues in the hinge region that links the N- and C-lobes [203, 204]. A general description of kinase domains will be given, but examples will often include descriptions of PLK1, as an introduction to this protein is required for the results section of this thesis.

Between beta-strand one and two of the N-lobe is a glycine rich loop with high flexibility, known as the P-loop. This region closes over the ATP molecule and aids in the coordination of the metal ions and phosphate groups of the ATP. The helix of the N-lobe (called the C-helix) is also required to be in a specific conformation for catalytic activity. A conserved glutamic acid residue from this helix and the catalytic lysine within the A-X-K motif interact and coordinate the alpha and beta phosphates of ATP [198, 205].

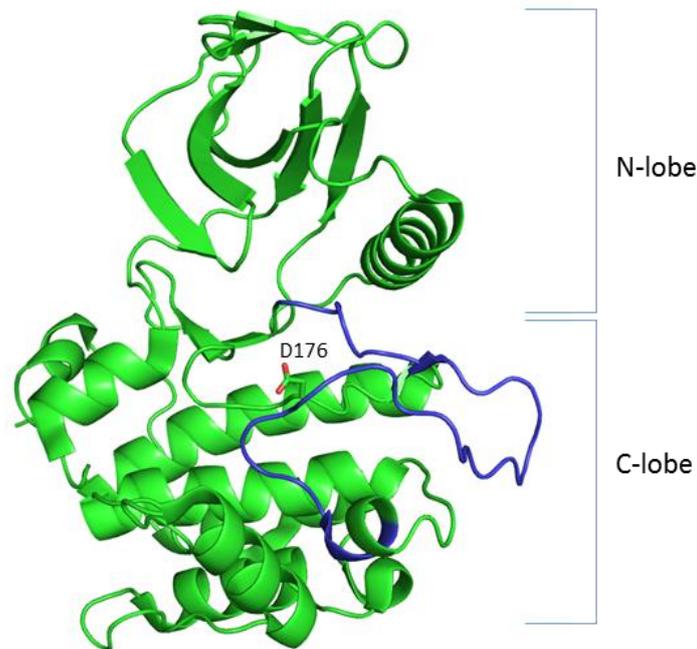


Figure 1.21. PLK1 catalytic domain. The catalytic domain of PLK1 contains the typical kinase fold. The activation loop (from the D-F-G motif to the A-P-E motif) is coloured blue. The catalytic residue, Asp176, is shown in stick form.

The C-lobe also contributes essential residues for catalysis. The [Y/H]-R-D motif and the D-F-G motif are found in this subdomain. The aspartic acid residue of the [Y/H]-R-D motif secures the substrate residue with donor hydroxyl group. The Y/H residue stabilises multiple residues, including the aspartic acid of the [Y/H]-R-D motif and the phenylalanine of the D-F-G motif. The D-F-G motif serves to coordinate the magnesium ion (Asp) and locks in the C-terminus of the C-helix (Phe). The residues from the D-F-G motif to the A-P-E motif form the activation segment. This is a primarily disordered loop that can alter in conformation to produce active and inactive states of the kinase [198, 205].

Kinases employ multiple methods to achieve target specificity. One method is to employ substrate motif recognition. It's common for this motif to bind to the active site of the kinase, where a specific Ser/Thr/Tyr residue will be phosphorylated [206]. Additional residues surrounding this residue interact with the catalytic domain, increasing the affinity and specificity of the interaction. For example, the consensus sequence of the catalytic domain of Plk1 is [D/E]-X-[S/T]-Φ-X-[D/E], where a hydrophobic residue (Φ) at +1 of the phosphorylation site and acidic residues at -2 and +3 are required for optimal phosphorylation [207]. This motif is observed in targets for Plk1, such as Cdc25C [208]. Also, ATR/ATM kinases have a consensus sequence [S/T]-Q [209, 210], where clusters containing multiple S-Q or T-Q sequences are often found in ATR or ATM substrates [211].

Additionally, there may be a distal docking site for the substrate protein; specialised recruitment domains that ensure substrate specificity. These accessory domains recognise motifs or domains on substrate proteins, which are regularly primed by phosphorylation by other kinases. Examples of these include Plk1, which recognises the [P/F]-[Φ/P]-[Φ]-[T/Q/H/M]-S-[pS/pT]-[P/X] motif, or the shorter S-[pS/pT]-[P/X] motif through its polo-box domains (Figure 1.12) [212, 213] and Chk2, which recognises the H-F-D-pT-Y-L-I motif through its FHA domain [214, 215]. By binding to specific linear motifs, these proteins can target multiple, but selected, substrates providing they contain the specific sequences. In addition, the actual peptide substrate that binds to the catalytic domain may have K_d s in the mM range. Having these extra groups ensures binding of their substrate proteins.

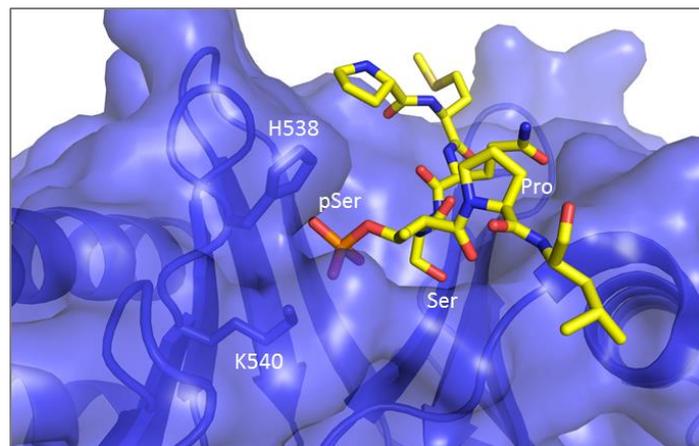
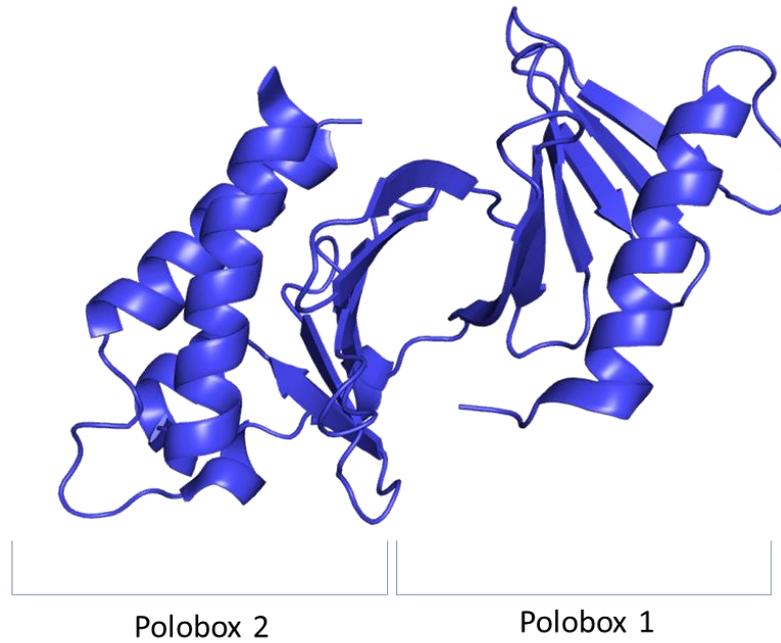


Figure 1.22. PLK1 poloboxes. Left shows the apo crystal structure of the poloboxes (PDB code 1Q4O). Each polobox domain can be seen clearly, with their β -sheets packing against each other. Right shows the crystal structure of the poloboxes bound to a peptide containing the Ser-pSer-Pro motif (labelled in image; PDB code 1UMW). The pincer residues of PLK1, H538 and K540, are shown in stick form. These residues are crucial for coordinating the the phosphoserine. The structure clearly shows the negatively charged phosphate buried and interacting with the positively charged lysine residue. The histidine residue contributes hydrogen bonds to the phosphate. The serine in the P-1 position is crucial for binding of phosphopeptides. It also projects into the pocket forming hydrogen bonds with the polobox domains.

1.6 Aims

Studies that have investigated the role of USP20 in the cell have only obtained cellular data, such as immunoprecipitation experiments or yeast two-hybrid. In only a few cases are the interactions mapped to USP20 or USP33, and there has been no investigation into the specific residues or even small regions of the protein that mediate the interactions. Therefore, obtaining structural data of USP20 would be invaluable to understanding how USP20 interacts with the proteins. In addition, further characterisation of the currently known interactions or new interactions would further the understanding of USP20's role in the cell.

The aims of this thesis are to characterise the deubiquitinating enzyme USP20. These include:

- Dissecting the domain structure of USP20
- Exploring the interactome of USP20

1.7 Objectives

For dissecting the domain structure of USP20 the following objectives will be attempted:

- Cloning and bacterial expression of full length USP20 and its domains
- Purification of these constructs
- Obtain crystals for X-ray crystallography

The interactome of USP20 will be investigated by:

- Using a yeast two-hybrid system and *in vitro* assays to characterise known and putative interactions of USP20.
- Using pull down assays and mass spectroscopy to identify new binding partners for USP20.
- Investigating the interaction between the USP20 zinc finger domain and ubiquitin.

To achieve these objectives, many techniques were used. In order to ensure that the results are clear, an introduction to methodologies used in this thesis will be given.

2 Introduction: Methods

2.1 Protein design and expression

Recombinant protein expression is the primary source of protein for crystallography. The basic process is as follows: a gene is cloned into an appropriate vector for bacterial expression. The vector is transformed into the bacteria and the cells are selected for using selectable markers, such as antibiotics. The cells are then grown in liquid cultures where induction of gene transcription is performed. The cells are harvested after a specific time period and lysed to release the cell contents prior to purification of the protein of interest [216].

The choice of vector is dependent on a number of factors including organism (most commonly *E. coli*), construct design, vector compatibility, selectable marker compatibility, induction method, expression conditions and restriction sites in the multiple cloning site. Depending on the nature of the protein, it may be necessary to express the protein of interest as a fusion with a protein tag.

2.1.1 Protein tags

The purposeful addition of amino acids to a protein of interest is commonly used to enhance its expression and purification. These additions are often called tags, which can consist of a few disordered amino acids or fully folded domains. The location may be on the N- and/or C-terminus, or inserted within a loop of the protein. The choice of tag depends on what the ultimate purpose is, and some tags are not compatible with certain proteins; a factor that is impossible to predict. Some tags are used only for purifying the protein (affinity tags), some only for

enhancing soluble protein expression (solubility tags), some can achieve both. A summary of tags that are commonly used is given in Table 2.1.

Of the peptide tags, the hexa-histidine tag is most commonly employed. It is used for purification of proteins using immobilised metal-ion affinity chromatography (explained further below); its effect on protein solubility is protein and location dependent, but generally produces a small negative effect on solubility [217]. In addition, poly-Arg-tags are used to aid in ion exchange chromatography. It achieves this by providing a large basic chain that increases a protein's isoelectric point so that it can effectively bind to a cation exchange column [218]. These two tags are the only commonly small tags with repetitive amino acid sequences, and they both boast cheap and effective purification protocols. Other peptide tags include those primarily used for purification by means of protein binding. These include Myc, FLAG and HA epitope tags that are detected by antibodies, and also include peptide sequences that have high affinity for non-antibody proteins, such as the Strep-tag for Streptavidin binding. Beads or columns loaded with antibodies or proteins are used to selectively isolate tagged proteins. For this reason, they are expensive to use, but highly specific.

Overall, the cost becomes a major factor in the decision of which tag to use. The estimated cost for purifying 10 mg of a 30 kDa protein with nickel NTA resin with a His-tag is £15. For GST- and MBP-tags, the costs are £25 and £8, respectively. These low costs contrast greatly with those of epitope tags, where such purifications can cost in excess of £1000 [219].

Table 2.1. Commonly used fusion tags.

Tag	Length	Size (kDa)	Purpose	Resin-bound ligand/protein
Histidine	6	0.84	Affinity	Metal ions (SM)
FLAG	8	1.01	Affinity	Anti-Flag (mAb)
Strep II	8	1.06	Affinity	Strep-Tactin (P)
HA	9	1.1	Affinity	Anti-HA (mAb)
C-Myc	11	1.2	Affinity	Anti-Myc (mAb)
SUMO1	101	11.5	Solubility	N/A
Thioredoxin	109	11.8	Solubility	N/A
T4 lysozyme	164	18.7	Stability/Crystallisation	N/A
GST	211	26	Affinity/Solubility	Glutathione (SM)
MBP	396	42	Affinity/Solubility	Amylose/Dextrin (SM)

SUMO1 [220] Table adapted from [221].

The use of larger fusion tags for crystallography purposes is termed carrier-driven crystallisation; the use of fusion proteins to enhance the crystallisation of proteins that won't typically crystallise. T4 lysozyme crystallises easily, and has been used as a crystallisation carrier. Major success has been observed in the crystallisation of G-protein-coupled receptors where T4 lysozyme is inserted into a loop of the transmembrane proteins to increase the surface available for crystal contacts to form [222, 223]. GST has also been used as a carrier, but successful crystallisation and structure determination is limited to very small domains and peptides only [224]. A novel take on carrier driven crystallisation includes the use of polymerisation modules by the formation of protein polymers where the crystal contacts are primarily formed by the carrier [225, 226].

As of 2015, only three fusion structures with thioredoxin and six with GST have been produced. The most commonly used carrier is MBP, with over 100 fusion structures [227]; a massive increase on the three MBP-fusion structures solved by 2003 [228]. Most of these crystal structures have been made using fusions of the short, fixed-arm linker form of MBP fusion designed by Moon *et al.* [229]. This highlights the issue with current protein carriers used for crystallisation. GST and MBP both work incredibly well as solubility tags, but they are both also relatively large proteins with inherent flexibility, among other issues. The MBP in these fusions has had its surface entropy reduced and its C-terminus made more rigid to improve its ability to crystallise. Although these modifications have improved the ability to crystallise low solubility proteins that require the presence of a solubility tag, their issues still remain a problem. It may be that the use of protein tags that are highly soluble, small, non-enzymatic and readily crystallisable may be the way forward in carrier-driven crystallisation. This could be achieved by screening the PDB, a database of solved structures, for proteins that could have potential as tags for carrier driven crystallisation.

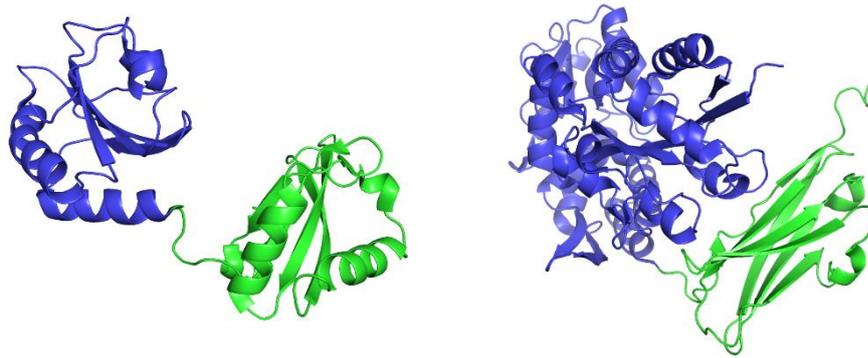


Figure 2.1 Fusion structures. (Left) Thioredoxin fusion with the UHM domain of Puf60 (PDB code 3DXB). **(Right)** MBP fusion with a fragment of myosin-binding protein c (PDB code 4EDQ). The thioredoxin and MBP tags are shown in blue, the fusion protein in green.

One issue with using fusion tags is the possibility that the tag could affect the fold of the fusion protein. For most small tags, such as the His-tag, this does not seem to occur [230], but in some cases small peptide tags can affect the protein structure [231]. A search of the PDB shows that over 23000 structures contain a His-tag sequence, further indicating the fact that small tags are not detrimental to the structure and crystallisation of proteins in the majority of cases [232]. For larger tags such as MBP, only one case has been identified where the structure was different to that of the un-fused protein. However, only the location of two independent domains was altered, not the independent fold of the domains [227].

2.2 Protein purification

2.2.1 Affinity chromatography

2.2.1.1 Immobilised metal-ion affinity chromatography

Immobilised metal-ion affinity chromatography is normally achieved using a polyhistidine tag (usually 6 consecutive histidines) and a resin that coordinates nickel ions. Other metals may be used, such as zinc or cobalt. The solid phase is usually nitrilotriacetic acid-agarose resin, which uses a nitrogen and three oxygens to coordinate nickel ions. Two histidines can then donate coordination bonds via the imidazole τ nitrogen. The polyhistidine tag has a nM affinity for the nickel-bound resin, and thus protein in the mobile phase binds to the column. The interaction between the poly-His-tag and the nickel is mediated by enthalpic contributions. The hexa-histidine tag is the optimal with an affinity (K_d) of free hexahistidine for nickel-NTA of 10 nM. In reality, through steric and electrostatic effects, the affinity of hexa-His-tagged protein is lower. Interestingly, over-increasing the length of the His-tag can reduce the affinity because, eventually, the entropic costs outweigh the enthalpic gain. The exact point at which this occurs is protein dependent, and would depend how what kind of self-interactions were present and how accessible the His-tag is [233, 234]. The high affinity of the His-tag for the nickel can be exploited for purposes of elution. Elution of the protein can be achieved by flowing EDTA, metal ions and histidine analogues (e.g. imidazole) through the resin, or lowering pH, which protonates the nitrogen and disrupts the coordination bond [235, 236]. Unlike most affinity tags, polyhistidine

can be particularly useful when purifying proteins in denaturing conditions. As the tag can still bind the nickel in these conditions, it allows purification of inclusion body proteins [237], repression of undesirable enzyme activity [238], and more sophisticated methods of refolding [239].

2.2.2 Size Exclusion Chromatography

Size exclusion chromatography, generally known as known as gel filtration (GF) in biological sciences, is a method of separating molecules in solution. It is mostly used for large biopolymers, such as proteins. The principle is that the solid phase has pores of variable sizes. As the mobile phase moves, molecules that are too large will pass through the column without entering the solid phase pores. The volume it takes for these molecules to elute is the total volume outside the porous matrix, known as the void volume (V_0). With decreasing size, the number and volume of pores accessible to the molecules increases. This means that smaller molecules have access to more volume of the solid phase pores, and therefore elute later. Unlike most affinity chromatography methods that are primarily enthalpy mediated, gel filtration is (in perfect conditions) mediated by the entropic component only. Molecules of larger size have reduced conformational freedom when entering the pores, and thus have lower entropy than small molecules that are not restricted in their freedom. This negative entropy mediates the exclusion of large molecules from the matrix producing the observed separation of molecules by size [240].

The elution volume (V_e) is the V_0 plus the fraction of the remaining column volume that the protein can access, defined by the equation:

$$V_e = V_0 + K_{av}(V_t - V_0)$$

Where K_{av} is the partition coefficient, the proportion of available matrix volume and V_t is the total column volume. Specifically, the elution volume will correlate with the hydrodynamic radius of the protein [241]. This is the radius of the encapsulating sphere formed by the tumbling of the hydrated protein in solution. The hydrodynamic radius correlates with the molecular weight of the protein because generally as the protein increases in mass, the size will increase too. However, the asymmetry of the protein also has influence on the hydrodynamic radius. For example for two proteins of exact mass, one of which is perfectly globular and the other is an elongated fibre, the elution volumes could vary dramatically due to the much larger hydrodynamic radius of the fibrous protein in comparison to the globular protein. Nonetheless, protein molecular weights can be approximated using gel filtration. Using elution volumes from a set of calibration proteins, the calculated K_{av} s can be plotted on a scatter graph against the log of their molecular weights. Linear regression can give an equation that can be used to estimate the molecular weights of a protein with a given elution volume by calculating its K_{av} . This process can also be reversed to give a predicted elution volume for a given molecular weight protein. As the calibrations are mostly

performed with globular proteins, approximations will be less accurate for elongated proteins.

The gel filtration of a protein sample does not only purify a protein of interest, it can also give valuable information with regards to the oligomeric state of the protein. For example, if the elution volume is lower than that of the predicted value, this could indicate an oligomeric form, provided that the protein is not expected to be particularly elongated. Also, if the protein elutes in the void volume, then this could indicate protein aggregation. This information is particularly useful when the end aim is crystallisation of the protein as it can give an indication of monodispersity, which is important for crystallisation [242, 243].

2.3 Crystallisation and X-ray-diffraction

X-ray crystallography is the use of X-ray diffraction by a crystalline substance to investigate the structure of the molecule or molecules that form the crystal. Using X-ray detectors, a series of images are obtained by rotating the crystal during its exposure to the X-ray source; most commonly synchrotron radiation. With good quality crystals, the images will show a diffraction pattern: arrays of spots formed by diffracted waves, known as reflections, which differ in their intensities. The intensities and locations of these reflections give most of the information about the structure of the crystal, such as the size and shape of the smallest building block of the crystal (unit cell), the symmetry of the diffraction pattern (Laue group), and the structure factor amplitudes (part of the structure factors – the mathematical description of diffracted waves from a crystal). With this

information alone, the structure of the molecules within the crystal cannot be determined. The other part of the structure factors, the phase, is not obtained in the diffraction images; known as the phase problem. Instead it must be derived by performing more experiments or by estimating the phases using a known structure that is similar to the protein forming the crystal. All of this data can then be used to produce a final 3D map of the electron density of the unit cell, in which atoms are built to produce the final structure.

2.3.1 X-ray crystallography history

X-ray crystallography was developed over a century ago by the culmination of research from multiple scientists, including Max von Laue, William H Bragg and William L Bragg. von Laue discovered the diffraction of X-rays from crystalline materials, and won the 1914 Nobel Prize in Physics. The Braggs, father and son, went on to use this to solve the first crystal structure – NaCl - in 1913 [244]. For their contributions to this field they won the 1915 Nobel Prize in Physics. This breakthrough paved the way for field of structural biology, however due to the complexity of proteins, it took a long time for the first protein structure to be solved. This was due to the phase problem, for which a new method of calculating the phases of the structure factors was required. Max Perutz found a way to do this - he used normal crystals and heavy metal derivatives to visualise differences in the diffraction patterns, which allowed calculation of the phases [245]. Subsequently, the first protein structure, myoglobin, was published in 1958 [246].

Since then thousands of protein structures have been solved by X-ray crystallography, and depositions into the centralised repository, the protein databank (PDB) [232], have been increasing at an exponential rate [247].

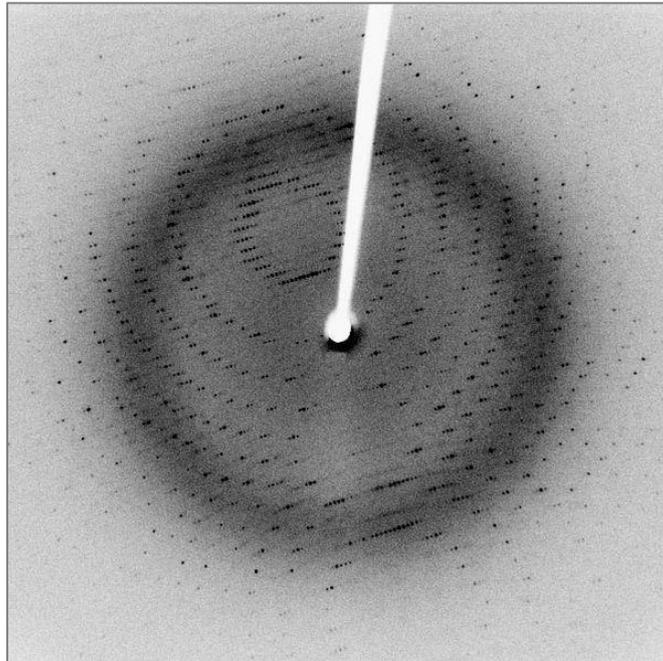


Figure 2.2. Diffraction pattern of crystallised 3Clpro. The diffraction pattern obtained from one exposure of 3Clpro crystals. The reflections can be seen as arcing spots. Author Jeff Dahl, via Wikimedia Commons.

2.3.2 Obtaining crystals

For protein crystallography, the vapour diffusion method is often used, but other less frequent methods include microbatch under oil, dialysis and free interface diffusion [248, 249]. There are hanging drop and sitting drop formats of vapour diffusion, but both use the same principle. A well contains the precipitant solution

and a small drop is produced that typically contains a ratio of 1:1 protein sample:well solution. The protein in the solution must be supersaturated, which is an issue for many proteins as achieving these concentrations can be difficult. It is usually performed in some of microtitre plate, often in a 96-wel format to maximise the different conditions for screening. Sitting drop vapour diffusion uses a chamber with a pedestal for the drop to 'sit' on. This is adjacent to the well solution, and the whole chamber is sealed. Vapour diffuses out of the drop, which concomitantly increases the protein and precipitant concentration. Vapour also diffuses out of the well solution, but to a lesser rate than the drop. When the osmolarity of the drop is equal with that of the well solution (equilibrium), there will be no net movement of water out of the drop, so it will stop dehydrating [250]. The hanging drop method differs only in that the drop does not sit in a well, but is put on the seal, and hangs above the well. The crystallisation process can be shown using a phase diagram (Figure 2.3A) [251]. Depending on the protein and precipitant concentrations, the likelihood of nucleation and crystal growth vary. If the concentrations are at the right levels at the initial stage of the crystallisation, then the concentrations may increase until the drop enters the nucleation zone. As nucleation occurs, protein is removed from the solution, so the effective protein concentration lowers. It may then enter the metastable zone where the crystals will grow until the concentration of protein drops out of this zone. This process is shown as a red arrow in 2.3B. The amount of nucleation is crucial to producing adequate crystals for diffraction. Too much nucleation, and the drop will contain too many small crystals [252, 253].

Large screens with many conditions are used because it is unknown what condition will favour protein crystallisation. Each condition will typically contain a precipitant. The most commonly used precipitants are salts, polymers and alcohols [254, 255]. In addition they may contain many other compounds such as buffers and additives. Once crystals form, optimisation of crystallisation is achieved by altering the conditions that crystals are grown in. The concentrations of any of the components of the condition, or pH, may be changed. This can alter the crystal form or size to improve diffraction. Nucleation can be controlled by altering many of the factors involved in vapour diffusion, including the rate of diffusion, temperature and materials used [252, 253].

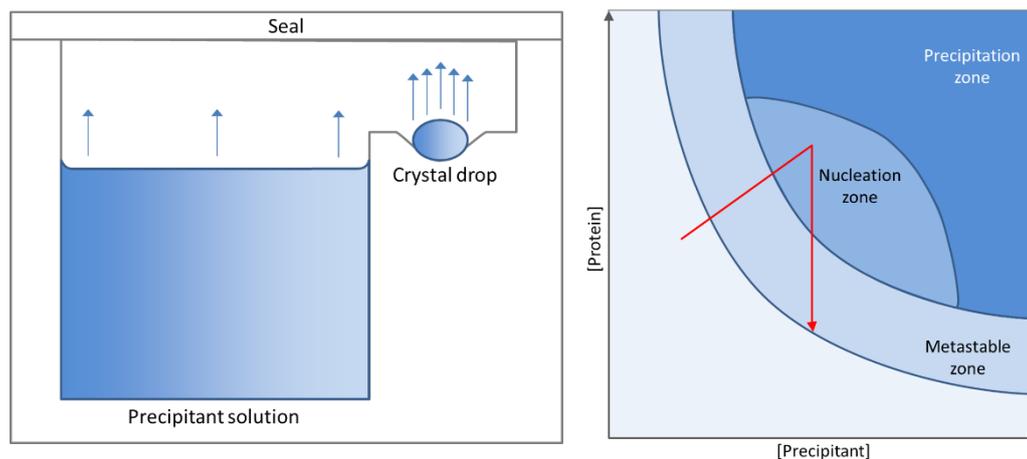


Figure 2.3. Vapour drop diffusion and the phase diagram. (A) A diagram of a single well in a crystallisation screen. **(B)** The phase diagram shows the concentration of protein and precipitant concentration in a crystallisation drop. The concentrations follow the path of the red arrowed line over time if crystallisation occurs. As the drop starts to evaporate, the protein and precipitant concentrations increase concomitantly. If nucleation occurs (the concentrations enter the nucleation zone), the protein is removed from the solution as the crystal grows.

2.3.3 Cryocooling

Cryocooling is required to reduce radiation damage to the crystal upon exposure to the X-ray beam. Radiation damage reduces the diffraction of the crystal, can alter the structure of the protein forming the crystal and can alter the unit cell size. To reduce this damage, the crystals are typically cooled to around 100K, where the diffusion of the damaging free radicals is slowed. The issue with cryocooling is that water can crystallise during the cooling process, which can lead to damage of the crystal and affect diffraction. The solvent in and surrounding the crystal must be vitrified to prevent damage. Anti-freeze agents, known as cryoprotectants, are added to the mother liquor that the crystal grew in and the crystal is soaked in this solution. The cryoprotectants prolong the time that the crystal can be cooled for before crystallisation of the water occurs. The most common cryoprotectant used in macromolecular crystallography is 20-30% glycerol, and works in around 2/3 cases. Other commonly used cryoprotectants include ethylene glycol, polyethylene glycols, salts and 2-Methyl-2,4-pentanediol. The cryoprotectant-soaked crystal is then 'fished' using a nylon cryoloop and plunged into liquid nitrogen prior to mounting on a goniometer and exposing to the X-ray beam (during which it remains inside a stream of liquid nitrogen) [256, 257].

2.3.4 The crystal and symmetry

The crystal is formed from an array of identical repeating units, known as the unit cell. This is in turn built up by symmetry repeats of an asymmetric unit. The unit

cell is defined by its edge length, a , b and c , and the angles between the axes, α , β , γ . These parameters define the overall lattice type of the crystal. Seven primitive crystal systems and additional lattices with translational centring (volume, base or face) make up the 14 Bravais lattices. Primitive lattices only have one lattice point per unit cell and non-primitive lattices contain either two lattice points (volume and base centred) or four (face centred) [258, 259].

Unit cells are built up using symmetry operations on the asymmetric unit. Crystallographic symmetry elements include identity, rotation, inversion and reflection. However, as proteins are chiral molecules, the only elements possible are identity and rotation. Point groups describe the whole symmetry of a finite object, showing the combinations of symmetry elements present. There are in total 32 point groups, only 11 of which can describe symmetry in chiral objects. The Bravais lattice (Figure 2.4) describes the translational symmetry in the crystal, and together with the point group gives the space group of the crystal. In total there are 230 space groups, 65 of which can describe chiral crystals [258, 259].

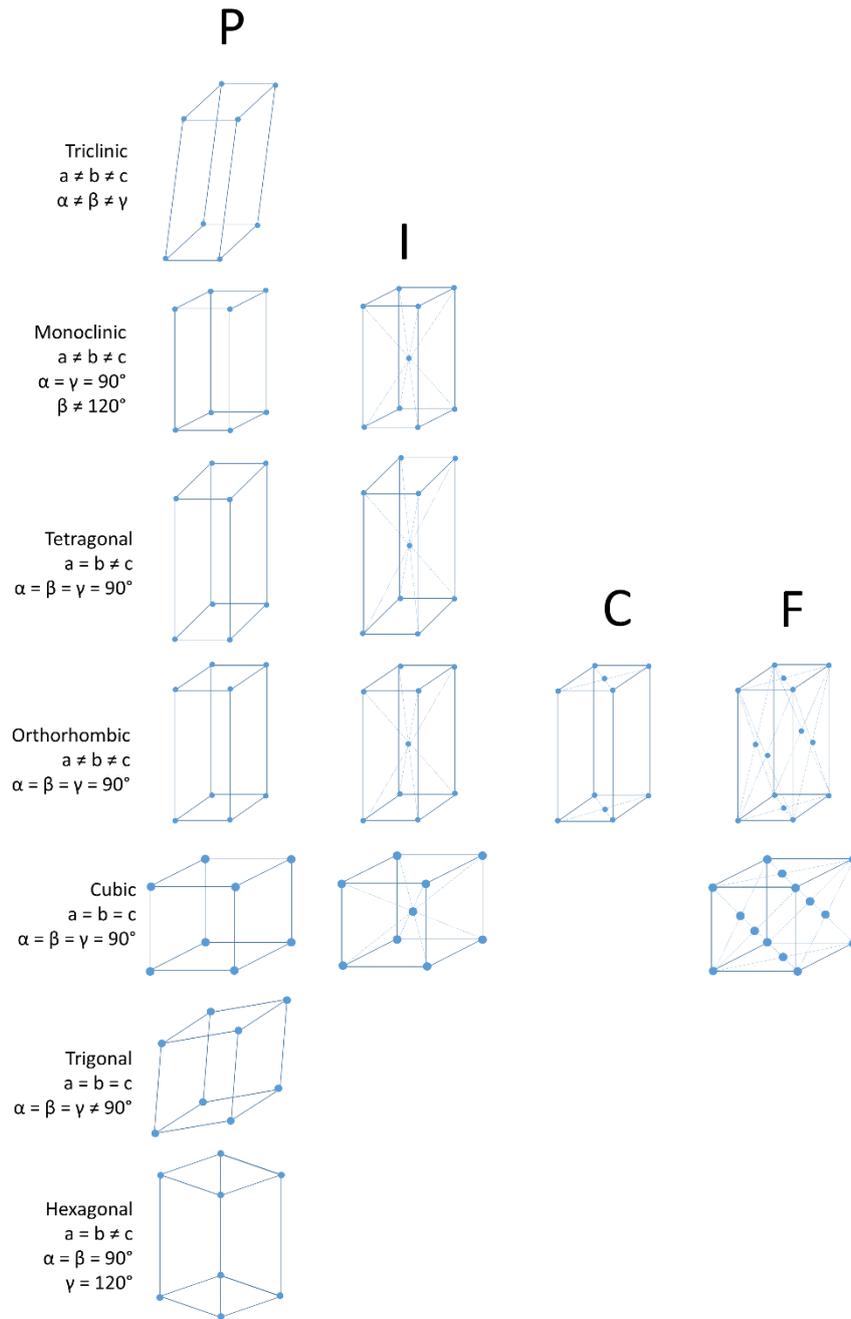


Figure 2.4. Bravais lattices. The 14 Bravais lattices showing the possible translational symmetry that can occur in a crystal.

2.3.5 X-rays and diffraction

X-rays used in crystallography may be obtained from home-source beams that use rotating anode X-ray tubes. Here, a metal cathode is heated and electrons are

released by thermionic emission. Electrons hit a rotating metal anode and cause the release of X-rays due to their rapid deceleration (Bremsstrahlung radiation) and knocking out inner-shell electrons of the anode atoms, which are replaced by higher energy outer-shell electrons (characteristic radiation). A beam is formed by collimation of the X-rays. [260, 261].

Synchrotrons offer a much brighter light than home sources and are more commonly used for macromolecular crystallography due to the speed in which datasets can be obtained. X-rays are produced in a different way in synchrotrons than in home sources. They are particle accelerators that consist of a linear accelerator and two rings: a booster ring and storage ring. Bending magnets in the storage ring of the synchrotron cause relativistic electrons to 'wobble', which causes the release of X-rays. Beam lines are found at tangents all around the ring, where the X-rays are focussed and collimated to produce small area, high intensity X-ray beams that can be used for crystallography.

X-rays interact with electrons in atoms within crystals and are scattered in all directions. When the X-rays are scattered by atoms on parallel planes, known as Miller planes, they may interact constructively to produce a reflection on an X-ray detector. The parallel planes may intersect any of the three axes a number of times, and is designated by the integral parameters h , k and l . The angle of the planes from the incident X-rays change when the crystal is rotated in the X-ray beam. When the difference in path length of reflected X-rays from two planes is an integral number of wavelengths (Figure 2.5), maximal constructive interference is

observed [258, 259]. This is known as Bragg's Law, and is defined by the following equation:

$$n\lambda = 2d \sin \theta$$

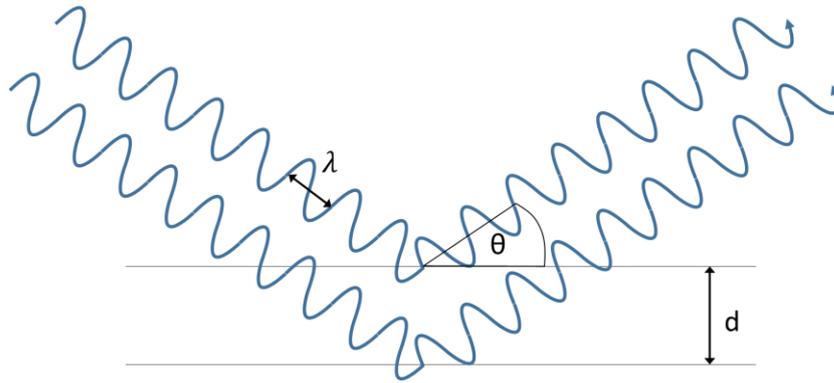


Figure 2.5. Bragg's law. Waves that are diffracted from two planes with a distance, d , will be in phase, provided that they satisfy Bragg's Law: $n\lambda=2d \sin\theta$.

The diffraction pattern obtained is indexed, meaning the reflections are assigned with their Miller indices. The structure factor equation $F(hkl)$ is used to describe the diffracted wave that produces the reflection. The summation of scattering is taken for all atoms (j) in the unit cell with coordinates x, y, z for miller indices, h, k and l . The result is a vector F with amplitude $|F(hkl)|$ and phase $\phi(hkl)$ [258, 259] (Figure 2.6).

$$F(hkl) = \sum_j^n f_j e^{2\pi i(hx_j + ky_j + lz_j)}$$

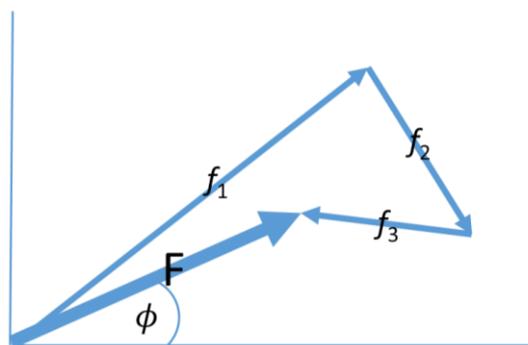


Figure 2.6. Structure factors. An Argand diagram is shown where the x-axis corresponds to the real axis, and the y-axis corresponds to the imaginary axis. Each atom contributes to the final structure factor. By plotting the vector of each atom, f_n then the final structure factor, F , can be seen as a sum of the individual vectors. The amplitude of the structure factor is the length of the vector. The phase of the structure factor is the angle between the vector and the real axis.

The electron density of the unit cell is calculated by the following equation:

$$\rho(xyz) = \frac{1}{V} \sum_{\substack{hkl \\ -\infty \\ +\infty}} |F(hkl)| e^{-2\pi i[hx+ky+lz-\phi(hkl)]}$$

This equation is a summation of the structure factors over three dimensions. The amplitude $|F(hkl)|$ is obtained through the single diffraction experiment. However, the phase $-\phi(hkl)$ must be acquired using either computational or experimental methods. More recently, the most common form of phasing is molecular replacement. Here, a model protein that has already been solved is used to estimate the phases by rotating and translating the model within the asymmetric unit. Statistical methods are used to identify whether a good placement is achieved. For this method to work, structural homology should be high; usually with a sequence identity cut-off of around 30%. Experimental methods include isomorphous replacement and anomalous dispersion. They all

utilise the fact that the presence of heavy atoms within the crystal can affect the diffraction observed. These differences can be used to solve the heavy atom sub-structure, which can then be used to solve the structure of the whole protein. Getting accurate phases is crucial, as the phases contribute more to the final electron density than the amplitudes [258, 259]. Figure 2.7 shows how vital calculating accurate phases is.

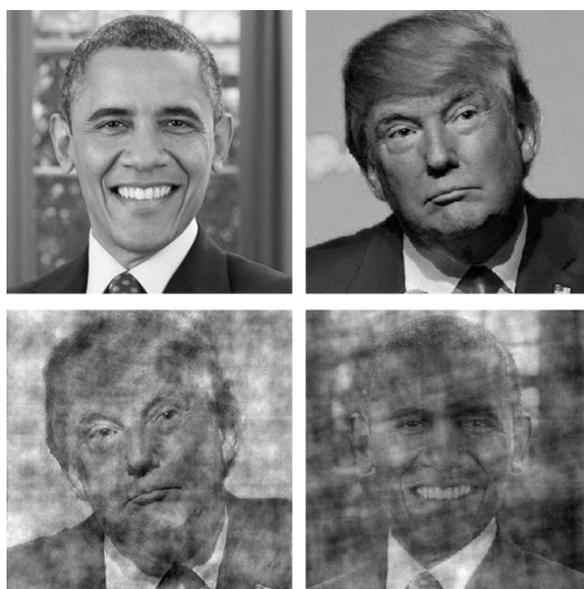


Figure 2.7. Phases and amplitudes. A discrete 2D Fourier transform can be made of the two top images. If the phases of Donald Trump's image are used with the amplitudes of the Barack Obama image, then the image bottom left is obtained. If the phases of Barack Obama are used with Donald Trump's amplitudes, then the image bottom right is obtained. By using the correct amplitudes, analogous to the reflection data obtained during crystallography, but the wrong phases, the final product mostly represents the image that was used for the phases. This is important for crystallography where the phases are obtained through experimental methods or molecular replacement. Poor data, or a poor model for molecular replacement will cause problems in the calculation of the true electron density. Images were made using FTL-SE software.

2.4 Hybrid assays

Hybrid assays are a simple, but elegant, tools for in-cell interaction studies. Many reviews into their biology and uses have been published [262, 263]. Their many variants allow them to be used for screening of interacting proteins (and other biological molecules) as well as characterising known interactions.

Hybrid assays are named depending on the organism of use (e.g. mammalian or yeast) and the number of interactions being assessed (as of 2016, up to a yeast four-hybrid has been described [264]). Most commonly, yeast two-hybrids are used for assessment of binary interactions. The premise of the assay is that a reporter gene is expressed upon interaction of two interacting proteins. This is achieved by fusing one of the interacting proteins to a DNA-binding domain, and the other to a transcription activating domain (Figure 2.8). This was first performed by Fields and Song [265] in 1989 where GAL4, a yeast protein that contains both DNA-binding and transcriptional activating domains, was split into its cognate parts and each fused to two proteins known to interact. These interacting hybrids reconstituted GAL4 and induced expression of the reporter gene by localising the transcription activating domain to its up-stream activating sequence (UAS).

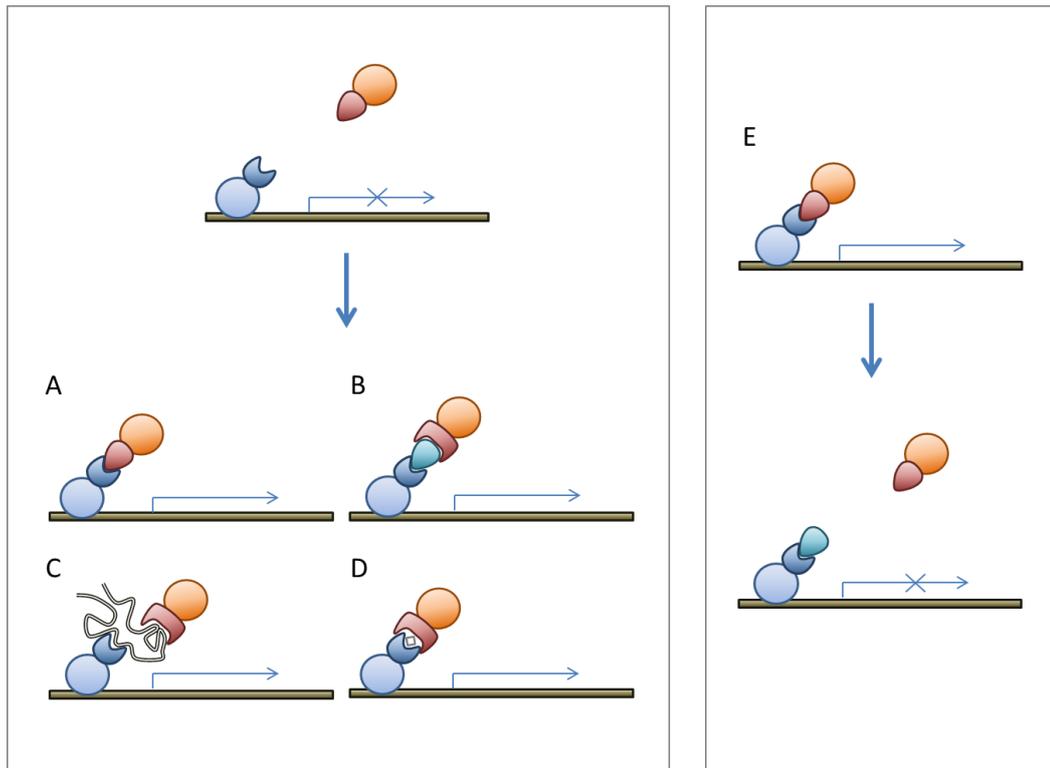


Figure 2.8. Yeast two-hybrid. (left) If the proteins do not bind, then transcription of the reporter gene is not activated. In a typical yeast two-hybrid, the two proteins interact and activate transcription (A), or yeast three-hybrid variations can utilise a bridging molecule of protein (B), nucleic acid (C) or small molecule (D). (E) A reverse yeast two hybrid can be used where interaction of a third molecule with the bait is detected by loss of transcription.

Since this initial demonstration of the two-hybrid method, many variations have been made. These include different reporter genes and alternative fusion systems – i.e. proteins that bind to different UASs or alternative transcription activating domains. Reporter genes that have been used include chromogenic reporters (LacZ [265], GusA [266], MEL1 [267], LacA3 [268]), prototrophic reporters (LEU2 [269], URA3 [270], HIS3 [271], ADE2 [272], LYS2 [266]), fluorescent reporters (GFP [273]) and antibiotic resistance reporters (AUR1-C [274]). In addition to the GAL4 DNA-binding domain (DBD), the LexA DBD [269], oestrogen receptor DBD [270],

bacteriophage repressor [266] and Tet repressor [275] have been used to bind to DNA upstream of the reporters. Also, B42 [269] and VP16 [271] have been used as transcriptional activation domains (TAD). DNA-independent, protein-fragment complementation assays have also been used [276-279]; where interaction of the fusions reconstitute two fragments of a protein reporter. Various methods are then used to detect the reconstituted protein. In this thesis, a LexA-VP16 Y2H system is used [269, 271] to characterise binding of USP20 domains and proteins that are thought to interact. This system uses the L40 strain of yeast with LexA binding sites upstream of HIS3 and LacZ reporter genes. This strain, in combination with the LexA proteins, allows the use of selective growth assays or β -galactosidase assays [280].

The yeast two-hybrid assay is often used to validate or characterise a specific binary interaction. Used in this way, the general process is to transform the yeast with the two DBD and TAD fusions. Once transformed, the interaction will be detected by expression of the reporter gene. In addition, the yeast two-hybrid system can be used to screen for interacting proteins. This can be achieved by using one protein as a bait for a cDNA library of prey fusions [281]. Here, a single bait and the library will be transformed into cells and then cells with interacting proteins will be identified by activation of the reporter gene(s). The protein or protein fragment that is producing an interaction will be identified by sequencing of the plasmid in the identified cells [282].

Yeast three hybrid systems are used to detect interactions between 3 or more proteins. In these assays, a non-fused protein will be transformed and act as a bridge between the two fusion proteins [283]. Three hybrids have also been used for non-protein bridging molecules, including small molecules [284] and RNA [285]. Interactions between DNA and proteins are achieved using a one-hybrid approach: a protein or library would be fused to a TAD, which would activate transcription of a reporter gene [286].

There are many benefits and disadvantages of using the yeast two-hybrid system. One of the main issues is autoactivation of the system by DBD fusions. This occurs when the bait protein fused to the DBD has transcription activating properties, irrespective of binding to the TAD fusion. This prevents its use as a DBD fusion, but can be resolved by reversing the system; swapping it to a TAD fusion.

The use of fusion proteins also provides issues on a physical level, steric issues and incorrect protein folding may occur, preventing true interactions from being detected. Steric hindrance can be rectified in some cases by alternative protein design, such as alternative fusion proteins, longer linking sequences, or fusing the bait/prey at a different terminus [287]. Depending on the host organism in the assay, correct post-translational modification that is required for the interaction may not be present, and therefore a valid interaction may not be detected. Modified hybrid assays such as coexpression [288] and fusion [289] of enzymes required for PTM have been successfully used to characterise systems in which modifications affect interaction.

Unlike some other interaction assays, the yeast two-hybrid cannot be used to calculate thermodynamics of the interaction. The signal of the reporter gene does correlate with the affinity of the interaction, but it varies with respect to the reporters, upstream activating sequences, DBDs, TADs, directionality of fusions and the proteins being investigated. This makes it difficult to conclude the affinity of the interaction with yeast two-hybrid data alone [290].

Advantages of the system over other protein interaction methods are that it is a relatively quick, cheap and labour non-intensive. Also, unlike most biochemical techniques, it does not require large quantities of pure protein, which is often difficult in recombinant systems. In addition, the yeast two-hybrid system is useful for the detection of weak and transient interactions (the GAL4 system can detect interactions with a K_d of $\sim 70 \mu\text{M}$ [291]).

2.5 ITC

Isothermal titration calorimetry (ITC) is a technique used to characterise the thermodynamics of binding interactions between biomolecules. In one assay, the enthalpy (ΔH), equilibrium constant (K) and stoichiometry (n) can be measured for an interaction. It can be used for interactions between proteins, DNA, RNA and small molecules.

It was first used for investigating the thermodynamics of acid-base titration and metal-ion coordination in the 1960s by Christensen *et al.* [292-294]. With better equipment, the first ITC paper using protein was performed in 1979 by Beaudette and Langerman where they investigated the binding of ADP with bovine liver glutamate dehydrogenase [295]. With modern day titration calorimeters, it is becoming much easier to analyse protein interactions using ITC.

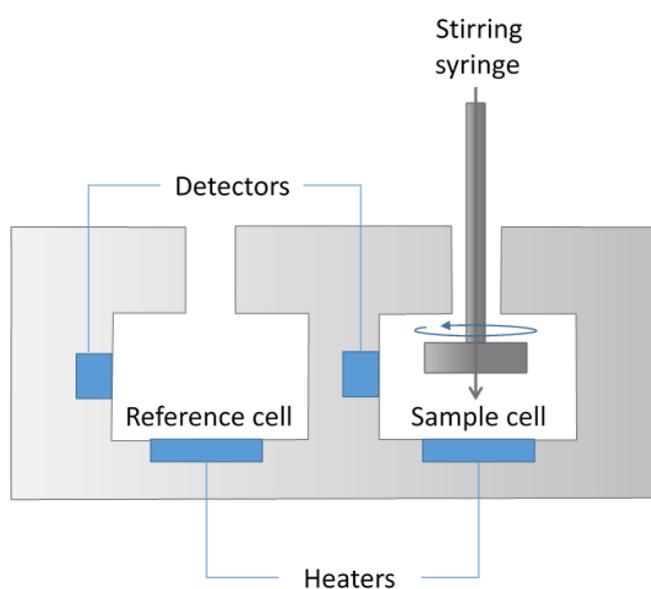


Figure 2.9. Isothermal titration calorimeter. The diagram shows a typical schematic of a calorimeter. The stirring syringe is found in the sample cell so that the sample is mixed adequately upon injection. The heaters maintain a specific temperature set by the user. The temperature of the cell is detected and the heaters are on a feedback loop to maintain the temperature. The difference in power output between the two heaters is used in the calculations of the binding parameters.

A general schematic of a titration calorimeter is shown in Figure 2.9. There are two cells present. The sample cell is filled with the titrand. The stirring syringe is filled with the titrant. A stepper motor is used to gradually titrate the titrant into the

cell and the stirring ensures homogeneity. The detectors constantly measure the temperature of the cells and the heater ensures that both cells are maintained at a constant temperature. Assuming binding of the molecules in the syringe and cell, heat will either be released or absorbed depending on whether the interaction is endothermic or exothermic, respectively. This will cause a decrease or increase in the temperature of the cell, which is detected, and the power to the heater of the sample cell is increased or decreased to maintain the set temperature [296]. The injection peaks on the thermogram are integrated and the plot of energy per mole (usually kcal/mol) against ratio of titrant to titrand concentrations are used to calculate ΔH (*difference between upper and lower plateaux*), K_a (gradient at mid-point) and stoichiometry (the ratio of concentration at the mid-point) (Figure 2.10).

In addition to the direct calculations of ΔH , K_a and stoichiometry, the entropy can also be calculated secondary to these values [296]. This can help describe the interaction by showing the individual contributions of enthalpy and entropy to the binding event. First, the Gibbs free energy (ΔG) of binding is calculated using the equation:

$$\Delta G = -RT \ln K$$

Where R is the gas constant, T is the temperature in Kelvin and K is the equilibrium constant (K_a). The entropy, ΔS , can be derived using:

$$\Delta G = \Delta H - T\Delta S$$

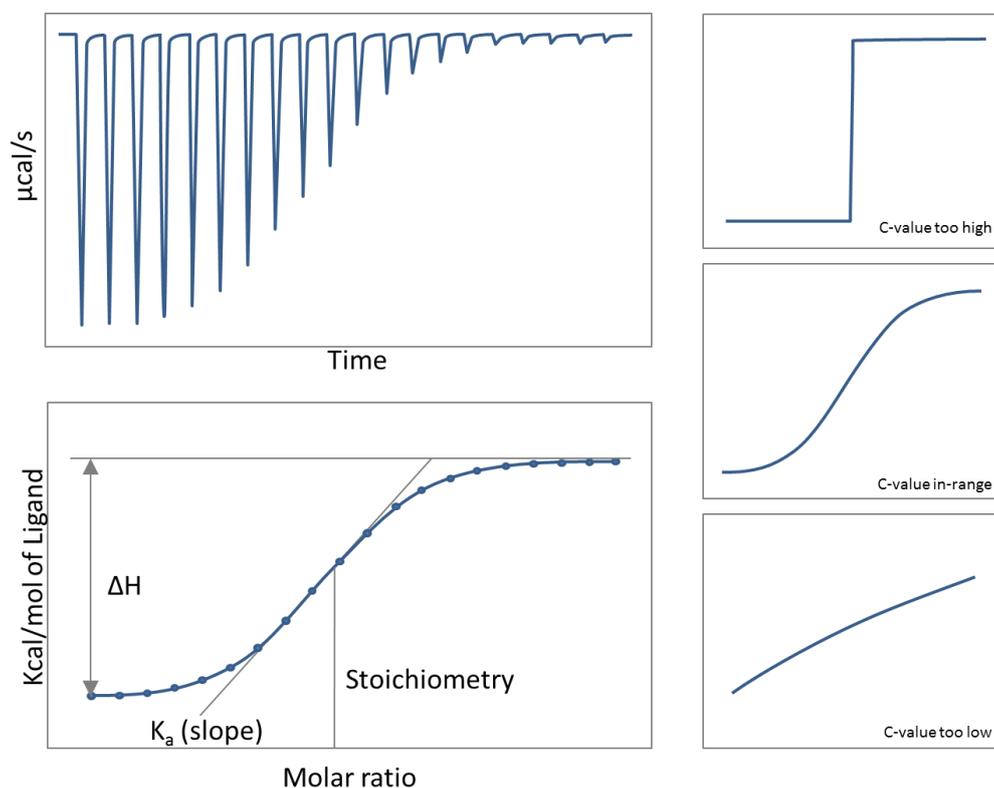


Figure 2.10. Example data from ITC. The left hand images show typical data that would be obtained from a suitable ITC experiment. Each peak from the top curve is integrated to obtain the energy per mole of ligand. The range of the plateaux, midpoint and slope of the sigmoidal curve allow calculation of binding parameters. The three right hand curves show how the c-value affects this curve.

ITC is a very effective technique for assessment of biopolymer interaction studies; often called the 'gold standard'. One of the main issues with ITC is the requirement for large quantities of pure protein, which is not always achievable. However, newer instrumentation, such as the MicroCal iTC200 calorimeter, uses far less

protein than its predecessors. This makes ITC a more accessible modality for protein interaction studies [297]. Also, ITC is not a high throughput assay; the instrumentation only performs one interaction study at a time, and can take up to a few hours to complete (in addition to the time for protein expression and purification). Due to the nature of the way the K_d is calculated, only interactions of affinities of approximately nM-mM in range can effectively be determined [298]. The reason for this is that the K_d is calculated from the slope of the midpoint curve in Figure 2.10. Sigmoidicity of the curve is essential, and can be determined by the c -value calculated by:

$$c = KMn$$

Where K is the K_a , M is the concentration of protein in the cell and n is the stoichiometry of the interaction. As shown in Figure 2.10, c -values should ideally fall within values of 10-500. Values higher and lower than this lose sigmoidicity of the curve [299]. To get good thermogram peaks and protein saturation, the recommended values for syringe concentrations are 10X that of the cell. So for a hypothetical low-affinity interaction with a K_d of 100 μ M, stoichiometry of 1 and using a low c -value of 10, the concentration required in the cell would need to be 1 mM and the syringe concentration would be 10 mM. For some instances of protein-ligand interactions this may be possible, but for most interactions, especially protein-protein, achieving these concentrations would be impossible.

Its advantages over other methodologies are that the thermodynamics can be evaluated in one experiment, and the assay is performed with free molecules in solution, rather than bound to a matrix or within the cell. This removes issues with

solid-phase immobilisation and labelling, and ensures that interactions are specific and binary between the interactors being tested.

2.6 Thermofluor

Thermofluor is a technique that is employed to measure temperature-dependent unfolding of a protein. This is a useful measure as it can be used to assess the effect of buffers on a protein [300], and investigate protein-ligand interactions [301]. Thermofluor presents an effective way to characterise proteins, and analysis of buffer composition on the melting temperature (T_m) is beneficial prior to crystallisation as increasing T_m correlates with success of crystallisation [300].

During a Thermofluor assay, the protein of interest is mixed with SYPRO Orange dye in an aqueous buffer. In its aqueous state, SYPRO Orange dye does not produce a high fluorescence at an excitation wavelength of 490 nm and an emission wavelength of 575 nm. However, during denaturation of the protein, the dye can bind to exposed hydrophobic residues from its core. Once bound, the dye is highly fluorescent, and can be detected by a rtPCR machine during a thermocycle.

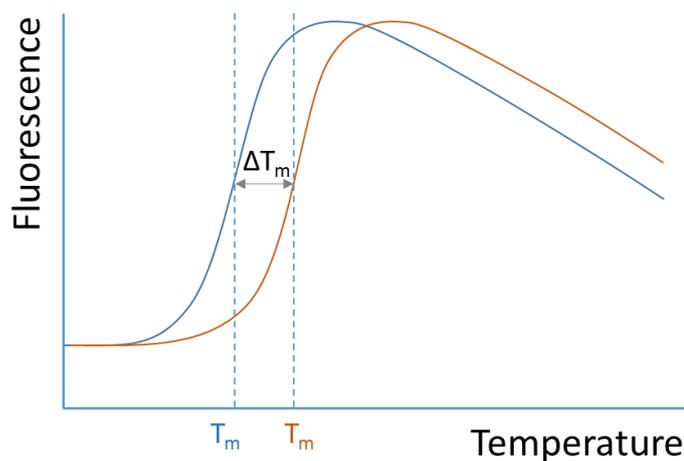


Figure 2.11. Thermofluor. Two thermograms are shown (red and blue curve). The melting temperature (T_m) is taken as the centre of the sigmoidal region of the trace.

Two thermograms are shown in Figure 2.11, and both proteins show a different melting temperature. The curves display an increase in fluorescence during which the protein is unfolding and then a decrease where it is thought that the dye and protein complex is precipitating. The melting temperature is the centre point of the sigmoidal region of the curve. In Figure 2.11, the red curve has a higher T_m than the blue curve. This shift can be used to analyse thermostabilising buffers (buffers where a positive shift is observed for the ΔT_m), or to look for the more thermostable of two proteins. It can also show whether a ligand is binding to the protein, as this can often lead to a shift in its T_m [300, 302].

3 Materials and methods

3.1 Cloning, mutagenesis and splicing by overlap extension PCR

All constructs made in this thesis were produced using standard molecular cloning, site directed mutagenesis and splicing by overlap extension PCR.

3.1.1 PCR

A 2X reaction mastermix containing 0.5 μ l template (approx. 25-250 ng of DNA), 22.5 μ l 5X HF buffer, 1.8 μ l dNTPs (10 mM each nucleotide), 56.25 μ l molecular grade H₂O. Each reaction contained 36.1 μ l mastermix, 2 μ l 20 μ M forward primer, 2 μ l 20 μ M reverse primer, 2 μ l 50% DMSO, 7.4 μ l molecular grade H₂O and 0.5 μ l PHUSION polymerase (Thermo Fisher Scientific).

Table 3.1. PCR protocol.

Step	Temp (°C)	Time	
Denaturation	95	5 minutes	
	Denaturation	95	30 seconds
24 cycles	Annealing	55-65	30 seconds
	Elongation	72	1 minute/kb
Extension	72	10 minutes	
Storage	4	∞	

The PCR protocol used is shown in Table 3.1. The gradient annealing temperature was used so that each PCR reaction was performed at 55 °C and 65 °C. Usually the PCR reaction produced the desired product at both annealing temperatures, but the 65 °C product was most often the purest.

3 µl of the PCR sample was subject to 1% agarose gel electrophoresis (1 g of agarose in 100 ml of TE buffer). 5 µl / 100 ml gel of NANCY-520 (Sigma-Aldrich) or 10 µl / 100 ml gel of SYBR Green was added to the gel. 3 µl PCR sample, 2 µl molecular grade H₂O and 1 µl 6X loading dye (30% glycerol, 0.25% bromophenol blue) was loaded into the wells of the gel. The gel was run at 100 V and imaged using a Dark Reader Transilluminator (Clarechemical).

If the product was pure (single band seen on agarose gel imaging), the PCR sample was subject to clean-up using a GenElute PCR Clean-Up Kit (Sigma Aldrich). If the sample was impure then it was cleaned-up and purified using a GenElute Gel Extraction Kit (Sigma Aldrich).

3.1.2 Restriction digestion

2 µl of the appropriate buffer for the two enzymes was added to 16 µl of DNA. 1 µl of each restriction enzyme (all from NEB) was added to the reaction and incubated for 4 h at 37 °C or overnight at room temperature. Up to 1 µg of DNA was used in each DNA restriction digestion reaction. If enzymes were not compatible, reactions with a single enzyme were performed sequentially, with a PCR clean-up in between.

The vector for gene insertion was also digested in the same way as the PCR product, and additionally dephosphorylated. 2 µl of Antarctic Phosphatase Buffer and 1 µl of Antarctic Phosphatase was added to the digestion reaction and incubated for 1-2 h at 37 °C. The PCR product and vector were subsequently cleaned-up using the GenElute PCR Clean-Up Kit.

3.1.3 Ligation

Ligation was performed using T4 DNA Ligase (NEB). Reactions were performed at 3:1 and 6:1 ratios. To calculate the appropriate volumes of DNA required to produce these ratios the following equation was used:

$$\frac{\left(\frac{\text{Size of insert} \times \text{Conc of vector}}{\text{Size of vector}}\right)}{\text{Conc of insert}} = \text{Volume of insert required for 1:1 ratio}$$

The volumes were then multiplied to correct the ratio and maximise the amount of DNA in the ligation reaction. 1 µl of T4 DNA Ligase Buffer was added to 8 µl of vector/insert DNA mix. 1 µl of T4 DNA ligase was then added and the reaction was incubated at 16 °C overnight. 2 µl of each reaction was then transformed into competent Nova Blue cells.

3.1.4 Transformation

Cells for transformation were prepared using the Inoue method for Ultra-competent cells [303]. The transformation was performed by adding up to 2.5 μl DNA into 50 μl of competent cells. The DNA/cell mixture was incubated on ice for 10 mins. The samples were heat shocked at 42 °C for 1 min and re-incubated on ice for 5 mins. 500 μl of LB was added to the cells and then the transformation mixture was incubated at 37 °C in a shaking incubator for 1 h. Up to 250 μl of this mixture was spread on an LB/agar plate containing the required antibiotic(s) and incubated at 37 °C overnight. Working concentrations of antibiotics are: Chloramphenicol – 35 $\mu\text{g/ml}$; Ampicillin – 100 $\mu\text{g/ml}$; Kanamycin – 50 $\mu\text{g/ml}$.

3.1.5 Cracking and analytical digest

The first pre-sequencing technique used to identify colonies with an insert was cracking. Cultures were made by picking single colonies off the transformation plate and inoculating 5 ml LB with required antibiotics. The culture was incubated overnight in a shaking incubator at 180 RPM and 37 °C. These cultures were initially checked using cracking. 20 μl of culture was added to 5 μl 5X cracking buffer (25 g sucrose, 5 ml 5 M NaOH, 2.5 ml 10% SDS, in 50 ml H₂O, bromophenol blue). The sample was run on agarose gel electrophoresis. Plasmids with the insert should appear at a higher molecular weight than those of a control empty plasmid. The DNA from cultures that appeared to contain an insert was purified using a GenElute Plasmid Miniprep Kit (Sigma-Aldrich).

Following cracking, Miniprep purified plasmids were digested to visualise an insert at the desired size. This analytical digestion was performed as the previous restriction digestion, but half the volumes of reagents and DNA were used, and digestion was performed for approximately 2 hours. The digestion was viewed following agarose gel electrophoresis and those plasmids with visible, digested inserts would be sequenced by the Sanger sequencing method (Source Bioscience).

3.1.6 Site-directed mutagenesis

Site directed mutagenesis was used to mutate the DNA sequence of a plasmid to alter the amino acid sequence of the final protein product. This was used routinely to introduce stop codons in constructs for protein expression, and also to alter amino acids in yeast two hybrid experiments. For each reaction, complementary primer pairs were produced where the mutated sequence was central to 15-30 template-complementary, flanking base pairs. The protocol used was taken from the Quickchange II manual. The reaction contained 10-50 ng of DNA, 5 μ l of 10X Pfu Ultra II reaction buffer, 2 μ l 20 μ M forward primer, 2 μ l 20 μ M reverse primer, 1 μ l dNTP mix, 1 μ l Pfu Ultra II polymerase (Agilent Technologies) and made up to 50 μ l with molecular grade H₂O. This mixture was cycled using the following protocol: 95 °C 0:30 ; [95 °C 0:30 ; 55 °C 1:00 ; 68 °C 1:00/kb]₁₆ ; 68 °C 3:00 ; 4 °C ∞ . 1 μ l of DpnI enzyme was added to the sample and incubated at 37 °C for 1 h. 2 μ l of this mixture was transformed into Nova Blue cells. Colonies were picked, overnight cultures were grown using suitable antibiotics, the DNA was purified using Genelute Miniprep kits and the DNA was sequenced to confirm mutation.

3.1.7 Cloning for catalytic domain-containing constructs.

In order to clone a gene and remove non-terminal sequences, splicing by overlap extension was used. It was used in the production of catalytic domain containing proteins that lack the 182-residue disordered loop. Initially, full length USP20 lacking the catalytic insert was cloned. To do this four primers were used, shown in Table 3.2. Splicing by overlap extension PCR was first performed in the same way as normal PCR. Two reactions were produced using the previous protocol, using forward primer 1, and reverse primer 1 for one reaction and forward primer 2 and reverse primer 2 for another reaction. In both cases the template plasmid was Flag-HA-USP20 (Addgene #22573). The primer 1 pair produce a PCR product that contains amino acid residues 1 to 250 of USP20. Primer pair 2 produces a product containing residues 432 to 914, with an additional 25 bp 5' extension that is complementary to the 3' of the first PCR product. This overlapping region, although complementary to each other, was made non-complementary to the template sequence to avoid binding of forward primer 2 upstream of its target sequence. Following PCR, the products were cleaned up using a PCR clean-up kit (Sigma Aldrich).

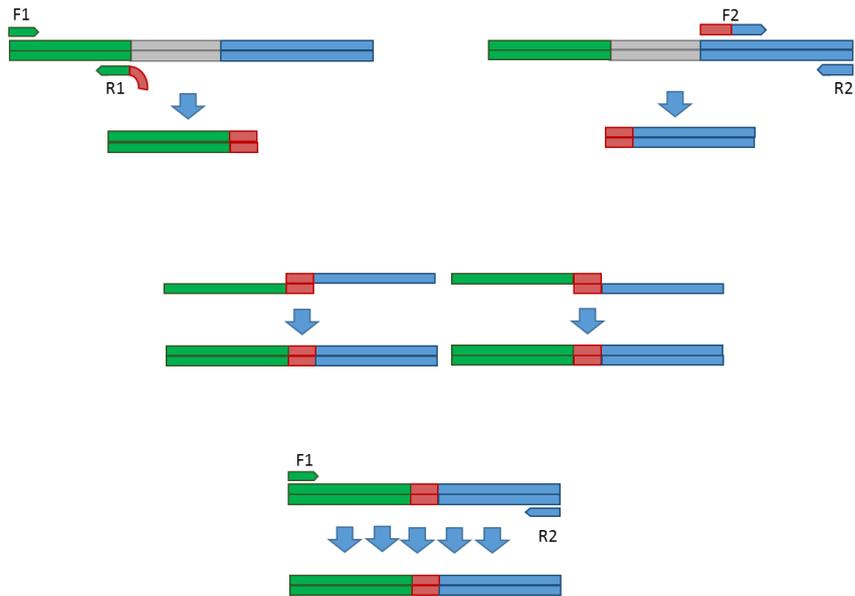


Figure 3.1 Principle of splicing by overlap PCR. The diagram shows the three stages of the procedure. First, the two (or more) fragments that are to be spliced together are amplified by PCR. Second, the two fragments are thermocycled, where they prime each other. Third, the 5' and 3' primers are added to and the mix is thermocycled with a high annealing temperature for specificity of the final, full length, spliced fragment.

Table 3.2. Splicing by overlap PCR primers

Primer	Sequence (5'-3')
Forward primer 1	GGAATTCATATGGGGGACTCCAGGGACC
Reverse primer 1	TGGTTCITTTAATTCTTCATGTAAGTGGTCCATCAGGCAGCGAAGG
Forward primer 2	GTTACATGAAGAATTAAGAACCAAGCCGGAGCGGAAGGAGCAGCG
Reverse primer 2	GCTCTAGATTACACGGCCCGTCTCGGCTTC

The second stage of splicing was overlap PCR. The concentration of the pure PCR products was measured by Nanoprop 1000 (Thermo Fisher Scientific) and mixed to equimolar concentrations in 25 μ l. The remaining constituents of the PCR reaction were added to this template: 5 μ l of 10X Pfu Ultra II reaction buffer, 1 μ l dNTP mix, 18 μ l molecular grade H₂O and 1 μ l Pfu Ultra II polymerase. This mixture was cycled using the protocol in Table 3.3.

Table 3.3. Overlap PCR.

Step		Temp (°C)	Time
Denaturation		95	2 minutes
24 cycles	Denaturation	95	20 seconds
	Annealing	60	20 seconds
	Elongation	72	30 seconds/kb
Extension		72	3 minutes
Storage		4	∞

The last stage of thermocycling was the purification PCR. 2 μ l of 20 μ M Forward primer 1 and 2 μ l of 20 μ M reverse primer 2 were then added to the reaction and further cycled according to the protocol in Table 3.3, but with an annealing temperature of 72 °C. The PCR product was separated by agarose gel electrophoresis and then the product corresponding to the correct molecular weight was gel purified using Genelute Gel extraction kit (Sigma Aldrich). To complete the cloning, this final product was treated as a normal PCR product and followed the previous protocols for molecular cloning.

The USP20 Δ Znf-UBP construct was made by normal cloning methods using primers for the 5' end of the catalytic domain sequence and the 3' end of the whole USP20 sequence. USP20 Δ DUSPs and USP20Catalytic only constructs were made by performing site-directed mutagenesis to insert a stop codon after the catalytic domains. All these constructs used the Flag-HA-USP20 plasmid for template. The corresponding constructs lacking the linker were all made as described, but using the USP20FL Δ insert plasmid as a template.

3.1.8 Summary of primers used

Tables 3.4, 3.5 and 3.6 show the details about cloning and site-directed mutagenesis.

Table 3.4. Cloning of USP20 *E. coli* expression constructs

Name	Protein and residues	Insertion Plasmid	Method	Template	Forward primer	Reverse primer
USP20 FL	USP20 1-914	pCOLD1	Molecular cloning	Flag-HA-USP20 (Addgene #22573)	GGAATTCATATGGGGAC TCCAGGGACC	GCTCTAGATTACACGGCCCGCTC TCGGCTTC
TF-USP20 FL	USP20 1-914	pCOLD TF	Molecular cloning	Flag-HA-USP20 (Addgene #22573)	GGAATTCATATGGGGAC TCCAGGGACC	GCTCTAGATTACACGGCCCGCTC TCGGCTTC
USP20 Znf-UBP-Cat	USP20 1-685	-	Site-directed Mutagenesis	USP20 FL	CGTACTCTTCTACAGGAAGA GCTGAGAGGAGGCCATGCG	CGCATGGCCCTCTCAGCTCTTCC TGTAAGAGTACG
USP20 Cat	USP20 147-685	-	Site-directed Mutagenesis	USP20 Cat-DUSPs	CGTACTCTTCTACAGGAAGA GCTGAGAGGAGGCCATGCG	CGCATGGCCCTCTCAGCTCTTCC TGTAAGAGTACG
USP20 Cat-DUSPs	USP20 147-914	pCOLD1	Molecular cloning	Flag-HA-USP20 (Addgene #22573)	CTCCATATGAAGAACCTCGG GAACTCC	GCTCTAGATTACACGGCCCGCTC TCGGCTTC
USP20 FLΔinsert	USP20 1-914	pCOLD1	Molecular cloning with splicing by overlap PCR	Flag-HA-USP20 (Addgene #22573)	1. GGAATTCATATGGGGAC TCCAGGGACC 2. GTTACATGAAGAATTA AAAGAACCGGAGCGGA AGGAGCAGCG	1. TGGTTCTTTAATCTTCATGTAAC TGGTCCATCAGGCAGGAAGG 2. GCTCTAGATTACACGGCCCGCTC TCGGCTTC
USP20 Znf-UBP-CatΔinsert	USP20 1-685	-	Site-directed Mutagenesis	USP20 FLΔinsert	CGTACTCTTCTACAGGAAGA GCTGAGAGGAGGCCATGCG	CGCATGGCCCTCTCAGCTCTTCC TGTAAGAGTACG
USP20 CatΔinsert	USP20 147-685	-	Site-directed Mutagenesis	USP20 Cat-DUSPsΔinsert	CGTACTCTTCTACAGGAAGA GCTGAGAGGAGGCCATGCG	CGCATGGCCCTCTCAGCTCTTCC TGTAAGAGTACG
USP20 Cat-DUSPsΔinsert	USP20 147-914	pCOLD1	Molecular cloning	USP20 FLΔinsert	CTCCATATGAAGAACCTCGG GAACTCC	GAGGGATCCTTAGTGGTGGTGGT GGTGGTGGCTGCCAGCACTCAAT ACC
USP20 Znf-UBP	USP20 1-101	pET21d	Molecular cloning	Flag-HA-USP20 (Addgene #22573)	CATGCCATGGGGACTCCA GGGACCTTTGC	GGAATTCCTAGTGGTGGTGGTGG TGGTGCAGCAGAGGGGCTGCCAG CC
USP20 Znf-UBP	USP20 1-92	pPROEx-Htb	Molecular cloning	Flag-HA-USP20 (Addgene #22573)	AAGCCATGGGGACTCCAG GGACC	ATTATGAATCTTACAGGAATACC TCCTTCTCACAGG
USP20 Znf-UBP	USP20 1-108	pPROEx-Htb	Molecular cloning	Flag-HA-USP20 (Addgene #22573)	AAGCCATGGGGACTCCAG GGACC	ATTATGAATCTTACAGGAATACC GAAGAGGAGCC
USP20 Znf-UBP	USP20 6-108	pPROEx-Htb	Molecular cloning	Flag-HA-USP20 (Addgene #22573)	AAGCCATGGACCTTTGCCCT CACCTTGACTCC	ATTATGAATCTTACAGGAATACC GAAGAGGAGCC
USP20 Znf-UBP	USP20 6-92	pPROEx-Htb	Molecular cloning	Flag-HA-USP20 (Addgene #22573)	AAGCCATGGACCTTTGCCCT CACCTTGACTCC	ATTATGAATCTTACAGGAATACC TCCTTCTCACAGG
USP20 DUSPs	USP20 684-914	pET26b	Molecular cloning	Flag-HA-USP20 (Addgene #22573)	GGAATTCATATGAGCGAG GAGGCCATGCGGGAG	CCGCTCGAGCAGGCCCGCTCTC GGCTTC
USP20 DUSPs	USP20 684-894	pET26b	Molecular cloning	Flag-HA-USP20 (Addgene #22573)	GGAATTCATATGAGCGAG GAGGCCATGCGGGAG	CCGCTCGAGCTGCGCCACTCTG GCGGATG
1BKR	SPTBN1 174-278	-	Site-directed Mutagenesis	pSIMON1	CCTATTATCATTATTTAGCT AGGCCGAGCTAGCGGTAC C	GGTACCGCTAGTGGCCCTAGCT AAAATAATGATAATAGG
1BKR-DUSPs	USP20 686-894	pSIMON1	Molecular cloning	Flag-HA-USP20 (Addgene #22573)	CTTGCGCCGCAAGCGAGG AGGCCATGCGGGAG	CCGCTCGAGTTACTGCGCCACT CTGGCGGATG
2GKG	frzS 3-123	-	Site-directed Mutagenesis	pSIMON2	GCTGATTGGCTTCCGTAGG CCGCAGCTAGCGGTACC	GGTACCGCTAGTGGCCCTACG GAAAGCCAATCAGC
2GKG-DUSPs	USP20 686-894	pSIMON2-his	Molecular cloning	Flag-HA-USP20 (Addgene #22573)	CTTGCGCCGCAAGCGAGG AGGCCATGCGGGAG	CCGCTCGAGTTACTGCGCCACT CTGGCGGATG
2GKG-his	frzS 3-123	-	Site-directed Mutagenesis	pSIMON2-His	GCTGATTGGCTTCCGTAGG CCGCAGCTAGCGGTACC	GGTACCGCTAGTGGCCCTACG GAAAGCCAATCAGC
MBP-DUSPs	USP20 686-894	pMALX(E)	Molecular cloning	Flag-HA-USP20 (Addgene #22573)	CTTGCGCCGCAAGCGAGG AGGCCATGCGGGAG	CTTAGATTAGTGGTGGTGGTGGT GGTGTGCGCCACTCTGGCGGA TG
MBP-PB	PLK1 367-603	pMALX(E)	Molecular cloning	GST-PLK1 (DNASU HsCD00630433)	AAGCGCCGCGGAGAACCT GTACTTCCAAGTTCTGCTG GTGAGGTGGTCTGACTGCC	TCAGGAATCTTAGGAGGCCTTGA GACGG

Details about proteins that were expressed in *E. coli* are given. For modifications, single letter amino acid code is used except for where whole or partial proteins are present (shown in brackets). All cloned proteins are the human orthologues, except for frzS, which is the *Myxococcus xanthus* orthologue.

Table 3.5. Cloning of solubility tag vectors

Name	Protein and residues	Method	Template	Insertion Plasmid	Forward primer	Reverse primer
pSIMON1	SPTBN1 174-278	Molecular cloning	Gene synthesis 1BKR + MCS in pUC57	pRSF-13	GCCCCATGGCGCATCATC ATCATC	GCGACTCGAGAAGCTTGAG CTCG
pSIMON2	frzS 3-123	Molecular cloning	Gene synthesis 2GKG + MCS in pUC57	pRSF-13	GCCCCATGGCGCATCATC ATCATC	GCGACTCGAGAAGCTTGAG CTCG
pSIMON2-His	frzS 3-123	Molecular cloning	Gene synthesis 2GKG-His + MCS in pUC57	pRSF-13	GCCCCATGGCGAAAAAA ATTCTG	GCGACTCGAGAAGCTTGAG CTCG

Details about proteins that were cloned to produce solubility tag vectors are given. For modifications, single letter amino acid code is used except for where whole or partial proteins are present (shown in brackets). SPTBN1 is the human orthologue; frzS is the *Myxococcus xanthus* orthologue.

Table 3.6. Cloning of yeast two-hybrid constructs

Name	Protein and residues	Method	Template	Insertion Plasmid	Forward primer	Reverse primer
USP20 Znf-UBP	USP20 1-101	Molecular cloning	Flag-HA-USP20 (Addgene #22573)	pBTM116mod	TCAGGAATTCATGGGGAC TCCAGGGACC	GGAGGATCCAGCAGAGGGGCT GCCAGCC
USP20 Cat	USP20 147-685	Molecular cloning	Flag-HA-USP20 (Addgene #22573)	pBTM116mod	TCAGGAATTCATGAAGAAC CTCGGGAATCC	GGAGGATCCGCTTCTCTGTAGA AGAGTACG
USP20 DUSPs	USP20 686-894	Molecular cloning	Flag-HA-USP20 (Addgene #22573)	pBTM116mod	TCAGGAATTCAGCGAGGAG GCCATGCGGGAG	GGAGGATCCCTGCGCCACTCT GGCGG
USP20 Znf-UBP	USP20 1-108	Molecular cloning	Flag-HA-USP20 (Addgene #22573)	pASV3mod	ATATCTCGAGATGGGGAC TCCAGGGACC	ACCCGGTTACTTGAAGAGGA GCCCAGCAGAGG
USP20 Cat	USP20 147-685	Molecular cloning	Flag-HA-USP20 (Addgene #22573)	pASV3mod	AATGGATCCATGAAGAACC TCGGGAATCC	AAAGATCTTAGCTTCTCTGTA GAAGATACG
USP20 Cat inactive mutant	USP20 147-685	Site-directed Mutagenesis	USP20 Cat	-	GCACGGCAGGCAGTGGGAA CTACATCGCCTACTGCCAGA ACG	CGTTCTGGCAGTAGGCGATGTA GTTCCCACTGCCTGCCGTGC
USP20 DUSPs	USP20 686-894	Molecular cloning	Flag-HA-USP20 (Addgene #22573)	pASV3mod	ATATCTCGAGAGCGAGGAG GCCATGCGGGAG	ACCCGGTTACTGCGCCACTCT TGGCGG
USP20 single DUSP	USP20 686-791	Site-directed Mutagenesis	USP20 DUSPs	-	GGAGATCGAGGCACTGTAA AAGCGCAGGAGGATCG	CGATCTCTCGCTTTTACAGT GCCTCATCTCC
USP20 DUSP D841R mutant	USP20 686-894	Site-directed Mutagenesis	USP20 DUSPs	-	CCCGGGCCATTGCAACA GCAGATTGCA	CCTGTGCAATCTGTGTTGCGA ATGGCCCGGGG
Beta arrestin 1	ARBB1 1-418	Molecular cloning	IMAGE:3604829	pBTM116mod	AATACTCGAGATGGGCGAC AAAGGGACCC	AACTGCAGTTATCTGTTGTAG CTGTGGAGAGC
Beta arrestin 1	ARBB1 1-418	Molecular cloning	IMAGE:3604829	pASV3mod	TCTGCTCGAGATGGGCGAC AAAGGGACCC	TCCGCGTCTGTTGTAGCTGT GGAGAGC
TRAF6	TRAF6 1-522	Molecular cloning	GST-TRAF6 (DNASU HsCD00077278)	pBTM116mod	TCAGGAATTCATGAGTCTGC TAAACTGTG	TCCGCGTTACAATACCCCTGCA TCAGTACTCG
RAD17	RAD17 1-681	Molecular cloning	RAD17 (DNASU HsCD0005955)	pBTM116mod	TCAGGAATTCATGAATCAG GTAACAGACTGG	TCCGCGTTACAATGTCCCATCA CTCTCG
PLK1 FL	PLK1 1-603	Molecular cloning	GST-PLK1 (DNASU HsCD00630433)	pBTM116mod	TCAGGAATTCATGAGTGCT GCAGTGACTGC	TCCGCGTTAGGAGGCCCTTGAG ACGG
PLK1 Cat	PLK1 1-408	Site-directed Mutagenesis	PLK1 FL	-	CCTGCCTGCATCCCCTAATT CTGGGTCAGCAAGTGG	CCACTGTGCTGACCAGAATTAGG GGATGCAGGCAGG
PLK1 Poloboxes	PLK1 367-603	Molecular cloning	GST-PLK1 (DNASU HsCD00630433)	pBTM116mod	TCAGGAATTCGGTGAGGTG GTCGACTGCC	TCCGCGTTAGGAGGCCCTTGAG ACGG
PLK1 Cat-PB1	PLK1 1-508	Site-directed Mutagenesis	PLK1 FL	-	GGTGATGAGCTCGCCCGGT AGCCCTACTACGGACC	GGTCCGTAGGTAGGCTACCGG GCGAGCTCATCACC
PLK1 Poloboxes H538A	PLK1 367-603	Site-directed Mutagenesis	PLK1 Poloboxes	-	CAACTCTTCCAGGATGCCA CCAAGCTCATCTGTG	CACAAGATGAGCTGGTGGCAT CCTGGAAGAAGTTG
PLK1 Poloboxes K540M	PLK1 367-603	Site-directed Mutagenesis	PLK1 Poloboxes	-	CTCCAGGATCACCCATGC TCATCTGTGCC	GGGCACAAGATGAGCATGGTGT GATCCTGGAAG

Details about proteins that were used in yeast two hybrid assays are given. For modifications, single letter amino acid code is used except for where whole or partial proteins are present (shown in brackets). All cloned proteins are human orthologues.

3.2 Protein expression

The protein expression methodologies varied depending on the construct, plasmid or cells. Information about specific IPTG concentrations, media, and induction times are given in Table 3.7, however they all followed the following general protocol. 100 ml of LB broth containing suitable antibiotic(s) in 250 ml, non-baffled Erlenmeyer flasks were inoculated with cells containing the plasmid of interest. These were grown overnight at 37 °C in a shaking incubator at 180 RPM. 1-2% of these starter cultures were used to inoculate 0.5-1L of media and appropriate antibiotic(s) in 2 L, non-baffled Erlenmeyer flasks. These cultures were grown to optical densities (OD) of 0.5-0.6 (non-pCOLD vectors) and 0.5-1 (pCOLD vectors), where they were induced with IPTG. The temperature was set to 10-37 °C and the cultures incubated for 4–68 h. Cells were harvested by centrifugation at 4000 RPM for 10 mins. The supernatant was removed and the pellets were stored in 50 ml tubes at -20 °C until needed.

Table 3.7. Expression conditions

Construct	Cell type	Media	Induction temp	IPTG concentration	Expression time
USP20 FL	BL21 codon plus	2YT	10	0.5	68
				0.2	68
	Arctic express	2YT	10	0.2	68
TF USP20	BL21 codon plus	2YT	10	0.2	68
Other catalytic constructs	BL21 codon plus	2YT	10	0.2	68
Znf-UBP (all)	BL21 codon plus	LB	25	0.5	20
DUSP domains 686-894	BL21 codon plus	LB	25	0.5	20
			37	0.5	4
DUSP domains 686-914	BL21 codon plus	LB	25	0.5	20
1BKR	BL21 codon plus	LB	18	0.5	20
1BKR-DUSPs	BL21 codon plus	LB	18	0.5	20
2GKG	BL21 codon plus	LB	18	0.5	20
2GKG-DUSPs	BL21 codon plus	LB	18	0.5	20
MBP-DUSPs	BL21 codon plus	LB	18	0.2	20
				0.5	20
MBP-poloboxes	BL21 codon plus	LB	18	0.5	20
Ubiquitin	BL21 codon plus	LB	37	0.5	4

3.3 Protein purification

3.3.1 Lysate preparation

The general protocol for lysate preparation was as follows. Cells were thawed and resuspended in 10 ml of buffer A per gram of pellet (if protease inhibitors were added they were dissolved into this resuspension buffer). Cells were sonicated on ice with an amplitude of 12 microns. The samples were sonicated for 20 seconds on, 20 seconds off for a total of 1 min 40 seconds of sonication time. Lysates were clarified by centrifugation at 14000 RPM at 4 °C for 45 mins. Supernatants were filtered with a 0.4 µm filter.

For the purification of ubiquitin, a modified lysate preparation was used. The cells are lysed, clarified and filtered as previously described. The pH of the lysate is dropped to pH 4.5-5 and then increased back to 5.1. The solution is then centrifuged at 14000 RPM again for 30 mins to pellet the insoluble protein.

3.3.2 Nickel affinity chromatography

A Hi-Trap chelating HP 5ml column was used as the initial purification step for His-tagged proteins. The initial nickel column buffer A for all constructs was 300 mM NaCl, 50 mM Tris pH 7.5 and 20 mM imidazole ± 1% glycerol. Buffer B was the same but with 500 mM imidazole. Modifications to the buffers to improve solubility were performed for many purifications. The exact modifications are discussed in the results section. Prior to loading the protein onto the column, it was stripped using 25 column volumes

(CV) of 100 mM EDTA and re-charged using 25 CV of 100 mM nickel sulphate. The cleared lysate was loaded onto the column after being equilibrated in 25 CV of buffer A using a peristaltic pump. The column was washed with 30 ml buffer A to clear the lysate and remove any non-specifically bound proteins. Typically, a 20 CV, linear concentration gradient of buffer B was used to elute the protein. This was performed using either an AKTA Purifier or an AKTA Prime. 2.5 ml fractions were collected during the elution and analysed by SDS-PAGE prior to pooling. The protein was concentrated using a centrifugal ultra-filtration unit of required molecular weight membrane.

3.3.3 Anion exchange

Samples for ion exchange were concentrated until 5 ml in volume. This was then diluted in 45 ml buffer A (20 mM Tris). The pH of the buffer was at least two units above the predicted pI of the protein as predicted by the protparam server. A Resource Q (GE Healthcare) 6 ml anion exchange column was used for additional purification due to its high resolution. Prior to loading of the sample, the column was pre-equilibrated with buffer A. The protein was loaded onto the column using a 50 ml superloop on an AKTA Purifier. Unbound protein was washed out with 30 ml buffer A. Protein was eluted using a 20 CV linear concentration gradient with buffer B (20 mM Tris, 1 M NaCl). 1 ml fractions were collected during the elution and analysed by SDS-PAGE.

3.3.4 Cation exchange

Cation exchange was only used for ubiquitin. The lysate was loaded onto a pre-equilibrated HiTrap SP 5 ml column at pH 5.1 using a peristaltic pump. The column was washed with 30 ml buffer A (20 mM ammonium acetate pH 5.1). The column was either connected to an AKTA Prime or an AKTA Purifier and the protein was eluted with a 20 CV gradient of buffer A to Buffer B (0.5 M NaCl, 20 mM ammonium acetate pH 5.1). 2.5 ml fractions were obtained and protein content was assessed by SDS-PAGE.

3.3.5 Size exclusion chromatography

Size exclusion chromatography was performed using Superdex gel filtration columns. The size of the protein determined the column used (Table 3.8). Columns were pre-equilibrated with GF buffer. The first gel filtration buffer used for most constructs was 150 mM NaCl, 20 mM Tris pH 7.5, \pm glycerol. Modifications to the buffers to improve solubility were performed for many purifications. The exact modifications are discussed in the results section. Samples were concentrated until 2.5 ml in volume - unless prohibited by the solubility of the protein. Samples were loaded onto a 5 ml loop and injected onto the column using an AKTA Purifier or AKTA Prime. Elution of the protein was performed using an isocratic elution using GF buffer. 2.5 ml fractions were collected from 40 ml in order to ensure collection of protein from void (44-47 ml; depending on the column) to the end of elution.

Table 3.8. Gel filtration columns

Column	Proteins used
Hiload Superdex 16/60 75 pg	USP20 Znf-UBP (1-101, 1-108, 6-92, 6-108), USP20 DUSP domains (686-894, 686-914), Ubiquitin, MBP-Poloboxes (post-cleavage), 1BKR, 2GKG.
Hiload Superdex 16/60 200 pg	USP20 catalytic domain-containing constructs (FL, Znf-UBP-Cat, Cat, Cat-DUSPs, FL Δ insert, Znf-UBP-Cat Δ insert, Cat Δ insert, Cat-DUSPs Δ insert), MBP-DUSPs, 2GKG-DUSPs, 1BKR-DUSPs, MBP-Poloboxes (pre-cleavage)

3.3.6 Hydrophobic interaction chromatography

Samples for hydrophobic interaction chromatography (HIC) were adjusted to 500 mM NaCl. A 5 ml HiTrap Butyl FF column was used for HIC. Prior to loading of the sample, the column was pre-equilibrated with buffer A (500 mM NaCl, 20 mM Tris pH 7.5). The protein was loaded onto the column using a peristaltic pump. Unbound protein was washed out with 30 ml buffer A. Protein was eluted using a 20 CV linear concentration gradient with buffer B (20 mM Tris pH 7.5). 1 ml fractions were collected during the elution.

3.4 SDS-PAGE

Gels of 0.75 or 1.5 mm thickness and 10 or 15 wells were prepared. The thicker gels were used in some cases for SDS-PAGE followed by Western blot, when large sample volumes were required. The 0.75 mm thickness gels were used for both SDS-PAGE, and SDS-PAGE for western blot. Discontinuous SDS-PAGE was used, where the gel comprises

resolving and stacking components. The resolving gels are formed with different concentrations of acrylamide, depending on the protein to be resolved. In this thesis, gels of 10-20% were made. The lower percentage gels (10-12 %) were used to resolve the high molecular weight proteins (those above 60 kDa. 15% gels were most commonly used, and resolved proteins between 20-60 kDa. Higher percentage gels (18-20%) were used to resolve the smaller proteins less than 20 kDa. A stacking gel of 6% was used in all cases. The gels were produced with the recipe in Table 3.9.

Table 3.9. SDS-PAGE gel recipes

Reagent	Resolving gel volume (ml)	Stacking gel volume (ml)
H ₂ O	1-6.3	5.4
30 % acrylamide/bis-acrylamide	5.33-10.67	2
1.5 M Tris pH 8.8	4	-
0.5 M Tris pH 6.8	-	2.5
10% SDS	0.16	0.1
10% APS	0.16	0.1
TEMED	0.016	0.01

The range of values for H₂O and acrylamide/bis-acrylamide indicate the range from 10-20% gels. The total sum of these volumes will be 11.66 ml.

SDS-PAGE samples were prepared mixing 3 X sample buffer (reducing or non-reducing) with the protein sample at a 1:2 ratio. The samples were boiled at 95 °C for 5 minutes prior to loading in the gel wells. The gels were run at constant 200 V until the band front reached the bottom of the gel. This usually took between 45-60 minutes. Gels were

stained with 0.1% Coomassie Brilliant Blue R-250, 50% isopropanol, 10% glacial acetic acid and destained with 50% isopropanol, 10% glacial acetic.

3.5 Western blot

Western blots were performed to view the presence of catalytic domain-containing USP20 constructs in the *E. coli* supernatant after sonication. Here, small scale expressions of 100 ml of 2YT were made, harvested and sonicated in the same process as described above. 20 µg of protein was loaded into the wells of 12% SDS-PAGE and the gel was run as described above. The gel, components of the transfer apparatus and membrane were equilibrated in 1 X Western blot transfer buffer (WBTB). The Mini Trans-Blot Cell (Bio Rad) was filled with 1 X WBTB and a freezer pack was inserted. The protein was then transferred onto the nitrocellulose membrane at 150 mA for 2 hours at 4 °C in.

Western blots were also used to ensure that the yeast two hybrid constructs were being expressed. The samples for SDS-PAGE were produced using a modified NaOH extraction method as described by Zhang *et al.* [304]. 1.5 ml YPD was inoculated with one transformed yeast colony (or untransformed for negative control). As positive controls, the cells were transformed with either empty pBTM116mod or pASV3mod. The culture was incubated at 30 °C overnight and harvested by centrifugation at 15000 RPM for 15 seconds. The cells were resuspended in 2 M lithium acetate and incubated on ice for 5 minutes. The cells were pelleted, resuspended in 0.4 M NaOH and incubated on ice for 5 minutes. The cells were once again pelleted and resuspended in 100 µl 3 X SDS-PAGE sample buffer. The tubes were boiled for 5 minutes at 95 °C and centrifuged at 15000 RPM for 10 minutes. 20-40 µl of the supernatant was loaded into the 15% gel and run as

previously described. For these western blots, PVDF (Thermo Fisher Scientific) was used as the membrane, which requires activation in methanol for 5 minutes prior to equilibrating in 1 X Western blot transfer buffer. All apparatus was equilibrated in 1 X WBTB and protein was transferred using a Trans-Blot Turbo Transfer system (Bio Rad). Transfer was achieved with the standard semi-dry method; 25 V for 20 minutes.

For both the USP20 and yeast two hybrid Western blots, the following procedure was the same. Membranes were blocked with 3% Skimmed milk TBST. The membrane was inserted into a 50 ml tube and incubated at 4°C overnight on a rolling mixer. The following morning, the blocking milk was poured away and the primary antibody was added. For USP20, the primary antibody was mouse IgG anti-tetra-His antibody (Qiagen). The antibody was used at a 1:1000 dilution in 3% Skimmed milk TBST. A rabbit polyclonal IgG anti-LexA antibody (Invitrogen) was used for LexA fusion detection at a 1:1000 dilution in 3% Skimmed milk TBST. A rabbit polyclonal IgG anti-VP16 antibody (abcam) was used for VP16 fusion detection at 1:500 dilution in 3% Skimmed milk TBST. 5 ml of the primary antibody was incubated with the membranes for 1 hour at room temperature on a rocking mixer. Blots were washed with 15 ml TBST for 5 minutes three times. Horseradish peroxidase (HRP)-tagged anti-mouse IgG antibody was used to detect the anti-His antibody and a goat polyclonal IgG anti-rabbit antibody (abcam). In all cases, secondary antibodies were diluted 1:2000 in 3% Skimmed milk TBST. 5 ml of the secondary antibody was incubated with the membranes for 1 hour at room temperature on a rocking mixer. Membranes were washed three times in TBST as before. Enhanced chemiluminescence (ECL) solution (1 ml 1 M Tris pH 8.6, 22 µl 90 mM *p*-coumaric acid, 50 µl 250 mM luminol and 3 µl 30% H₂O₂) is added to the membrane and incubated for 90 seconds at room

temperature. The ECL solution is removed and the blot is imaged using a Biospectrum Imaging System. The CCD camera automatically calculates the required exposure time. The resulting images are manipulated with ImageJ and GIMP 2.8.

3.6 Analysing protein concentration

A Nanodrop 1000 (Thermo Scientific) was used to analyse the concentration of purified proteins. For samples from the nickel column, pooled fractions were analysed using a blank at the average % B concentration over the pooled peak. This was essential as imidazole has fluorescence at 280 nm so using either buffer A or buffer B would over- or underestimate the protein concentrations, respectively. Extinction coefficients were calculated by inputting the primary sequence into Protparam [305]. The value obtained from the Nanodrop 1000 was divided by the correction factor from Protparam to obtain the true concentration. Ubiquitin and 2GKG concentrations were difficult to assess as they contain no tryptophans. Therefore, the measurement and correction may produce an error, systematically either over- or underestimating the concentration at any stage.

3.7 Crystallisation

Crystallisation was performed using 96-well block screens. The vapour diffusion method was used in all crystallisation trials. Both sitting drop and hanging drop variants were used. The wells of the crystal trays were filled with 80 μ l of precipitant solution using either multi-channel pipettes or Matrix Hydra II (Thermo Scientific). 0.1-0.4 μ l of protein was dispensed into the wells of a 96-well MRC plate or onto a plastic adhesive sheet by

a mosquito or Hydra liquid handling robots. The same volume of precipitant solution was taken from the well and mixed with the protein solution.

Seeding of crystals was performed by crushing the crystals in drops PACT B5, C9, and D6. The drops were pipetted into a 1.5 ml Eppendorf tube and approximately 45 μl of the well solution from PACT B5 was added to the solution. A pre-chilled seed bead was added to the tube and the solution was vortexed for three minutes total (30 seconds vortexing and 30 seconds on ice). The crystal screens were made as above, but 25 nl of the seed stock solution was added to the drops prior to sealing the plate.

3.8 X-ray diffraction

Crystals were soaked in cryoprotectant consisting of 30% glycerol in addition to the condition of the mother liquor. They were then cryocooled in liquid nitrogen (100K) and kept frozen throughout the experiment. Diffraction data from the crystals of 2GKG were obtained at the IO2 beamline at Diamond Light Source. The data were collected at 100K at the beamline. The wavelength of the X-rays were 0.97 \AA and the detector distance was set to 267.02 mm from the crystal. For imaging 3 test shots were taken to visualise diffraction. These consisted of 0.5° oscillations at 45° apart. The datasets were then collected by a 180° rotation of the crystal with images consisting of 0.1° oscillations. This produced datasets with 1800 images. Indexing was attempted using iMosflm and XDS. XDS is generally thought to be better for fine slice data than iMosflm.

3.9 DUSP domains solubility assays

3.9.1 Native PAGE assay

One method to analyse the effect of buffer composition on the DUSP domains is through the use of native PAGE. This method shows the different species in the protein solution by electrophoresis of the natively charged proteins. If a specific buffer stabilises the protein, then it may be observed by a change in the profile of the gel. 8% native PAGE gels were produced with the recipe in table 3.10.

Table 3.10. Native PAGE gel recipes

Reagent	Native gel volume (ml)
30 % acrylamide/bis-acrylamide	2
0.3 M Tris pH 8.8	7.89
10% APS	0.1
TEMED	0.01

The protein from the DUSP domains nickel column was pooled and concentrated until 1 mg/ml. The protein was then diluted 4 X by 1.33 X test buffer. 4 X native PAGE sample buffer was mixed with the protein sample at 1:3 ratio. The test buffers are given in Table 3.11.

3.9.2 Cell lysis assay

A solubility screening protocol as described by Lindwall *et al.* [306] was also used to investigate buffer composition on the solubility of the DUSP domains. A 600 ml flask of LB was inoculated with BL21 cells containing the DUSP domains protein (USP20 686-894). Just prior to induction, 1 ml of cells was taken to act as a negative control. The cells were induced at OD₆₀₀ with 0.5 mM IPTG. Prior to harvesting, 1 ml of cells was taken to act as a positive control. The cells were harvested by centrifugation at 4600 RPM for 30 minutes. The pellet was resuspended in 30 ml wash buffer (10 mM Tris pH 8.5, 100 mM NaCl, 1 mM EDTA) and aliquoted into 1 ml samples. Aliquots and control samples were centrifuged at 13000 RPM for 10 minutes and the supernatant was discarded. The control samples were resuspended in 100 μ l denaturing buffer (15 mM Tris pH 6.8, 1% SDS, 2 M Urea, 1.25% β -mercaptoethanol, 2.5% glycerol). The remaining aliquots were resuspended in 1 ml of test buffer (see Table 3.11). The sample was lysed by addition of lysozyme and sonication for 30 seconds. The samples were incubated on ice for 10 minutes and clarified by centrifugation at 13000 RPM for 10 minutes at 4 °C. 15 μ l of lysate was loaded into 15% gels and run on SDS-PAGE. Gels were stained with Coomassie Brilliant Blue. Gels were scanned and densitometry was performed using ImageJ software.

3.9.3 Buffer compositions

The buffer compositions used in the native PAGE and cell lysis assays are given in

Table 3.11.

Table 3.11. Screening conditions

Name	Composition
Standard buffer	500 mM NaCl, 20 mM Tris pH 8.0, 20 mM imidazole, 5% glycerol, 1.5 mM β -mercaptoethanol
KCl	500 mM KCl, 50 mM Tris pH 7.5
Glucose	100 mM glucose, 500 mM NaCl, 50 mM Tris pH 7.5
Low glycerol	1% glycerol, 500 mM NaCl, 50 mM Tris pH 7.5
High glycerol	10% glycerol, 500 mM NaCl, 50 mM Tris pH 7.5
Urea	100 mM Urea, 500 mM NaCl, 50 mM Tris pH 7.5
L-arginine	10 mM L-arginine, 500 mM NaCl, 50 mM Tris pH 7.5
Citric acid	100 mM citric acid, 500 mM NaCl, 50 mM Tris pH 7.5
Ammonium sulphate	150 mM ammonium sulphate, 50 mM Tris pH 7.5
LiCl	10 mM LiCl, 500 mM NaCl, 50 mM Tris pH 7.5
Guanodinium hydrochloride	100 mM guanodinium hydrochloride, 500 mM NaCl, 50 mM Tris pH 7.5
MgCl ₂	10 mM MgCl ₂ , 500 mM NaCl, 50 mM Tris pH 7.5
Low salt	150 mM NaCl, 50 mM Tris pH 7.5
Ethylene glycol	3 % ethylene glycol, 500 mM NaCl, 50 mM Tris pH 7.5
MES pH 5.8	MES pH 5.8, 500 mM NaCl
Bis-Tris propane pH 6.5	Bis-Tris propane pH 6.5, 500 mM NaCl
HEPES pH 7	HEPES pH 7, 500 mM NaCl
HEPES pH 7.5	HEPES pH 7.5, 500 mM NaCl
HEPES pH 8.5	HEPES pH 8.5, 500 mM NaCl
Bicine pH 9.2	Bicine pH 9.2, 500 mM NaCl

3.10 Thermofluor

Thermofluor was used to assess the thermostability of the Znf-UBP domain and DUSP domains in different buffers. It was also used to assess the binding of the Znf-UBP domain to ubiquitin. The fluorescent dye SYPRO orange binds to exposed hydrophobic regions of proteins in this manner it is used to detect the melting temperature of proteins by constantly observing the fluorescence of the dye while the protein is subjected to a temperature gradient. When the protein melts (unfolding) the core hydrophobic residues are exposed, which leads to an increase in fluorescence. At the protein's melting temperature, a sigmoidal fluorescence curve over temperature is seen. The midpoint of this curve is taken as the melting temperature.

A 96-well buffer screen was made that consisted of an NaCl screen, pH/buffer screen and an additive screen in triplicate. The 96 well block and conditions are shown in Figure 3.2.

	1	2	3	4	5	6	7	8	9	10	11	12
A	NaCl 0 mM	NaCl 50 mM	NaCl 100 mM	NaCl 150 mM	NaCl 200 mM	NaCl 250 mM	NaCl 300 mM	NaCl 500 mM	NaCl 750 mM	Chaps 0.1%	Chaps 0.1%	Chaps 0.1%
B	NaCl 0 mM	NaCl 50 mM	NaCl 100 mM	NaCl 150 mM	NaCl 200 mM	NaCl 250 mM	NaCl 300 mM	NaCl 500 mM	NaCl 750 mM	Glucose 100 mM	Glucose 100 mM	Glucose 100 mM
C	NaCl 0 mM	NaCl 50 mM	NaCl 100 mM	NaCl 150 mM	NaCl 200 mM	NaCl 250 mM	NaCl 300 mM	NaCl 500 mM	NaCl 750 mM	EDTA 1 mM	EDTA 1 mM	EDTA 1 mM
	+ Tris pH 7.5 + NaCl + 0.02% sodium azide											
D	Citric acid pH 5.5	Citric acid pH 6	MES pH 6.5	MOPS pH 7	Tris pH 7.2	Tris pH 7.5	Tris pH 8	Bicine pH 8.5	Glycine pH 9	DTT 1 mM	DTT 1 mM	DTT 1 mM
E	Citric acid pH 5.5	Citric acid pH 6	MES pH 6.5	MOPS pH 7	Tris pH 7.2	Tris pH 7.5	Tris pH 8	Bicine pH 8.5	Glycine pH 9	L-arginine 10 mM	L-arginine 10 mM	L-arginine 10 mM
F	Citric acid pH 5.5	Citric acid pH 6	MES pH 6.5	MOPS pH 7	Tris pH 7.2	Tris pH 7.5	Tris pH 8	Bicine pH 8.5	Glycine pH 9	Urea 100 mM	Urea 100 mM	Urea 100 mM
	+ 150 mM NaCl + 50 mM buffer + 0.02% sodium azide											
G	150 NaCl + 50 Tris pH 7.5 + 0.02% Sodium azide			MgCl ₂ 10 mM	MgCl ₂ 10 mM	MgCl ₂ 10 mM	KCl 10 mM	KCl 10 mM	KCl 10 mM	ZnCl ₂ 1 mM	ZnCl ₂ 1 mM	ZnCl ₂ 1 mM
H	Empty			A mM SO ₄ 100 mM	A mM SO ₄ 100 mM	A mM SO ₄ 100 mM	Glycerol 5%	Glycerol 5%	Glycerol 5%	Ethylene glycol 5%	Ethylene glycol 5%	Ethylene glycol 5%
	+ 150 mM NaCl + Tris pH 7.5 + additive + 0.02% sodium azide (Note: A mM SO ₄ well does not contain NaCl)											

Figure 3.2. Thermofluor screening conditions. This 96 well block was made to use for Thermofluor buffer screening.

It includes an NaCl screen (green), pH/buffer screen (blue) and additive screen (yellow), all in triplicate.

For the buffer screening, 5 µg of protein was pipetted into clear-capped PCR tubes (5 µl of protein at 1 mg/ml). The 5000 X SYPRO orange dye was diluted to 10 X by a 1:500 dilution in H₂O. 40 µl of the test buffer was added to the PCR tube and mixed well. 5 µl of the 10 X SYPRO orange dye was added to the tube and the tube was again mixed. The total volume of the tube was 50 µl. The assay was performed using a Stratagene Mx3005p RT-PCR machine. The program performed 70 cycles with 1 °C increments from 25 to 95 °C. The amplification plots from the RT-PCR machine were converted into the correct format using excel and Prism Graphpad was used to calculate the Boltzmann sigmoid using the template files from <ftp://ftp.sgc.ox.ac.uk/pub/biophysics>. The V50 from this curve was taken as the melting temperature. Excel was then used to analyse the data. T-tests and linear regression were performed in Prism GraphPad.

3.11 Ubiquitin-AMC assays

Ubiquitin-AMC (Ub-AMC; Enzo Life Sciences) was used for investigation of deubiquitinating enzyme activity. A deubiquitinating enzyme and the substrate are incubated together and the release of ubiquitin from AMC by enzymatic cleavage can be observed using the increase in fluorescence of AMC at 460 nm when excited at 380 nm. Activity of the enzyme can be assessed by observing the fluorescence at multiple time points. If cleavage is occurring, and thus the enzyme has activity, an increase in fluorescence will be observed. All USP constructs that contained a catalytic domain were assessed by analysis with Ub-AMC assays. The constructs included USP20FL, TF-USP20FL, USP20 Δ DUSPs, USP20Catalytic, USP20 Δ Znf-UBP, USP20FL Δ insert, USP20 Δ DUSPs Δ insert, USP20Catalytic Δ insert, USP20 Δ Znf-UBP Δ insert.

30 μ l volume assays were performed in 384-well black plates (Nunc). 25 μ l of 1.2 X enzyme was equilibrated at room temperature in assay buffer (150 mM NaCl, 50 mM Tris-Cl pH 7.5, 1 mM DTT). 4.5 μ M Ub-AMC was also incubated at room temperature. 5 μ l of Ub-AMC was added to the enzyme to start the reaction. The final concentrations in the 30 μ l reaction were 100-500 μ M enzyme and 0.75 μ M substrate. The plates were instantly transferred to an EnVision 2104 multilabel plate reader with either a monochromator or filters. Using the plate reader with a monochromator, activity was assessed using an excitation wavelength of 355 nm, emission wavelength 426/428 nm, excitation light 100%, detector gain 750 and number of flashes 10. These values were optimisations by Hayley Gratton and gave the best signal/noise ratio on this machine. For the filter plate reader, Fura2 excitation and Umbelliferone emission filters were used

with wavelengths of 380 nm (bandwidth 19) and 460 (bandwidth 25), respectively. The excitation light was set to 10%, with a detector gain of 10 and 3 flashes. The plates were read at various time-points, which reflects the expected activity of the protein. The data were plotted on excel to produce a scatter graph of time and fluorescence.

3.12 Yeast two hybrid

3.12.1 Yeast media

Media for yeast consists of non-selective media (yeast extract, peptone and dextrose (YPD) \pm agar) and selective media (synthetic derived (SD) \pm agar). The YPD and SD media are for the growth of liquid cultures of yeast. The addition of agar provides a solid medium for the isolation of single colonies. The compositions are given below in Table 3.12. All media are autoclaved at 121°C for 20 mins.

Table 3.12. Yeast media

	400ml YPD (g)	400ml YPD + Agar (g)	400ml SD - L -T (g)	400ml SD - L -T + Agar (g)	400ml SD - T (g)	400ml SD - T + Agar (g)
Yeast Extract	4	4				
Peptone	8	8				
D-Glucose	8	8	8	8	8	8
Yeast Nitrogen Base Without AA (Y2062)			2.67	2.67	2.67	2.67
Yeast synthetic drop out medium (Y0750)			0.62	0.62	0.62	0.62
Agar		8		8		8
Leucine					0.048	0.048
Autoclaved MilliQ H ₂ O	Up to 400 ml	Up to 400 ml	Up to 400 ml	Up to 400 ml	Up to 400 ml	Up to 400 ml

3.12.2 Genotype

The strain used for all yeast two hybrid assays was L40. The genotype for this strain is *trp1 leu2 his3 ade2 LYS2::(lexAop)4x-HIS3 URA3::(LexAop)8x-LacZ*. The *trp1* and *leu2* inactivations were utilised for selection of colonies that have transformed pBTM116-mod (provides TRP1 gene) and pVP16-mod (provides LEU2 gene). The *URA3::(LexAop)8x-LacZ*. Modification allows quantitative assessment of β -galactosidase activity for evaluating interaction of proteins.

3.12.3 Transformation

An overnight culture was made by taking a single L40 yeast colony from a streak plate and inoculating 10 ml YPD under sterile conditions. The culture was incubated at 30°C for 16 hours while shaking at 250 RPM. The culture was diluted to 0.2-0.3 OD₆₀₀ and returned to the shaking incubator. At OD₆₀₀ 0.8, the culture was centrifuged at 4000 RPM for 5 mins. The supernatant was removed and the pellet resuspended in 25 ml sterile H₂O. This wash step was repeated twice. It was then resuspended in 1 ml H₂O and put into a sterile 1.5 ml Eppendorf tube. It was centrifuged for 15 secs at 15000 RPM and the supernatant was removed. The washed pellet was resuspended in 250 μ l 0.1 M lithium acetate (sterile filtered). The cells were then incubated at 30°C for 15 mins while shaking at 250 RPM. For a 10 X transformation mastermix, the following reagents were mixed: 1.2 ml 50% PEG 3350 (sterile filtered), 180 μ l 1 M lithium acetate (sterile filtered), 250 μ l 2 mg / ml salmon/herring sperm DNA (boiled at 95°C for 10 mins and

cooled rapidly on ice) and 100 μ l autoclaved H₂O. 173 μ l of this mastermix was added to each transformation tube. 100 ng of each vector and 25 μ l of the cells were added to the mastermix. The mixture was incubated at 30°C for 30 mins while shaking at 250 RPM. The cells were then heat shocked at 42°C and plated on SD -L -T or SD -T plates. Typically colonies would be a reasonable size in 2-4 days. Single colonies were re-streaked in triplicate and incubated in the same manner.

3.12.4 B-galactosidase assay

Half a loop of each transformant was picked from the SD transformation plate and used to inoculate 15 ml SD -L -T or SD -T. The cultures were briefly vortexed prior to incubation at 30°C for 16 hours while shaking at 250 RPM. The cultures were harvested by centrifugation at 4000 RPM for 5 mins. The supernatant was removed and the pellet resuspended in 1 ml autoclaved MilliQ H₂O. The cells were transferred to a 1.5 ml Eppendorf tube and centrifuged at 15000 RPM for 15 secs. The washed pellet was resuspended in 150 μ l Z buffer (60 mM Na₂HPO₄, 40 mM NaH₂PO₄, 10 mM KCl, 1 mM MgSO₄, 50 mM β -mercaptoethanol, adjusted to a final pH of 7.0). 150 μ l of glass beads were added to the cells and lysis performed using a Disrupter Genie. Cells were disrupted for a total of 3 minutes – broken into 1 minute disruption and 1 minute incubation on ice. The lysate was clarified by centrifugation for 20 mins at 4°C.

For the assay, 2-50 μ l of lysate was added to Z buffer to make a total of 500 μ l in a spectrophotometer cuvette. To start the assay, 100 μ l of 4 mg / ml ortho-Nitrophenyl- β -galactoside was added to the cuvette. It was lightly vortexed to ensure homogeneity

and the timer was started. Once the solution has turned a strong yellow colour, or 30 minutes has elapsed the reaction was stopped with 250 μ l 1 M Na₂CO₃ and the OD₄₂₀ was measured.

The Bradford assay was used to measure the concentration of protein in the lysate. For this, a serial dilution of BSA was made from 1 to 0.03125 mg / ml to use as a standard curve. 10 μ l of each lysate or BSA standard was added to 200 μ l of Bradford reagent in a clear 96-well microtiter plate. After 10 mins the OD₅₉₅ of the well was read using envision plate reader. If the lysate was above 1 mg / ml then it was diluted until it fell into this range. The standard curve was plotted and the concentrations of the lysates were calculated.

To calculate the specific activity, the following equation was used:

$$\text{Specific activity} = \frac{OD \times 0.85}{0.0045 \times \text{concentration} \times \text{volume} \times \text{time}}$$

Where OD is the absorbance at 420 nm after stopping the reaction, 0.85 is the volume in ml, 0.0045 is the optical density of a 1 nmol/ml solution of o-nitrophenol at 420nm, concentration is the lysate concentration in mg/ml, volume is the volume in ml of the lysate in the assay and time is the reaction time in mins before the reaction was stopped. Specific activity is, therefore, measured in nmol/mg/min.

3.13 ELISA

ELISA was performed to investigate the interaction between ubiquitin and the Znf-UBP domain, PLK1 poloboxes and the USP20 DUSP domains, and FL-PLK1 and TF-USP20FL. To assess binding of proteins, an ELISA was performed using an un-tagged protein to

capture the His-tagged probe. Anti-histidine primary antibodies was used to detect the probe and an enzyme labelled anti-IgG antibody was used as the secondary antibody. 96-well clear Maxisorp plates were used as these have a high binding capacity. 100 μ l of test and control proteins were bound to the plate at 10 μ g/ml in TBS (150 mM NaCl, 50 mM Tris pH 7.5). Blank controls were made by binding 100 μ l 3% (w/v) skimmed milk TBS to the wells. The protein, or milk, was incubated at 4 °C overnight to ensure proper binding.

The following day, wells were emptied and washed once with 400 μ l TBS per well. They were then blocked with 400 μ l 3% (w/v) skimmed milk TBST (150 mM NaCl, 50 mM Tris pH 7.5, 0.025% Tween 20). Blocking milk was incubated at room temperature for 1 hour. Blocking milk was washed once with 400 μ l TBST per well and once with 400 μ l TBS per well.

The wells were then probed with 100 μ l of the hexa-histidine-tagged protein or control for 2 hours at room temperature. The probe or control wells were washed twice with TBST and once with TBS. 100 μ l mouse anti-tetra-His primary antibody (Qiagen) was used as the primary antibody, diluted 1:1000 in 3% (w/v) skimmed milk TBST. The antibody was incubated for 1 hour at room temperature and washed twice with TBST and once with TBS.

For detection, 100 μ l horseradish peroxidase (HRP)-tagged anti-mouse IgG antibody was used at 1:2000 dilution in 3% (w/v) skimmed milk TBST. The antibody was incubated for 1 hour at room temperature and wash three times with TBST and once with TBS.

3,3',5,5'-tetramethylbenzidine was used as a colourimetric substrate. The HRP on the secondary antibody causes the conversion of 3,3',5,5'-tetramethylbenzidine (TMB) to 3,3',5,5'-tetramethylbenzidine diamine, which forms a blue colour. After a strong blue colour develops, the reaction is stopped with 100 μ l 1M H₂SO₄. This converted the blue colour to a yellow colour. Absorbance was measured using an EnVision 2104 multilabel plate reader with a monochromator detecting the absorbance of the wells at 450 nm.

3.14 ITC

Isothermal titration calorimetry (ITC) experiments were used to assess binding between ubiquitin and the Znf-UBP domain of USP20. A VP-ITC microcalorimeter was used, which requires approximately 2 ml of cell sample and 0.5 ml syringe sample. Both the Znf-UBP domain and ubiquitin were run on a gel filtration Superdex 75 column using the same buffer to prevent buffer mismatch. Ubiquitin was used in the syringe as it's very soluble and can reach the required concentrations for the syringe; the Znf-UBP domain, although it is also substantially soluble, would likely not achieve 300-400 μ M. 400 μ M ubiquitin was degassed and loaded into the syringe. The cell was filled with 40 μ M Znf-UBP domain. Ubiquitin was titrated into the sample cell in 8 μ l, 8 second injections or 16 μ l, 8 second injections. The temperature was set to 25°C with a stirring speed of 300 rpm. Injections were set to 200 second intervals.

3.15 Far-western blot

PVDF membrane was activated with methanol for 5 minutes, then equilibrated in 1 X Western blot transfer buffer. This was allowed to dry and the proteins were spotted on using a pipette. The total masses of poloboxes and control proteins were 0 (buffer only control), 0.625, 1.25, 2.5, 5, 10, 20 and 40 pmol. 5 pmol of the positive control spots (His-tagged DUSP domains) were spotted on to ensure the assay was working. The DUSP domains and DU15 were used to probe the poloboxes and BSA. As a further control BSA was used to probe the poloboxes to ensure no non-specific signal was arising. The spotted on proteins were allowed to dry and the membrane was blocked with 3 % Skimmed milk TBS for 2 hours at room temperature on a rolling mixer. The blocking buffer was poured of and the probes were incubated with the membrane at 100 µg/ml for 1 hour at room temperature on a rolling mixer. The membrane was washed twice with 10 ml TBST for 5 minutes on a rolling mixer. Following this, detection was performed exactly as the normal Western blot.

3.16 Mammalian lysate

HEK293 cells were used to produce the lysate for pull down assays. Complete MEM was made by combining 60 ml foetal bovine serum, 16 ml minimal essential medium, 12 ml L arginine, 6 ml non-essential amino acids and 6 ml penicillin/streptomycin. One tube of frozen HEK293 cells was taken from the cell bank. It was partially defrosted at room temperature then placed on ice to thaw the remaining cells. The cells were added to 9 ml pre-warmed 37 °C complete MEM. The cells were pelleted by centrifugation at 1000

g for 2 minutes and resuspended in 15 ml pre-warmed 37 °C complete MEM. This was then put into a T75 flask and incubated at 37 °C in a static incubator. At 80% confluency, the cells were subcultured into a new T75 with new medium. For subculture, the media was poured off and the cells were washed with 10 ml sterile PBS. This was poured off and replaced with 15 ml of complete MEM. HEK293 cells are only semi-adherent, so the cells do not need to be trypsinised; a gentle tap of the flask releases the cells. Cells were separated with gentle pipetting and 1.5 ml of cells was added to 13.5 ml of complete MEM. This 15 ml subculture was put into a new T75 for maintaining the culture.

To make the lysate, the culture was subcultured into T225 flasks with a total volume of 45 ml of media. At 90% confluency, the cells were harvested. The medium was removed from the T225, the cells were washed with 15 ml sterile PBS twice and 3 ml of ice-cold RIPA buffer (50 mM Tris pH 7.5, 150 mM NaCl, 1% SDS, 1% Na deoxycholate, 1% Ipegal 630, protease inhibitor cocktail tablets) was added. A cell scraper was used to remove all cells from the flask. The lysate was passed through a 23 Gauge hypodermic needle 5 times to break any remaining cells and shear DNA. The lysate was clarified by centrifugation at 13000 RPM for 20 minutes at 4 °C. For storage, the lysate was snap-frozen in liquid nitrogen and kept at -80 °C until required.

3.17 Pull down assays

0.1 g of cyanogen activated beads was used per pull down assay. Dry beads were put into a 5 ml polypropylene column (Qiagen). They were washed with 20 ice-cold 1 mM HCl, 25 ml distilled molecular grade H₂O and 0.5 ml buffer (0.1 M NaHCO₃ pH 8.4 and 0.5 M NaCl) using the gravity flow column. The Znf-UBP domain (USP20 1-101) and the DUSP domains (USP20 686-894) were gel filtered with the 0.1 M NaHCO₃ pH 8.4 and 0.5 M NaCl buffer. The beads were incubated with 2 mg of protein overnight at 4 °C on a rolling mixer. The beads were washed with 5 ml carbonate buffer to remove residual protein, and blocked with 1 M ethanolamine pH 8.0 for 2 hours at room temperature on a mixing roller. Ethanolamine was washed away with 5 ml carbonate buffer. 5 cycles of washing with 5 ml carbonate buffer and 5 ml 0.1 M sodium acetate pH 4 was performed and the column was finally equilibrated in carbonate buffer. The affinity column is now formed with the USP20 proteins covalently bound to the agarose beads.

The beads were incubated with 5 ml of HEK293 lysate for 2 hours at 4 °C on a rolling mixer. The lysate was then allowed to pass through the column and the beads were washed with 0.5 ml RIPA buffer. The beads were then incubated with 5 ml RIPA buffer and incubated for 20 minutes at 4 °C on a rolling mixer. This was allowed to pass through the column and the beads were washed with 1 ml RIPA buffer. The beads were then incubated with 5 ml 50mM Tris pH 7.5, 150mM NaCl, 5mM DTT. This was

allowed to pass through the column and the beads were washed with 1 ml 50mM Tris pH 7.5 then 1 ml 50mM Tris pH 7.5, 1M NaCl. To elute the proteins, the beads were boiled in 100 μ l of SDS-PAGE sample buffer for 5 minutes. The samples were run on a large gel. Specific bands were excised and sent for LC-MS/MS at Cambridge Centre for Proteomics.

4 Results: expression, purification and crystallisation

4.1 Construct design

There is currently little information on the domain architecture of USP20, and no structural information is available. A bioinformatical analysis of USP20's domain structure was conducted to characterise its domain architecture and inherent disorder. First, the domain boundaries were determined. This was done using online bioinformatics servers and looking at homologous proteins that have already been characterised. Second, disorder plots helped to ensure the designed constructs contained as few disordered residues as possible, as this is preferential for crystallisation. Once these characteristics of the protein were determined, specific domains or domain groups were expressed and used for crystallisation trials. Once crystals formed in these trials, X-ray crystallography was to be used to determine the structure of the protein.

4.1.1 USP20 bioinformatics

4.1.1.1 Domain prediction

Bioinformatics servers recognise four domains in USP20: a disordered central catalytic USP domain flanked N-terminally by a Znf-UBP domain and C-terminally by two DUSP domains (Figure 4.1). These domains can be identified by multiple bioinformatics servers including Pfam, SMART and PROSITE and GENE3D. Each server predicts similar domains, but they have slightly different boundaries (Table 4.1). Therefore, other evidence must be used in conjunction with these data to

properly identify suitable domain boundaries for recombinant expression of crystallisable protein.

Figure 4.1. Domains of USP20. A schematic diagram of USP20 is shown. There is an N-terminal Znf-UBP domain (blue), a central catalytic domain (red) and two C-terminal DUSP domains (orange). Grey regions represent non-domain portions of USP20.



Table 4.1. Domain boundary predictions

Domain	Server	Residues
Znf	GENE3D	7-92
Znf-UBP	PROSITE	28-92
	SMART	29-81
	Pfam	30-91
USP3	PROSITE	145-685
Peptidase C19 UCH	Pfam	145-249 ; 425-682
Peptidase C19 UCH, DUSP domain	SUPERFAMILY	711-779 ; 777-888
	PROSITE	687-779 ; 788-891
	pFAM	816-887
	GENE3D	724-780 ; 781-900
	SMART	702-784 ; 809-894

USP20 DUSP 1 is not detected by the pFAM search. This is most likely due to the lower sequence homology with other DUSP domains. The first DUSP domains of both USP20 and USP33 along with DUSP domains 2 and 3 of USP48 form a separate clade to the domains of USP4, USP11, USP15, and the other USP20, USP33 and USP48 DUSP domains.

The Conserved Domain Architecture Retrieval tool [307] analyses protein sequences, and uses the Conserved Domain Database and RPS-BLAST to identify other proteins with the same or similar domain architecture. Interestingly, this domain architecture – Znf-UBP, C19 Peptidase, DUSP, DUSP – appears to have arisen in the Eukaryota. This exact architecture can be seen in 817 sequences in the database. Additionally, in some placental mammals, there are 56 sequences with a predicted additional domain interposing the catalytic domain. However, upon assessing phylogeny (Figure 4.2) and alignments (not shown) it is likely that these predicted additional domains are not present in the proteins, and are actually an artefact from the statistical prediction methodologies. These sequences show reasonable conservation, even in the regions where the additional domains are predicted.

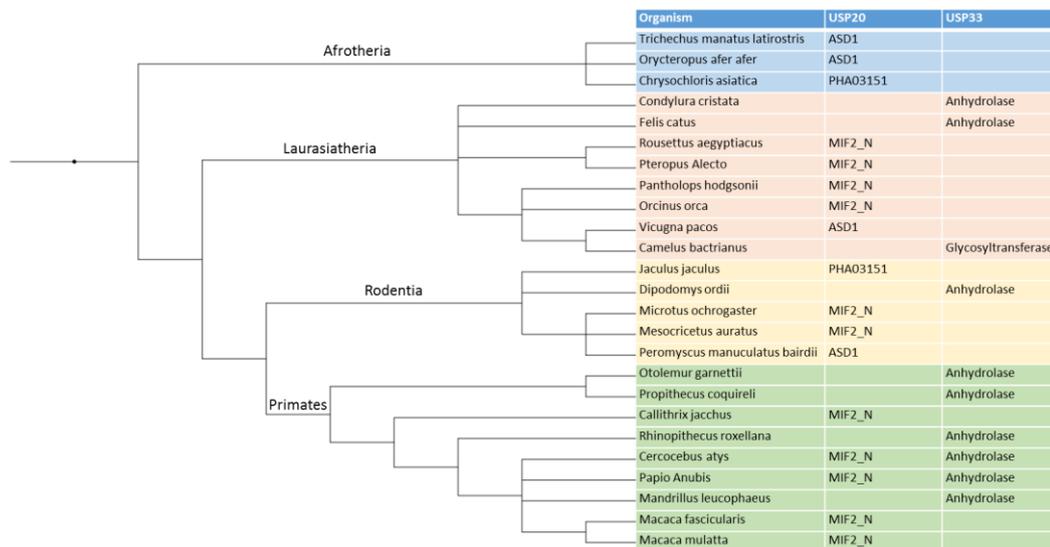


Figure 4.2. Phylogeny of species and inserted domains. A phylogeny of species that were predicted to have an additional domain within the catalytic domain of either USP20 or USP33 is shown. 5 different domains were predicted sporadically throughout the phylogeny. Either convergent evolution has occurred multiple

times, or the predictions are untrue and are artefactual. Blue boxes represent the Aftorheria, Red boxes Laurasiatheria, Yellow Rodentia and Green Primates.

803 of the sequences with exact matching architecture are found in the animalia, the remaining 14 sequences include diplomonads, fungi, Capsaspora, cercozoans, ciliates, Ichthyophonida, Longamoebia and Fonticula. OrthoDB, a database that identifies gene orthologs also finds 141 genes in 92 species that include Fungi, arthropods, vertebrates and metazoa. Although most of these are uncharacterised proteins and the Uniprot entries are unreviewed, it shows there has been a necessity for a deubiquitinating enzyme with this domain architecture since the presence of early eukaryotes. Whether or not the functions of these genes are analogous to that of human USP20 or USP33 is unknown. Indeed, high conservation of the USP20 protein sequence is observed from fish to mammalia (Figure 4.3), which suggests that its role may also be conserved. USP20 also shows reasonable similarity to its paralogue, USP33. A sequence alignment (Figure 4.4) of human USP20 and USP33 shows high sequence identity of their predicted domains, but lower homology in their non-domain sequences. Again, this suggests that their domains are involved in similar processes, and their non-domain regions may contribute to the individual roles of the two proteins.

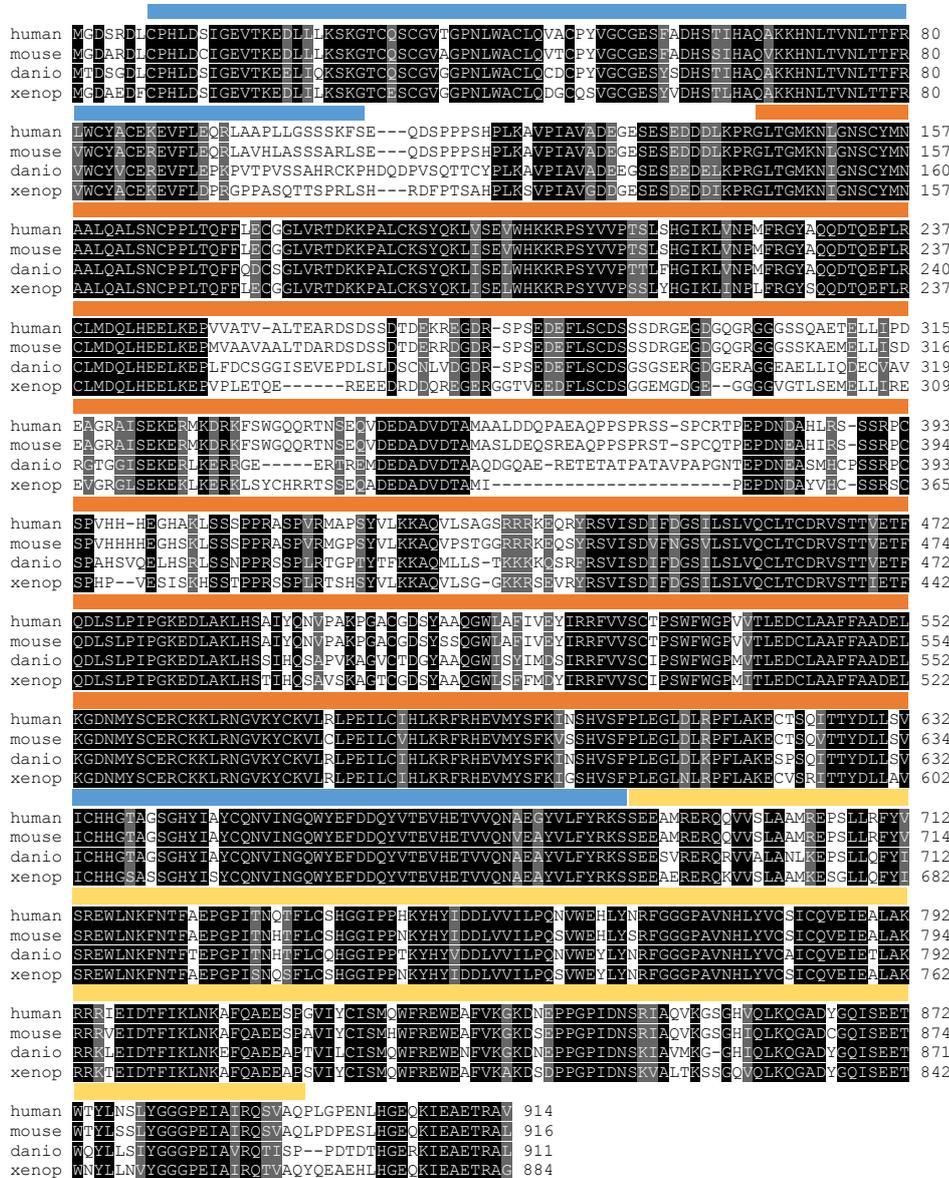


Figure 4.3. Sequence alignment of USP20 across four evolutionary divergent species. The sequence alignment shows high conservation of the USP20 sequence from fish to mammals. Human (*Homo sapiens*), mouse (*Mus musculus*), Zebrafish (*Danio rerio*) and Western clawed frog (*Xenopus tropicalis*) were aligned with 65.6% identity. Identical residues are labelled in black, similar residues are in grey and non-similar residues are in white. The coloured bar above the sequence shows the domains of USP20. Blue = Znf-UBP; Red = Catalytic domain; Yellow = DUSP domains.

```

USP20 -----MGDSRLCPHLDSIGEVTKEDLILKSKGTCQSCGWTGPNLWACLQVACP 49
USP33 MTGNSNSHITILTTLKVLPHFESLGGQEKIPNKRVSARFNHCPHLDSVGEITKEDLILKSLGTCQSCKWTGPNLWACLNRCS 80

USP20 YVCGGESFADHSTIHAQAKKHNLTVNLTFRLWCYACBKEVFLEORLAPLLSS----SKF---SEQDSPPF--SHPLKA 121
USP33 YVCGGESQVDHSTIHSQETKHVLTVNLTFRLVWCYACSKKEVFLRKLQTPSTPHVRQPHQIQENSVQDFKIEBNTTLKT 160

USP20 VFLTAVADDEGESESEDDDLKPRGLTGMKNLGNSCYMNAAALQALSNCPPLTQFFLECGGLVRTDKKPAICKSYCKLIVSEVW 200
USP33 PLVAVFDDLDIEADEDEELRARGLTGLKNTGNFCYMNAAALQALSNCPPLTQFFLDCGGLARTDCKPAICKSYLKLMLLW 240

USP20 HKKRFSYVVPVPTSLSHGTLKLVNEMFRGYAQQDTQFFLRCLMDQLHEELKEPVAIVATVALTEARSDSDTTEKREKDRSPSE 280
USP33 HKSRFGSVVPTLIFCGLIKTVNEMFRGYSQODAOBFLRCLMDQLHEELKEQVMEVEE-----DFQITITTEETMDEKQSQS 315

USP20 DEFLSCDSSSDRGEGL--GQGRGGGSSCAETELLIPDEAGRAISEKERMKDRKFSWQQQR-TNS-----EVDDEADV 350
USP33 VDFQSCSNCNSDRAPNENGSRCFSEDNNEETMLIQDDENNSEM-----SKDWQKEKMCNKINKVNSEGEFDKQRDS 387

USP20 DITAMALDDCPAEAQPSPRSSSPORTEPEPDNAHLRSSSRFCSPVHHHEHAKLSLSPPRASPFRMAPSYVLKKAQVLS 430
USP33 IGETVLDLNNQETVKVQ-----IHSRASEYITVHSNDLSTEQILSPNEGKPNRLSSPPEKSGNWPGLAPPHKKAQ-SA 460

USP20 AGSRERKEQRYRVSVIDIFDGSILSLVQCLTCDRVSTVETFDQLSLPIPGKEDLAKLHSAIQVNPVAKPFAAGDSYAA 510
USP33 SPKRKQHKKYRVSVIDIFDGSILSSVQCLTCDRVSTVLETFDQLSLPIPGKEDLAKLHSSSHPTSIVKACSCGEAYAP 540

USP20 GWLAFIVEYLRFRVVSCTPSWFWGPPVTLDCLAFFAADDELKGDNMYSCEKCKKLRNGVKYCKVLRLEPILCIHLKRFR 590
USP33 GWLAFMEYVYKRFVVSCTPSWFWGPPVTLDCLAFFARDELKGDNMYSCEKCKKLRNGVKYCKVQNPPEILCIHLKRFR 620

USP20 HEVVMSEPKINSHVSFPLEGLDLPFLARECTSQTITVDLLSVICHHGTAAGSGHYIAYCQNVINGQWYEFDDQYVTEVHEI 670
USP33 HELMPSIKISTHVSFPLEGLDLPFLAKDSPAQIVTYDLLSVICHHGTAAGSGHYIAYCRNLLNINLWYEFDDQSVTEVSES 700

USP20 WQNAEGYVLFYRKSSEEAARFQQVVSLEAMRFPSLIRFYVSRWLNKFRFFAEPGPIINQTFLCSHGGLPPHKYHYID 750
USP33 WQNAEGYVLFYRKSSEEAQKERRRISNLNIMRPSLLOFYVSRWLNKFRFFAEPGPIINQTFLCLHGGVPPKACGYID 780

USP20 DLVWILPONVWEHLVNRFGGPAVNHLVYCSICQVELEAIAKRRRIEIDTFIKLNKAFQAEESPVIYCI SMQWFRWEA 830
USP33 DLVWILPONVWNLVYSRYGGGPAVNHLVYCHTCQLEAEKIEKRRTELEIFIRLNRAFOKEDSPATFYCISMQWFRWEBS 860

USP20 FVKGKDNPPGPIDNSRIAQVKGSCHVQLKQCADYQOISEETWTLNLSYGGGPEIATRQSVVQPIGPNHGEOKIEAB 910
USP33 FVKGKDCDPPGPIDNFKIAVT-KCENVMRLROGADSQOISEETWTLQSYGGGPEVILRPFVW-HVDFPIIQAEKIEVE 938

USP20 FRAV 914
USP33 FRSI 942

```

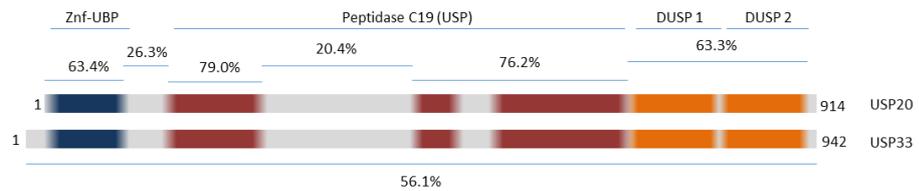


Figure 4.4. Alignment of USP20 and USP33. (A) Sequence alignment of USP20 and USP33. Identical residues are labelled in black, similar residues are in grey and non-similar residues are in white. **(B)** Schematic showing the identity observed between USP20 and USP33. High levels of sequence homology are observed in all the domains of USP20 and USP33. Low homology is observed in the non-domain regions of the proteins: between the Znf-UBP domain (blue) and catalytic domain (red), and within the insert inside the catalytic domain.

4.1.1.2 Disorder

Protein disorder is detrimental to crystallisation and disordered residues or loops should be avoided if possible. Also, the disordered regions of a protein can help to identify the domains within the protein as they are generally well folded with the

presence of secondary structure. Disorder plots using DisEMBL [308] were produced to identify these regions within USP20 (Figure 4.5).

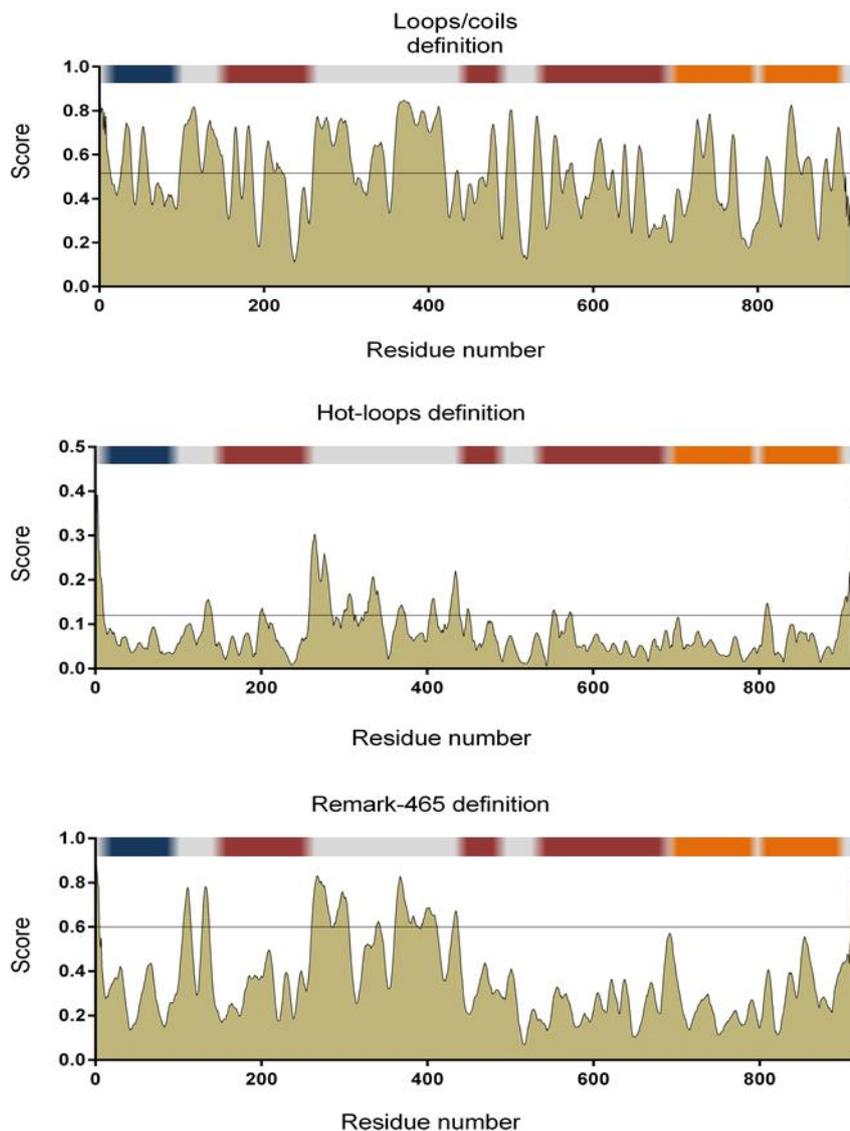


Figure 4.5. Disorder plots of USP20. Three disorder plots are given. **A.** Loops/coils definition. **B.** Hot-loops definition. **C.** Remark-465 definition. Above each graph is the USP20 schematic showing the Znf-UBP domain in blue, Catalytic domain in red, DUSP domains in orange, and non-domain regions in grey. The threshold for each disorder definition is shown as a horizontal line on the graph. Residues with disorder values above this line are predicted to be disordered.

The hot-loops and Remark-465 definitions seem to be most informative for identifying regions of disorder in USP20. In particular, the N- and C-termini show high probabilities of disorder. Also, the region between the Znf-UBP domain and catalytic domain, and the large insert between boxes two and three of the catalytic domain show increased probability of disorder (Figure 4.6). For constructs to be expressed and crystallised, it may be important to omit these residues.

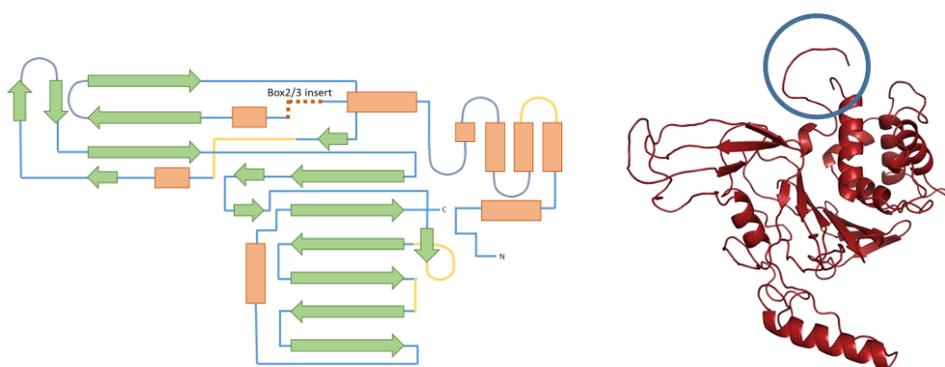


Figure 4.6. Topology of USP20. (A) The general topology of the core catalytic USP domain is shown (adapted from Ye *et al.* [87]). Helices are shown as red boxes, β -strands as green arrows, common insertion points are coloured yellow. The insert between box 2 and three is shown as a dotted red line. USP20 has a 182-residue insert at this point. (B) A homology model produced by PHYRE2 shows the typical fingers-palm-thumb structure of the USP domain. The insert region for USP20 is shown with a blue circle.

4.1.2 Znf-UBP domain design

The N-terminal domain of USP20 is the Znf-UBP domain and we chose to express this as a single domain for crystallisation. Using sequence and structure alignments of USP20 and USP33, a construct was designed that comprised residues 1-101 of USP20. Residue 101 follows the end of the sequence similarity between USP20 and USP33, but also precedes the predicted disordered residues. The USP33 Znf-

UBP domain structure has previously been solved by NMR [84]. The USP20 and USP33 Znf-UBP domains show 63% identity, and therefore should be similar in their structures. The USP20 1-101 construct is analogous to the USP33 protein used to produce the NMR structure. Although the N-terminus was also predicted to be disordered, according to the alignments with USP33, the secondary structure appears to initiate on residue eight. These disordered residues could be removed from the construct; however, they are few and a methionine would still have to be incorporated to initiate translation. For this reason, the construct was designed from residue one.

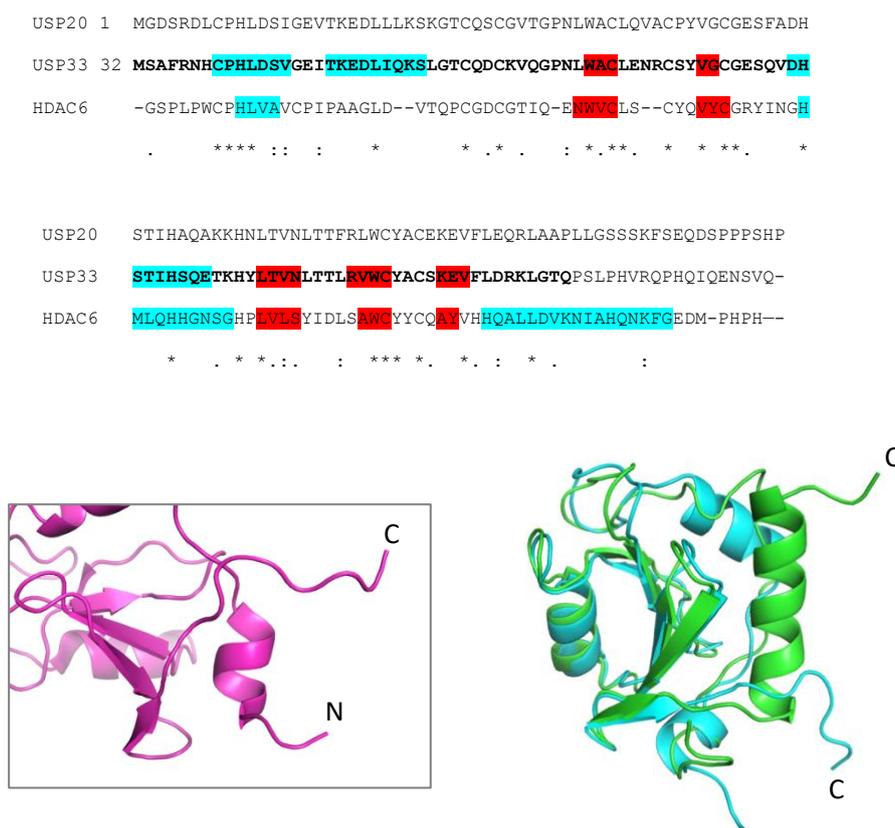


Figure 4.7. Znf-UBP domain structure and boundaries. (A) Sequence alignment of USP20, USP33 and HDAC6 Znf-UBPs. **(B)** N- and C-termini of USP33 Znf-UBP NMR structure. **(C)** Two superimposed homology models of the USP20 Znf-UBP produced by SWISS-MODEL. The two homology models are based on HDAC6 (green) and USP33 (cyan).

It was noticed that the structures of HDAC6's zinc finger domain (3C5K and 3GV4) show a helix after the last beta-strand of the fold. This helix is not found in the USP33 NMR model, although according to the alignment, the model would terminate during this helix (Figure 4.7; residues 32-130 in bold). It may be that the USP33 zinc finger does not form this secondary structure due to primary sequence differences, but it's also possible that if all of the residues were present, then this helix would also be present in the NMR structure. As the USP20 1-101 construct may also terminate during this helix, and to ensure that the N-terminal disorder does not affect crystallisation, four more constructs were produced: 1-92, 1-108, 6-92 and 6-108. According to the alignments, these constructs would terminate either before or at the end of this possible helix. Also, the first five amino acids were removed in two of the constructs to reduce the N-terminal disorder.

4.1.3 Catalytic domain design

4.1.3.1 Catalytic domain boundaries

Four constructs were designed that contained a full catalytic domain. These are USP20 full length (USP20FL), USP20catalytic, USP20 Δ DUSPs, and USP20 Δ Znf-UBP. A multiple sequence alignment of the C- and N-termini of multiple USPs catalytic domains show a highly conserved N-terminus. Both PROSITE and Pfam predict that the domains start at residue T145 (sequence T-G-M-K), however sequence

conservation starts at residue G146 (Figure 4.9). Homologous terminal residues are located exiting the domain between a β -sheet and helical bundle in structures of USP4 [309], USP5 [310] and USP8 [311]. The construct was to be cloned into a pCOLD1 vector to enhance the expression of properly folded enzyme. The constructs in pCOLD1 have a 16-residue N-terminal tag, and to minimise residues in the final constructs, residue M147 was used as the N-terminal USP20 residue for USP20catalytic and USP20 Δ Znf-UBP. Comparing USP20 homology models and the known structures of the other catalytic domains also show that the new 16 residue tag should be free and exposed from the catalytic domain, ensuring binding to the nickel column (Figure 4.8).



Figure 4.8. Catalytic domain N-terminus. The crystal structures of USP4, USP5 and USP21 are superimposed and shown in red. The yellow residue is the first residue taken for USP20 constructs that have the start of the catalytic domain as their N-terminus. The residues of the tag from the pCOLD1 vector would be exposed and be disordered like the N-termini shown in the figure.

The C-terminus of the catalytic domain is predicted to be between residue 682 and 685. Known structures show that the C-terminal secondary structure ends within one or two residues after the homologous residues to the USP20 V-L-F-Y motif. Using all of this information, residue 685 was used as the C-terminus for USP20 catalytic and USP20ΔDUSPs

```

USP20 -----ESESEDDDLKPRGLTGMRNLGNMNAALQALSNCPLTQFFLECGGLVR 181
USP5  RIGEWELIQESGVPLKPLFGPGYTGIRNLGNMNAALQALSNCPLTQFFLECGGLVR 362
USP21 SSEPFYS-DDKMAHHTLLLGSGHVGLRNLGNTCFNAVLQCLSSTRPLRDFCLRRDFRQE 248
USP45 -----CETDEIQKGGKCRNLSVRGITNLGNTCFNAVMQNLAQTYTLTDLMNEIKESST 226
USP46 -----E--KDIGPEQFPINEHYFGLVNFVNTCYCNSVLQALYFCRPFRENVLAYKAQQK 71
USP2  APGFSRSSSPGRDGMNSKSAQGLAGLRNLGNTCFMNSILQCLSNTRELRYCLQLRYMRD 303
USP1  -----S--SPINCEKRENLLPFVGLNMLGNTCYLNSILQVLYFCPGFKSGVKHLFNIIIS 117
          *: **:*:*: * : * *      :

USP20 --NGQWYFDDQYVTEVHET-----VVQNAEGYVLFYRKSSEAMRERQV-- 696
USP5  ---GRWVIYNDQKVCASEKP-----PKDLGYIYFYQRVAS----- 858
USP21 ---TGWHVYNDSRVSPVSEN-----QVASSEGYVLFYQLMQEPPRCL----- 565
USP45 ESAGQWVHVSPTYLQVVPES-----RALSAQAYLLFYERVL----- 814
USP46 ---GFWLLFDDDIWEKIDAQAIEEFYGLTSDISKNSSEGYILFYQSRE----- 366
USP2  --TGEWHTFNDSSVTPMSS-----QVRTSDAYLLFYELASPPSRM----- 605
USP1  ---DLNSLELDKGNFVVDQM---CEIGKPEPLNEEEARGVVENYNDEEVSIIRVGGNTQPS 655
          *                          . : * .

```

Figure 4.9. Catalytic domain alignments. (A) Catalytic domain N-terminus. **(B)** Catalytic domain C-terminus.

The N-terminus shows a highly conserved region for all USP domains. The C-terminus shows some homology, ending with a conserved hydrophobic sequence containing a Tyr residue.

4.1.3.2 Removal of the catalytic domain insert

Ye *et al.* [87] showed that the USP domains can be split into sections known as boxes. The intervening sequences between the boxes are the most common insertion points in the catalytic domains of USPs. USP20 has a 182 residue insert

between boxes two and three. It also has a 52 residue insert between boxes three and four. The disorder plot shows high disorder in the large insert, which may contribute to poor expression, stability and crystallisation of diffraction quality crystals. For this reason four constructs were made by removing this large insert: USP20FL Δ insert, USP20catalytic Δ insert USP20 Δ DUSPs Δ insert and USP20 Δ Znf-UBP Δ insert.

USP20 must be able to fold properly following removal of its large insert, so it is essential to look whether the amino acids can be removed without complication, or whether some residues would be required to connect the two halves of the catalytic domain properly. Using a predicted structure produced by SWISS-MODEL and PHYRE2, structural superpositions were made with the solved structures of USP4, USP5 and USP21 (Figure 4.10). Between boxes two and three, USP4 and USP21 have inserts of 22 and 37 residues, respectively, and USP5 has no insert in this location (and is lacking an additional 10 residues from the start of box three). The alignments suggest that, following removal of the insert between boxes two and three and joining of the flanking sequences, USP20 should still be able to fold. It is most likely that the remaining residues will adopt an intermediate sized loop when comparing the USP4/USP21 and USP5 inserts.

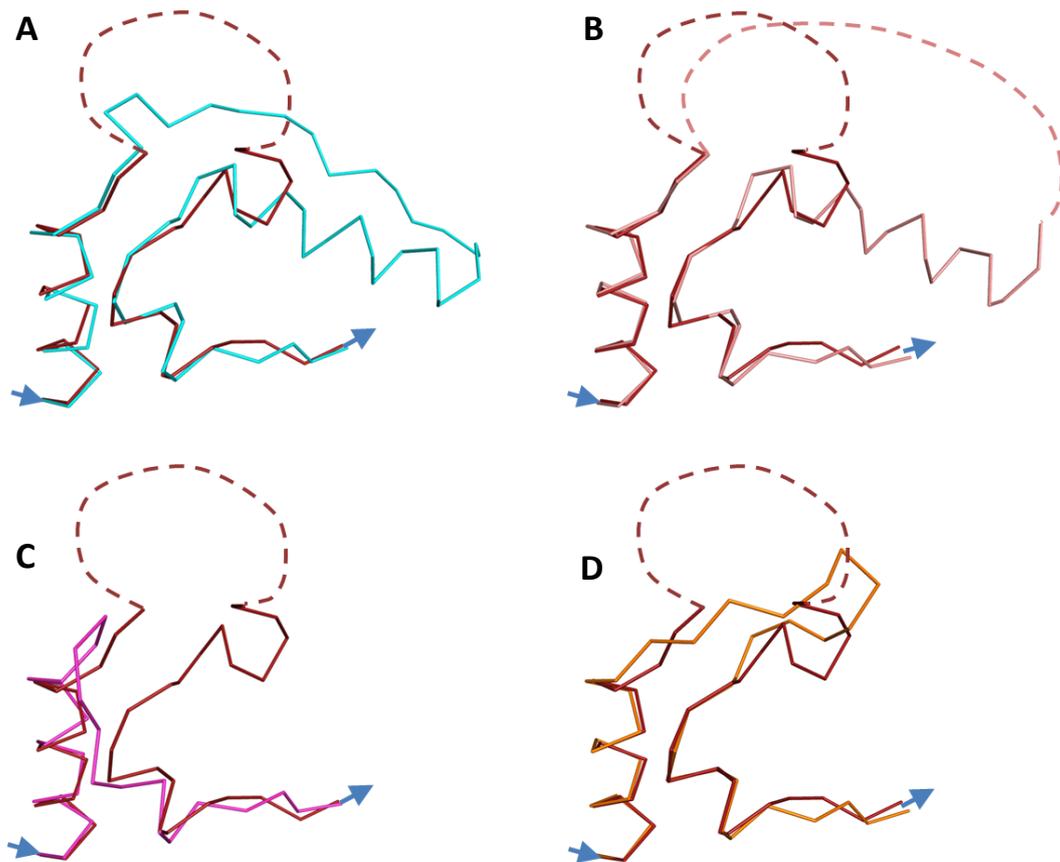


Figure 4.10. Structural alignments of Box 2/3. A homology model of USP20 (Red) has been superimposed with **(A)** USP4 PDB code 2Y6E, **(B)** USP21 PDB code 2Y5B and **(C)** USP5 PDB code 3IHP. **(D)** shows a superposition of the USP20FL model and USP20FL Δ insert. The arrows indicate direction of peptide chain from N- to C-terminus. The dashed lines show where a loop would join the chains, but are not observed in structures or models.

4.1.4 DUSP domains

Alignments using the DUSP domains from USP20 and USP33 show high sequence conservation across all vertebrates (Figure 4.11). The C-terminus of the catalytic domain has already been identified as most likely to be S685. The predictions for the

start of the DUSP domains vary widely, with N- and C-terminal residues ranging from 687-724 and 888-900, respectively. The predicted structures produced by PHYRE2 homology modelling suggest that the secondary structure initiates on A689 (helix) and ends on Q889 (β -strand) (Figure 4.12).

```

USP20 Homo sapiens          DQYVTEVHETVVQNAEGYVLFYRKSSEEAMRERQQVVS LAAMREPSLLRFYVSREWLNK 59
USP33 Homo sapiens          DQSVTEVSESTVQNAEAYVLFYRKSSEEAKERRRISNLLNIMEPSLLQFYISRQWLNK 59
USP20 Mus musculus         DQYVTEVHETVVQNV EAYVLFYRKSSEEAMRERQQVVS LAAMREPSLLRFYVSREWLNK 60
USP20 Xenopus tropicalis   DQYVTEVHETVVQNAEAYVLFYRKSSEEAEERERQVVS LAAMKESGLLQFYISRWLNK 60
USP20 Danio rerio          DQYVTEVHETVVQNAEAYVLFYRKSSEESVREERQVVALANLKEPSLLQFYISRWLNK 59
** **** *:.***.*.*****: :*::: * : * .*:*:*:*:**

USP20 Homo sapiens          FNTFAEPGPITNQTFLCSHG GIPPHKYHYIDDLVVL P QNVWEHLYNRFGGGPAVNHLVY 119
USP33 Homo sapiens          FKTF AEPGPISNNDFLCIHGGVPPRKAGYIEDLVLM LPQNIWDLNLSRYGGGPAVNHLVY 119
USP20 Mus musculus         FNTFAEPGPITNHTFLCSHG GIPPNKYHYIDDLVVL P QSVWEHLYSRFGGGPAVNHLVY 120
USP20 Xenopus tropicalis   FNTFAEPGPISNQSF LCSHG GIPPNKYHYIDDLVVL P QSVWEYLYNRFGGGPAVNHLVY 120
USP20 Danio rerio          FNTFTEPGPITNHTFLCQHG GIPPTKYHYVDDLVL P QNVWEYLYNRFGGGPAVNHLVY 119
*:*:*****:*. *** **:* * *:*****:***.:*: **.*:*****:

USP20 Homo sapiens          CSICQVEIEALAKRRRIEIDTFIKLNKAFQAEESPGVIY CISMQWFREWEAFVKGKDNEP
USP33 Homo sapiens          CHTCQIEAEKIEKRRKTELEIFIRLNRAFQKEDSPATFYCISMQWFREWEAFVKGKDGDP 179
USP20 Mus musculus         CSICQVEIEALAKRRRVEIDTFIKLNKAFQAEESPAVIY CISMHWFREWEAFVKGKDSFP 180
USP20 Xenopus tropicalis   CSICQVEIEALAKRRKTEIDTFIKLNKAFQAE EAPSVIY CISMQWFREWEAFVKAKDSDP 180
USP20 Danio rerio          CAICQVEIETLAKRRKLEIDTFIKLNKEFQAE EAPT VILCISMQWFREWFVKGKDNEP 179
* **:* * : ***: *:: **:*: * * :*: * .: *****:***** **:* * :*

USP20 Homo sapiens          PGPIDNSRIAQVKSGSHVQLKQGADYGGQISEETW T YLNSLYGGGPEIAIRQVAQPLGPE 239
USP33 Homo sapiens          PGPIDNTKIAVTK-CGNVMLRQGADSGQISEETWNFLQSIYGGGPEVILRPPVHVH-DPD 237
USP20 Mus musculus         PGPIDNSRIAQVKSGSHIQLKQGADCGQISEETW T YLSSLYGGGPEIAIRQVAQLPDPPE 240
USP20 Xenopus tropicalis   PGPIDNSKVALTKSSGQVQLKQGADYGGQISEETW N YLLNVYGGGPEIAIRQVAQYQEA 240
USP20 Danio rerio          PGPIDNSKIAVMK-GGHIQLKQGADYGGQISEETW QYLLSIYGGGPEIAVRQTISP--PDT 236
*****::* * *.: *:*** ***** :* .:*****: :* :

USP20 Homo sapiens          NLHGEQKIEAETRAV 254
USP33 Homo sapiens          ILQAEEKIEVETRSL 252
USP20 Mus musculus         SLHGEQKIEAETRAL 255
USP20 Xenopus tropicalis   HLHGEQKIEAETRAG 255
USP20 Danio rerio          DTHGERKIEAETRAL 251
:.*.***.***:

```

Figure 4.11. Sequence alignments. Multi-species sequence alignments of the DUSP domains of USP20 were produced by Clustal Omega. The domain is well conserved between species. There is a small stretch of variable sequence just following the terminus of the predicted end of the second DUSP domain, USP20 Q894 (Q234 in alignment). The C-terminal disordered tail is also well conserved, suggesting that it has some functional role in USP20.

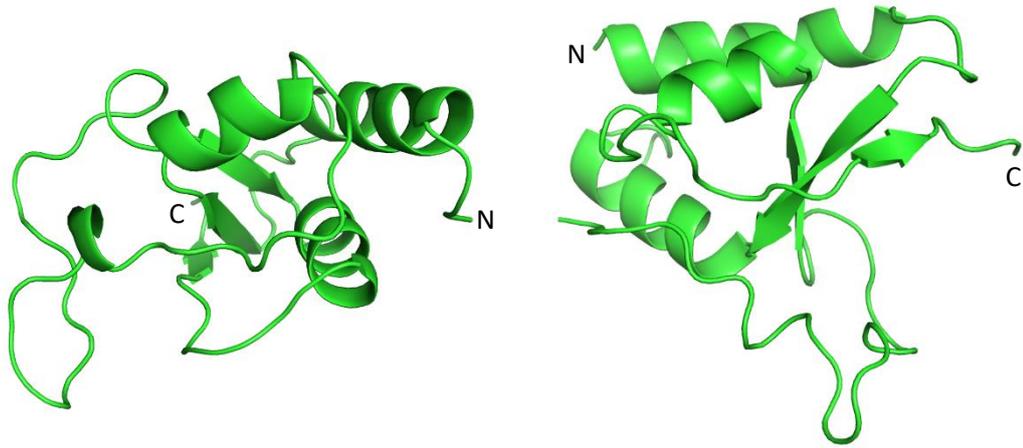


Figure 4.12. Homology models of the DUSP domains of USP20. Homology models of the DUSP domains were produced by PHYRE2. For both DUSP1 (left) and DUSP2 (right) the typical DUSP AB3 fold is predicted. When inserting the whole double DUSP sequence into PHYRE2, a model with reasonable confidence cannot be produced.

4.2 Cloning and mutagenesis

A summary of all constructs made by cloning and mutagenesis in the course of this thesis are tabulated below (Tables 4.2, 4.3 and 4.4) and further illustrated in Figure 4.13.

Table 4.2. *E. coli* expression constructs.

Name	Protein and residues	Plasmid	N-term modification	C-term modification	Internal modifications
USP20FL	USP20 1-914	pCOLD1	MNHKVHHHHHHIEGRH		
TF-USP20FL	USP20 1-914	pCOLD TF	MNHKVHHHHHH-(Trigger Factor)-AKVTEKETTFFNELMNQQASAGLEVLFGQPSAGLVPRGSGGIEGRH		
USP20 ΔDUSPs	USP20 1-685	pCOLD1	MNHKVHHHHHHIEGRH		
USP20 Catalytic	USP20 147-685	pCOLD1	MNHKVHHHHHHIEGRH		
USP20ΔZnf-UBP	USP20 147-914	pCOLD1	MNHKVHHHHHHIEGRH		
USP20FLΔinsert	USP20 1-914	pCOLD1	MNHKVHHHHHHIEGRH		Δ251-431
USP20ΔDUSPs Δinsert	USP20 1-685	pCOLD1	MNHKVHHHHHHIEGRH		Δ251-431
USP20 CatalyticΔinsert	USP20 147-685	pCOLD1	MNHKVHHHHHHIEGRH		Δ251-431
USP20ΔZnf-UBPΔinsert	USP20 147-914	pCOLD1	MNHKVHHHHHHIEGRH		Δ251-431
USP20 Znf-UBP	USP20 1-101	pET21d		HHHHHH	
USP20 Znf-UBP	USP20 1-92	pPROEx-Htb	MSYHHHHHHHDYDIPTTENLYGQGA		
USP20 Znf-UBP	USP20 1-108	pPROEx-Htb	MSYHHHHHHHDYDIPTTENLYGQGA		
USP20 Znf-UBP	USP20 6-108	pPROEx-Htb	MSYHHHHHHHDYDIPTTENLYGQGAM		
USP20 Znf-UBP	USP20 1-92	pPROEx-Htb	MSYHHHHHHHDYDIPTTENLYGQGAM		
USP20 DUSPs	USP20 684-914	pET26b		LEHHHHHH	
USP20 DUSPs	USP20 684-894	pET26b		LEHHHHHH	
USP20 DUSP	USP20 684-791	pCOLD1	MNHKVHHHHHHIEGRH		
1BKR	SPTBN1 174-278	pSIMON1	MAHHHHHHS		
1BKR-DUSPs	USP20 686-894	pSIMON1	MAHHHHHHS-(1BKR)-AAA		
2GKG	frzS 3-123	pSIMON2	MAHHHHHHA		K94A, K 96S
2GKG-DUSPs	USP20 686-894	pSIMON2-his	MAHHHHHHA-(2GKG)-AAA		
2GKG-his	frzS 3-123	pSIMON2	MAHHHHHHA		R46H, K71H, D73H, D74H, K76H, N77H, K94H, K96H
MBP-DUSPs	USP20 686-894	pMALX(E)	(MBP)-AAA		
MBP-PB	PLK1 367-603	pMALX(E)	(MBP)-AAAENLYGQGSA		

Details about proteins that were expressed in *E. coli* are given. For modifications, single letter amino acid code is used except for where whole or partial proteins are present (shown in brackets). All cloned proteins are the human orthologues, except for frzS, which is the *Myxococcus xanthus* orthologue.

Table 4.3. Solubility tag vectors.

Name	Protein and residues	Plasmid	N-term modification	C-term modification	Internal modifications
pSIMON1 (1BKR)	SPTBN1 174-278	pRSF-13	MAHHHHHHS	Multiple cloning site	
pSIMON2 (2GKG)	frzS 3-123	pRSF-13	MAHHHHHHA	Multiple cloning site	K94A, K 96S
pSIMON2-his (2GKG-His)	frzS 3-123	pRSF-13		Multiple cloning site	R46H, K71H, D73H, D74H, K76H, N77H, K94H, K96H

Details about proteins that were cloned to produce solubility tag vectors are given. For modifications, single letter amino acid code is used except for where whole or partial proteins are present (shown in brackets). SPTBN1 is the human orthologue; frzS is the *Myxococcus xanthus* orthologue.

Table 4.4. Yeast two hybrid constructs.

Name	Protein and residues	Plasmid	N-term modification	C-term modification	Internal modification
LexA-Znf-UBP	USP20 1-101	pBTM116mod	(LexA 1-202)-EF	GS	
LexA-Catalytic	USP20 147-685	pBTM116mod	(LexA 1-202)-EF	GS	
LexA-DUSPs	USP20 686-894	pBTM116mod	(LexA 1-202)-EF	GS	
VP16- Znf-UBP	USP20 1-108	pASV3mod	M-(L29 22-32)-(VP16 411-490)-GAILE		
VP16-Catalytic	USP20 147-685	pASV3mod	M-(L29 22-32)-(VP16 411-490)-GAILEKLGs		
VP16-Catalytic H643Q	USP20 147-685	pASV3mod	M-(L29 22-32)-(VP16 411-490)-GAILEKLGs		H643Q
VP16-DUSP domains	USP20 686-894	pASV3mod	M-(L29 22-32)-(VP16 411-490)-GAILE		
VP16-DUSP1	USP20 686-791	pASV3mod	M-(L29 22-32)-(VP16 411-490)-GAILE		
VP16-DUSP domains D841R	USP20 686-894	pASV3mod	M-(L29 22-32)-(VP16 411-490)-GAILE		D841R
LexA-B-arrestin-1	ARBB1 1-418	pBTM116mod	(LexA 1-202)-EFPGL		
VP16-B-arrestin-1	ARBB1 1-418	pASV3mod	M-(L29 22-32)-(VP16 411-490)-GAILE		
LexA-TRAF6	TRAF6 1-522	pBTM116mod	(LexA 1-202)-EF		
LexA-RAD17	RAD17 1-681	pBTM116mod	(LexA 1-202)-EF		
LexA-PLK1	PLK1 1-603	pBTM116mod	(LexA 1-202)-EF		
LexA-PLK1ΔPoloboxes	PLK1 1-408	pBTM116mod	(LexA 1-202)-EF		
LexA-Poloboxes	PLK1 367-603	pBTM116mod	(LexA 1-202)-EF		
LexA-PLK1ΔPolobox2	PLK1 1-508	pBTM116mod	(LexA 1-202)-EF		
LexA-Poloboxes H538A	PLK1 367-603	pBTM116mod	(LexA 1-202)-EF		H538A
LexA-Poloboxes K540M	PLK1 367-603	pBTM116mod	(LexA 1-202)-EF		K540M

Details about proteins that were used in yeast two hybrid assays are given. For modifications, single letter amino acid code is used except for where whole or partial proteins are present (shown in brackets). All cloned proteins are human orthologues.

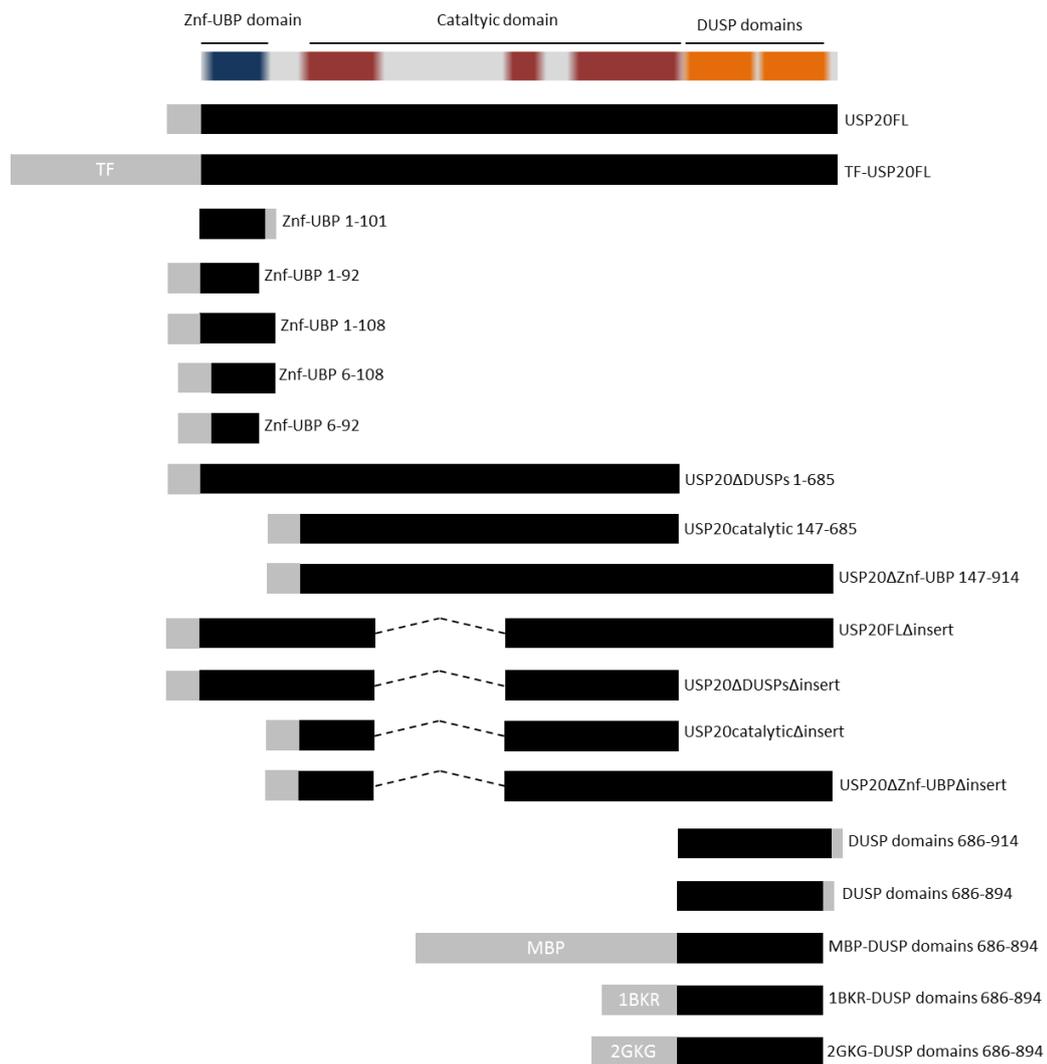


Figure 4.13. Summary of expression constructs for crystallisation. The USP20 schematic is shown at the top. The black region of each construct below represents the region of USP20 is used in reference to the schematic. The grey regions show additional sequences that are either affinity tags, restriction-site amino acids, cleavage sites and/or solubility tags. The dashed lines show spliced regions that are not present in the construct.

4.3 Purification and crystallisation of USP20 domains

4.3.1 Znf-UBP

Four constructs of Znf-UBP were expressed for crystallisation. These are shown in Figure 4.14.

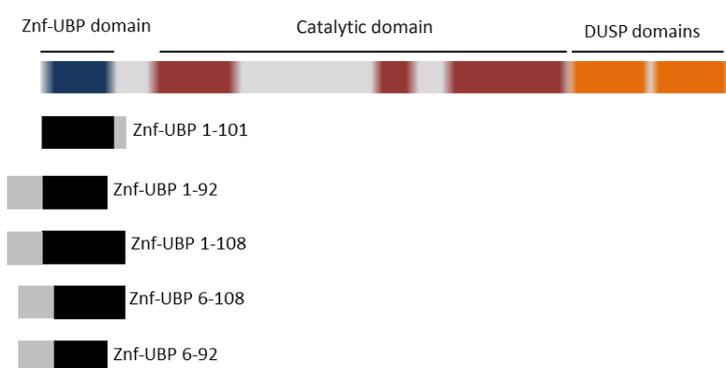


Figure 4.14. USP20 Znf-UBP domain constructs. The regions of USP20 expressed for the Znf-UBP domain are shown in black. Extra residues (His-tag, cleavage sites and restriction site residues) are shown in grey.

4.3.1.1 Znf-UBP 1-101

The Znf-UBP domain of USP20 construct (residues 1-101) had a six residue His-tag with no linker between the domain and the tag. It expressed well and there was very little protein observed in the pellet following sonication when compared to the lysate supernatant. The protein had good solubility in typical nickel and gel filtration buffers; achieving 10 mg/ml without visible signs of precipitation or aggregation.

It eluted as a single peak from the nickel column at 29.3% concentration of buffer B (160.6 mM imidazole), and was relatively pure at this point. Following this it eluted as a monodisperse peak from the HiLoad 16/600 Superdex 75 prep grade column at 79.5 ml, giving it a predicted mass of 12.4 kDa; very close to its actual mass of 11.8 kDa. The final yield of protein for the Znf-UBP 1-101 was approximately 2 mg/L of culture. This sample was used in crystallisation trials, but no crystals formed. A common occurrence in the crystal trials was the presence of skins on the drops. This may be caused by a layer of denatured protein on the top of the crystallisation drop, although the drops were still clear underneath the skin in many cases [312]. Chromatograms and SDS-PAGE gels from the purification are shown in Figure 4.15.

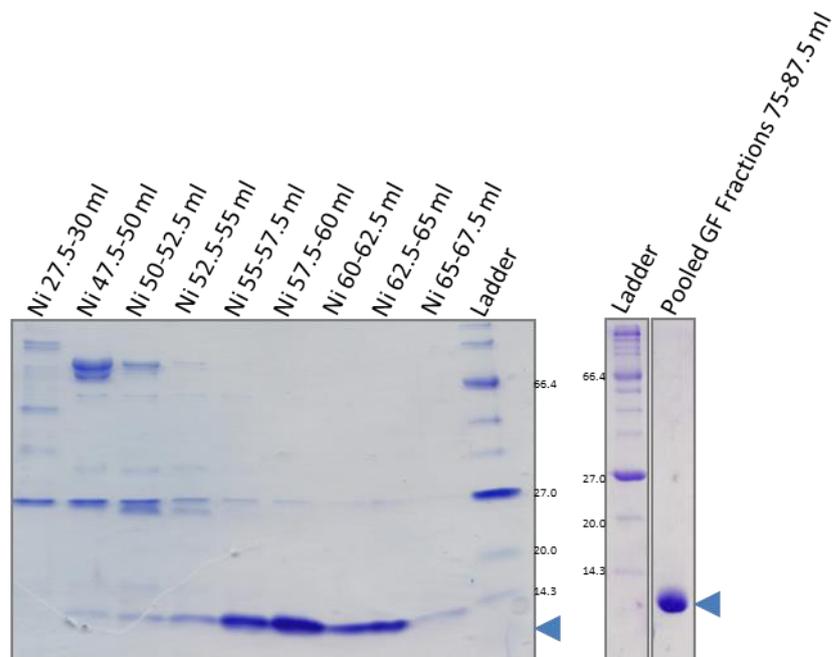
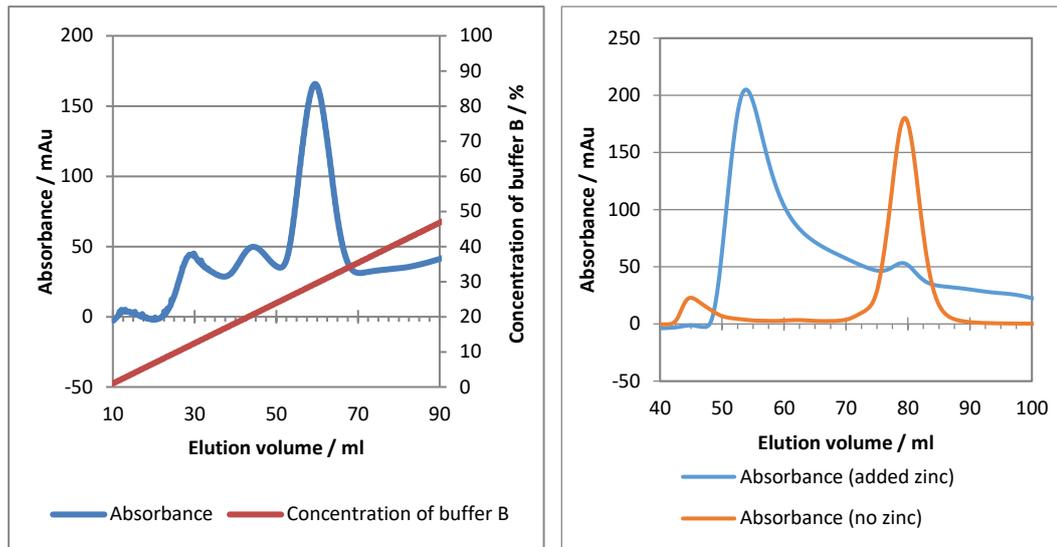


Figure 4.15. Purification of Znf-UBP 1-101. (A) Nickel column chromatogram. Curves of UV absorbance and the concentration of buffer B are shown. Multiple low percentage B peaks are visible, most likely *E. coli* contaminant proteins. (B) Gel filtration chromatogram showing the profile of elution in the presence (blue) and absence (red) of $10 \mu\text{M Zn}^{2+}$. (C) SDS-PAGE of Nickel column fractions and final sample for crystallisation. The SDS-PAGE shows that the protein is relatively pure after the nickel column. Only a few higher molecular weight proteins are observed on the gel. After the Gel filtration, the sample is very pure; no other bands are visible. The blue arrow indicates the Znf-UBP domain.

As the protein is predicted to contain three Zn^{2+} ions, like its paralogue USP33, it may be possible that the bacterial expression of this construct was unable to supply enough Zn^{2+} to fully load all sites on the protein, or that the Zn^{2+} binding is weak and therefore is being lost during the purification. Removing Zn^{2+} with EDTA massively destabilises the protein (see thermal shift assays section 4.4) so the gel filtration buffer was supplemented with 10 μ M $ZnSO_4$ in order to maintain full occupancy of the Zn^{2+} binding sites. This had an unusual effect on the profile of the gel filtration chromatogram, producing a polydisperse peak with a peak elution volume of 53.9 ml. This corresponds to a predicted molecular weight of 60.4 kDa, which is 5.1 times the molecular weight of the monomer. The reason for this shift is unclear - it could be that there is an interaction via the his-tags with the free Zn^{2+} ions; it could destabilise the bound Zn^{2+} causing transient interactions with the free ions; or it could be that Zn^{2+} ions are loading onto the Znf-UBP domain, but rather than coordinating with residues of one molecule, they become coordinated by multiple Znf-UBP domains. The two peak fractions of this higher molecular weight product were also subjected to crystallisation trials, but again no crystals formed.

4.3.1.2 **Znf-UBP 1-92, 6-92 and 6-108**

To overcome possible early termination of a helix at the C-terminus of the Znf-UBP domain, four new constructs were designed. Boundaries of USP20 residues 1-92, 6-92, 1-108 and 6-108 were chosen as they minimised the extra N- and C-terminal residues that may increase disorder in these region while still allowing complete secondary structure. In addition, the genes were cloned into a plasmid that

produced a protein with an N-terminal, cleavable His-tag in case the non-cleavable, disordered His-tag was preventing crystallisation of the Znf-UBP 1-101 construct.

Small-scale experiments were performed to quickly analyse differences in levels of soluble protein (Figure 4.16). These showed that very little protein from the 1-92, 6-92 and 6-108 constructs was expressed. However, the 1-108 expression showed an intense band at the correct molecular weight of the construct. This observation was corroborated by the full-scale expressions. The 1-92 construct was not visible after the nickel column purification so gel filtration could not be performed on this protein. The 6-92 and 6-108 constructs were marginally better than the 1-92 construct, and very similar to one another in their expression and purification. An example purification of Znf-UBP 6-108 is shown in Figure 4.17. A small amount of protein that is likely to be the 6-108 construct can be seen in the nickel fractions. Gel filtration showed a small peak at 85.3 ml with a predicted mass of 8.7 kDa. This is much lower than the actual mass of 14.5 kDa. The reason for this is unclear as it migrates to an appropriate size on SDS-PAGE and it would be expected to have a larger hydrodynamic radius than just the domain alone because of the large disordered His-tag and linker. It is possible that this construct misfolds, but forms a compact globular structure; this may also explain the low yield. The 1-92, 6-92 and 6-108 Znf-UBP constructs were not subjected to crystallisation trials due to the low yields obtained

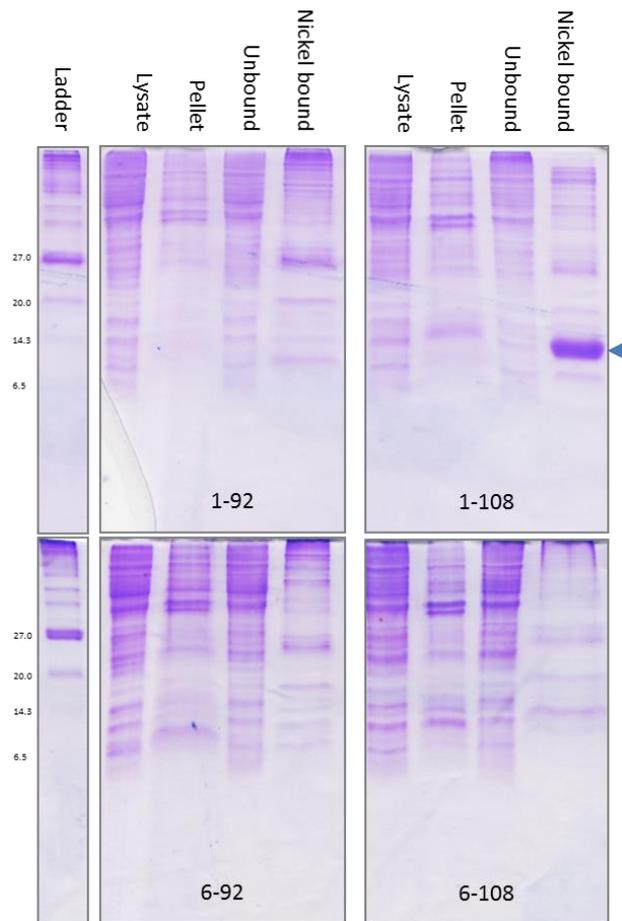


Figure 4.16. Small scale expressions of Znf-UBP domains. For each of the four Znf-UBP constructs, the lysate, pellet, unbound and 100% buffer B elution are shown. Only in the case of USP20 1-108 is an obvious band (blue arrow) found in the elution from the nickel column.

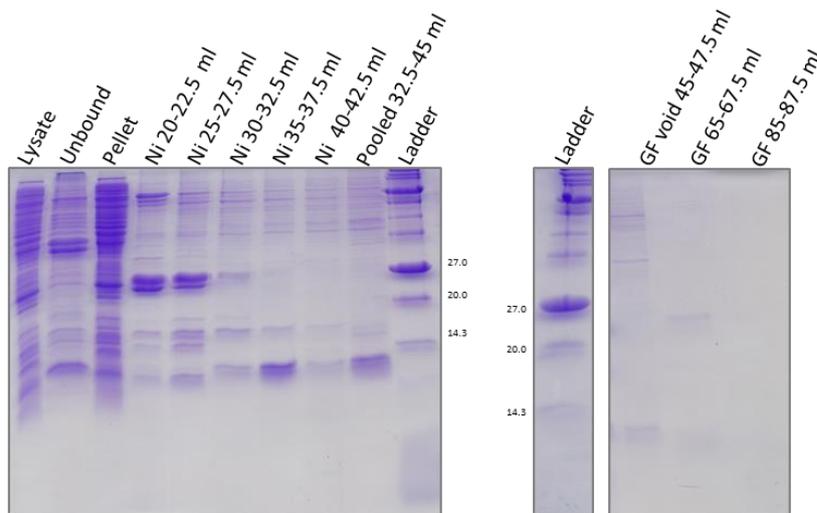
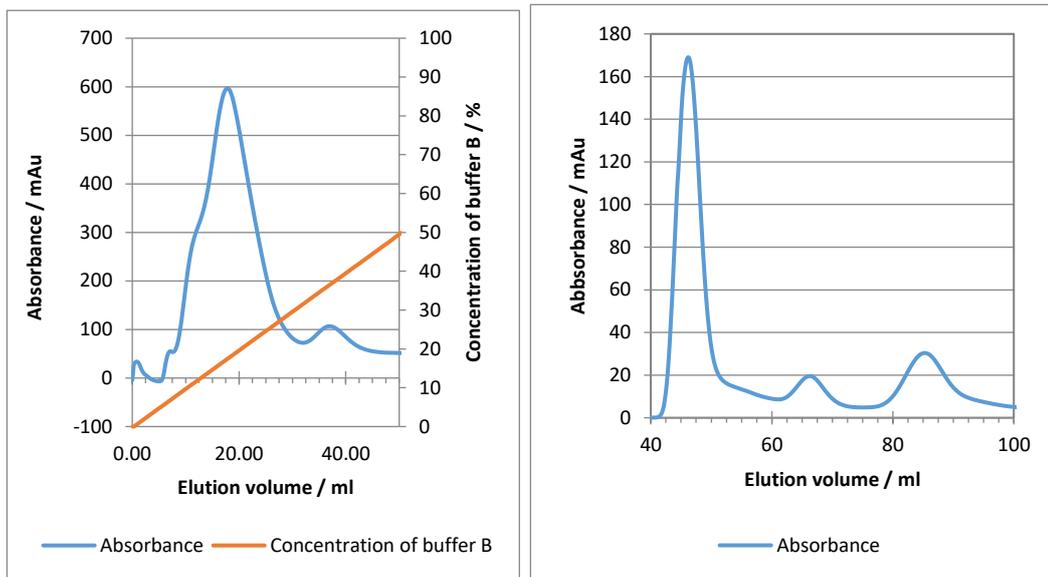


Figure 4.17. Purification of Znf-UBP 6-108. (A) Nickel column chromatogram. Curves of UV absorbance and the concentration of buffer B are shown. A large *E. coli* contaminant protein peak is observed at 17.8% of buffer B, followed by a very small peak that appears to contain the Znf-UBP domain according to SDS-PAGE. (B) Gel filtration chromatogram showing a small peak at the correct elution volume for Znf-UBP 6-108. The concentration is so low that it's hardly detectible by SDS-PAGE (C) SDS-PAGE of Nickel column fractions (left) and gel filtration fractions (right). The SDS-PAGE shows that there is a very tiny amount of the protein at the correct molecular weight, but it is also contaminated with the 20-27 kDa protein.

4.3.1.3 **Znf-UBP 1-108**

The Znf-UBP 1-108 domain expressed well. The protein required 500 mM NaCl buffers for stability, as precipitation occurred in 150 mM NaCl at higher concentrations. Following nickel affinity chromatography and gel filtration, pure, stable protein was obtained.

The protein eluted as a non-base resolved peak at 37.9 % buffer B on from the nickel. Contaminating proteins were mainly of much higher molecular weight so they could be resolved by gel filtration. The protein eluted at 81.7 ml on gel filtration, corresponding to a molecular weight of 10.8 kDa. Again this is lower than the 14.9 kDa of the actual construct, but is proportionally higher than the predicted and actual masses observed for the 6-108 construct. The His-tag was cleaved using TEV protease, which was fully cleaved after 1 hour. Following cleavage, the tag, residual uncleaved protein and the TEV protease were removed using a nickel column. The flow through was collected and concentrated. Chromatograms and SDS-PAGE gels from the purification are shown in Figures 4.18 and 4.19. The final yield of protein at the end of purification was approximately 1 mg/L of culture. The protein was then used for crystallisation trials. Crystallisation trials were performed using the uncleaved and cleaved Znf-UBP 1-108 protein, but no protein crystals were obtained. The trials looked very similar to that of the Znf-UBP trials; skins were observed in many crystallisation drops and microprecipitate was observed in around 50%. Heavy amorphous precipitate was rarely seen. Also, small salt crystals were obtained in many conditions, likely due to the high NaCl concentration in the buffers.

In summary, the Znf-UBP 1-101 was more stable and produced a slightly higher yield of protein than the 1-108 construct. Although 500 mM NaCl was used for both, it was used to increase the stability of the 1-101 construct, whereas it was a requirement for the 1-108 construct. Ultimately, no protein crystals were obtainable from either protein.

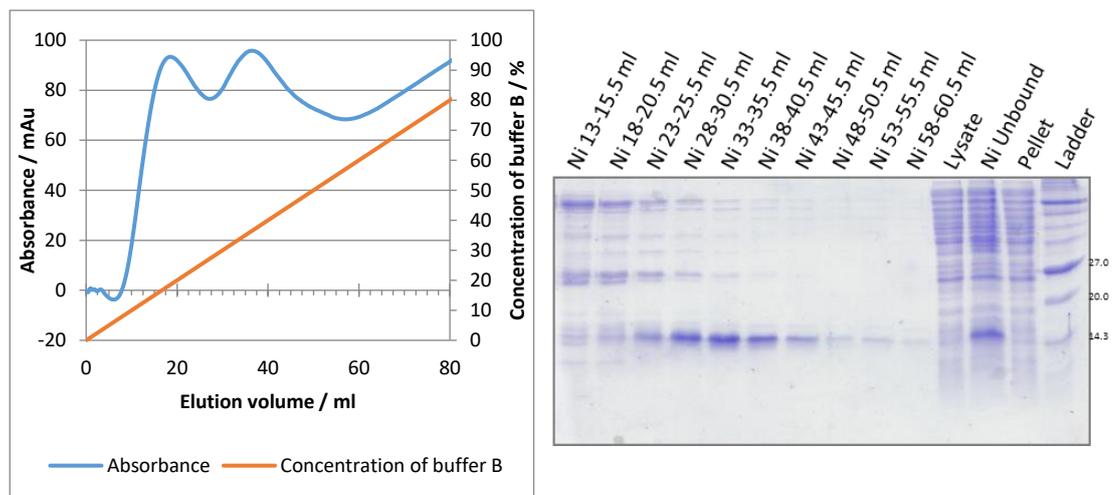


Figure 4.18. Nickel column of Znf-UBP 1-108. (A) Nickel column chromatogram. Curves of UV absorbance and the concentration of buffer B are shown. There is a peak on the chromatogram that shows produces a band at the correct molecular weight on SDS-PAGE.

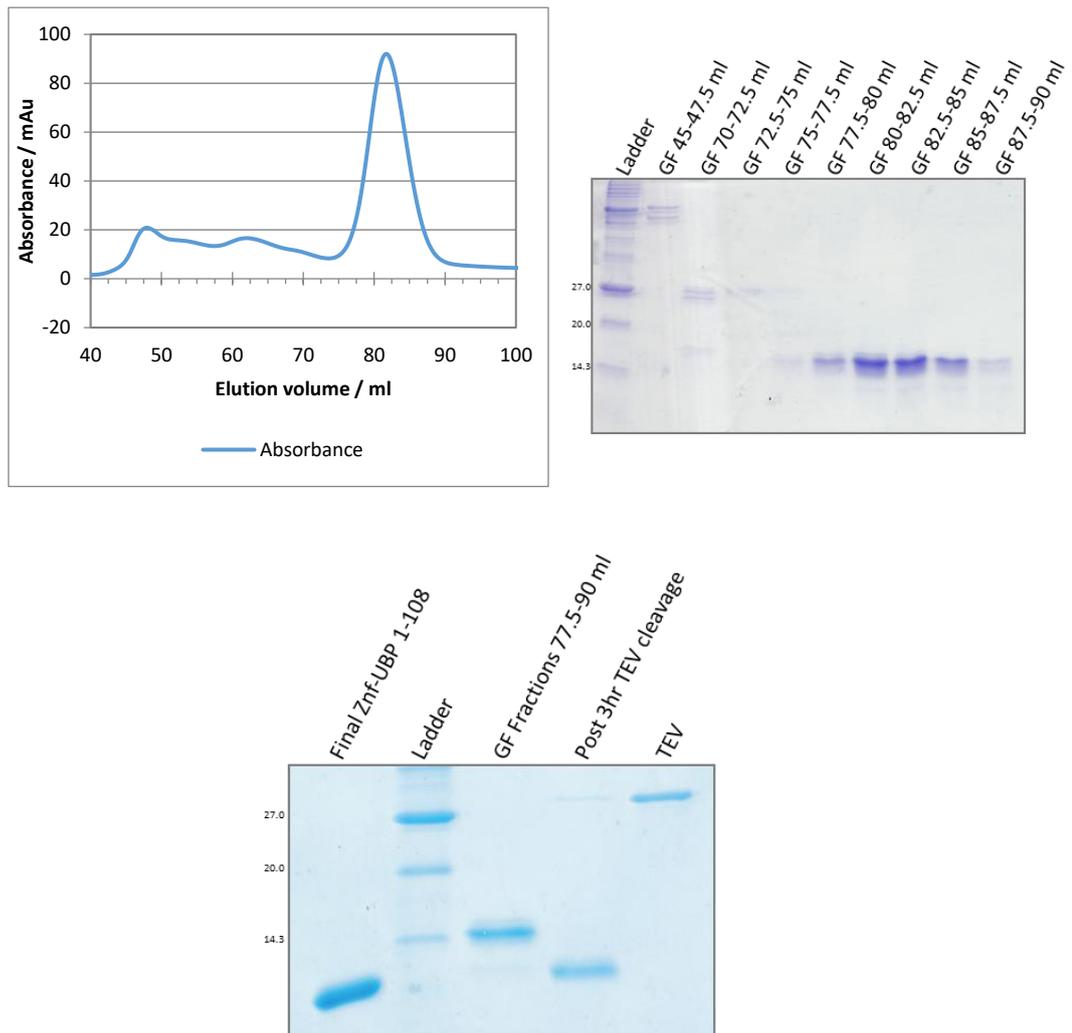


Figure 4.19. Gel filtration and cleavage of USP20 1-108. (A) Gel filtration chromatogram of the pooled nickel fractions. The trace shows a monodisperse peak at 81.7 ml, which is relatively pure on SDS-PAGE (shown in **B**). Some breakdown is observed which is most likely in the tag region of the protein (**C**) cleavage of the pooled GF fractions. Following clean-up of the cleavage reaction, the final sample is very pure and appears as a single band.

4.3.2 Catalytic domains

Nine constructs containing the catalytic domain of USP20 were expressed. These are shown in Figure 4.20.

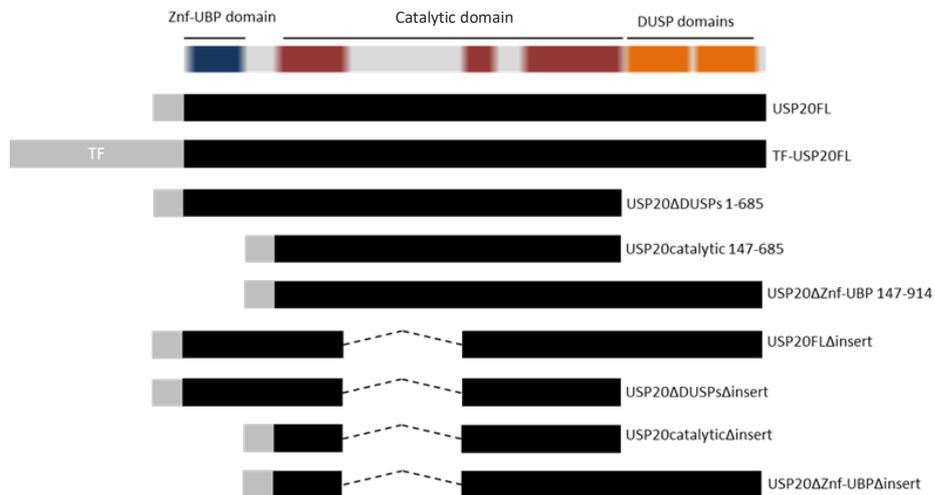


Figure 4.20. USP20 catalytic domain constructs. The regions of USP20 expressed are shown in black. Extra residues (His-tag, cleavage sites and restriction site residues) are shown in grey. Dotted lines show spliced sequences.

4.3.2.1 USP20FL

The full length USP20 enzyme was expressed as an N-terminally His-tagged pCOLD1 construct. It was thought that the cold expression vector would be beneficial to the folding of the domain, as this had been used successfully in the lab previously with other full length USPs. Following sonication, the protein is visible in the lysate supernatant on SDS-PAGE, and is not observed as an intense band in the pellet indicating that the protein does not form inclusion bodies. The protein elutes from the nickel column at 45.6 % buffer B (211.5 mM imidazole). The total mass of protein from the nickel column was 5.5 mg, which is good for a

protein of this size. However, the protein eluted in the void volume from gel filtration, suggesting that the protein is aggregated. In addition, there is an intense band that co-elutes with the full length protein in all fractions. This is likely a contaminant protein from *E. coli*, such as GroEL or DnaK due to the nature of the aggregated protein, but could also be some form of breakdown product.

The aggregation and the presence of the extra band was not alleviated by using high salt (500 mM NaCl) but the aggregation was partially reversible by the addition of DTT to the sample prior to gel filtration. Upon addition of DTT, a peak at 75.4 ml was obtained, corresponding to a molecular weight of 80.1 kDa; 3.1 ml later than the predicted elution volume of the 104 kDa construct. SDS-PAGE analysis shows that the protein in the peak contains the full length protein, but also still contains the 55-66 kDa protein that is observed from the nickel column. Also, it shows that the protein is still very impure, containing the same ladder of bands as the void samples and pre-gel filtration samples do. This peak and the void sample was taken and concentrated for use in activity assays to ensure that the full length protein is properly folded and active.

As the aggregation of the protein could be caused during expression, variables of the expression were optimised to improve yield of non-aggregated protein. Typically, one of the first parameters to optimise for better expression is the IPTG induction concentration. It is thought that large rates of translation could lead to aggregation of partially folded proteins. Reducing the rate by lowering the IPTG

concentration could reduce the concentration of folding protein, allowing the proteins to fold properly without aggregation [313, 314]. The concentration was altered from 0.5 mM IPTG to 0.2 mM, but no increase of non-void protein was observed. As the protein was already being expressed in a cold-expression system at 10 °C, this was not altered. In addition, a routine procedure for full length USP purification in the laboratory is to co-express the protein with ubiquitin. With USP4, USP11 and USP15, this produces a complex and seems to stabilise the protein, improving the yield of active protein. However, the expression of USP20 alone, or with co-expression of ubiquitin did not reduce the amount of void protein. Arctic express cells were also used to increase monomeric protein expression. The arctic express cells produce cold-adapted chaperonins from *Oleispira antarctica* that should aid in the folding of the recombinant protein. However, the profile of the gel filtration was almost identical to that of the BL21 codon plus expression; with no difference in the ratio of void to monomer. Chromatograms and SDS-PAGE gels from the purification are shown in Figure 4.21.

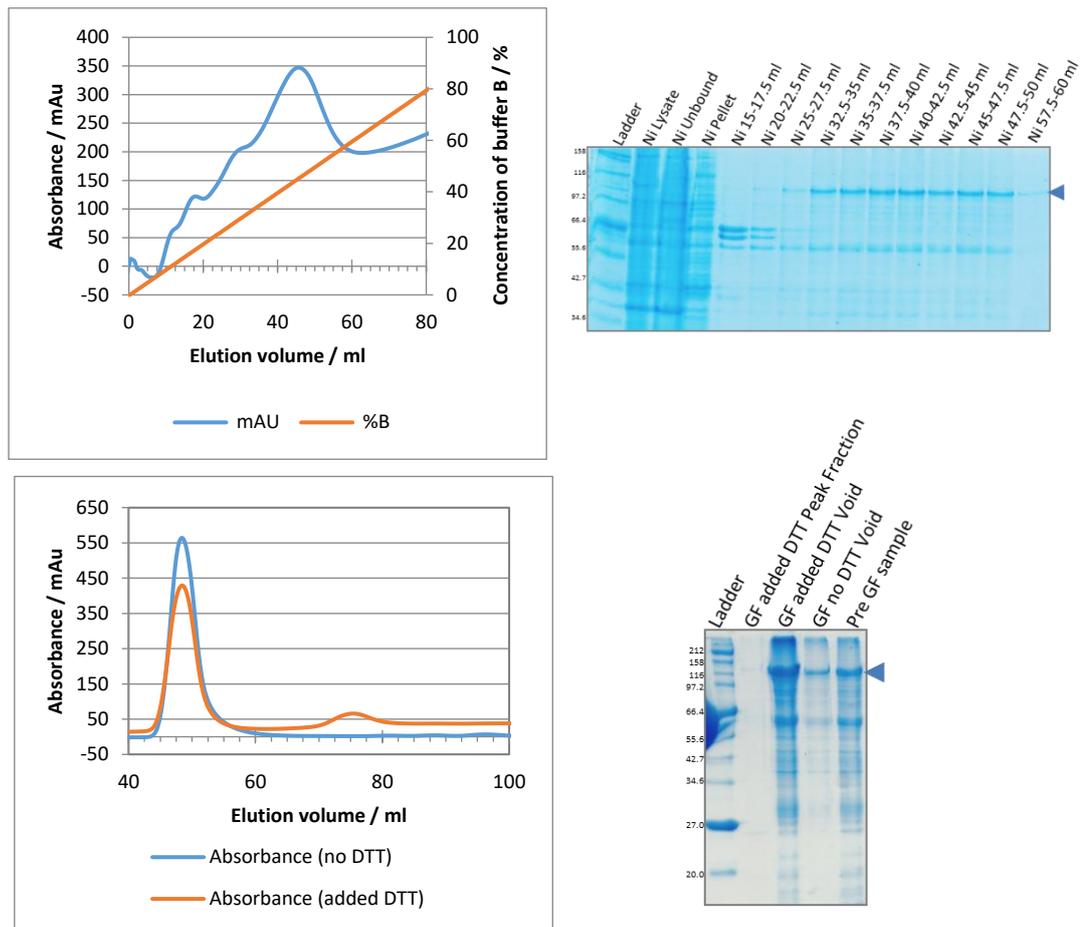


Figure 4.21. Purification of USP20FL. (A) Nickel column chromatogram. Curves of UV absorbance and the concentration of buffer B are shown. A peak at 45.6% of buffer B was obtained, which was spread over a large number of fractions. **(B)** The SDS-PAGE shows a major band at the correct molecular weight for USP20FL. However, there are also many other bands present, even though the protein eluted at a relatively high concentration of imidazole. **(C)** The gel filtration chromatogram shows the profile of USP20 elution in the presence (red) and absence (blue) of 1 mM DTT. **(D)** The SDS-PAGE shows that both the void volume sample, and the peak at 75.4 ml contain a high number of additional proteins/breakdown products, and was no more pure than prior to gel filtration.

There are multiple assays available for assessing the activity of deubiquitinating enzymes, which include fluorescence and gel-based assays. To analyse the activity of USP20, ubiquitin-7-amido-4-methylcoumarin (Ub-AMC) substrate was used. The Ub-AMC assay measures the cleavage of ubiquitin from a C-terminally bound AMC molecule. This produces an increase in fluorescence at 460 nm when excited at 380 nm. The Ub-AMC assay essentially analyses the ability of the DUB to cleave a peptide bond as the AMC is bound to the C-terminus. USP20 has low activity when cleaving linear-, K6- and K27-linked chains, however cleavage has been observed for linear chains [313], which means that USP20 should, in principle, cleave Ub-AMC.

The void and 'monomer' peaks of USP20FL were tested for activity using Ub-AMC assays. Activity of most USPs can be detected within 30 mins at 50-100 nM enzyme concentrations in the assay. For the USP20 assays no activity was detected after 1 hour at 100 nM for either gel filtration peak (not shown). As USP20 has weak activity with linear chains, or due to the contaminant protein reduced the effective USP20 concentration in solution, the assay was repeated with higher protein concentrations and a longer assay time. No activity was also detected for the 'monomer' peak at 500 nM within 16 hours (Figure 4.22). However, a low activity was detected for the protein in the void volume at 500 nM (Figure 4.22), suggesting that, although aggregation is occurring, at least some of the protein is properly folded and able to cleave ubiquitin. Due to the lack of activity and impurity of the monomer peak, and the aggregation of the void, pCOLD1 FL-USP20 could not be used for enzyme kinetic assays or crystallisation.

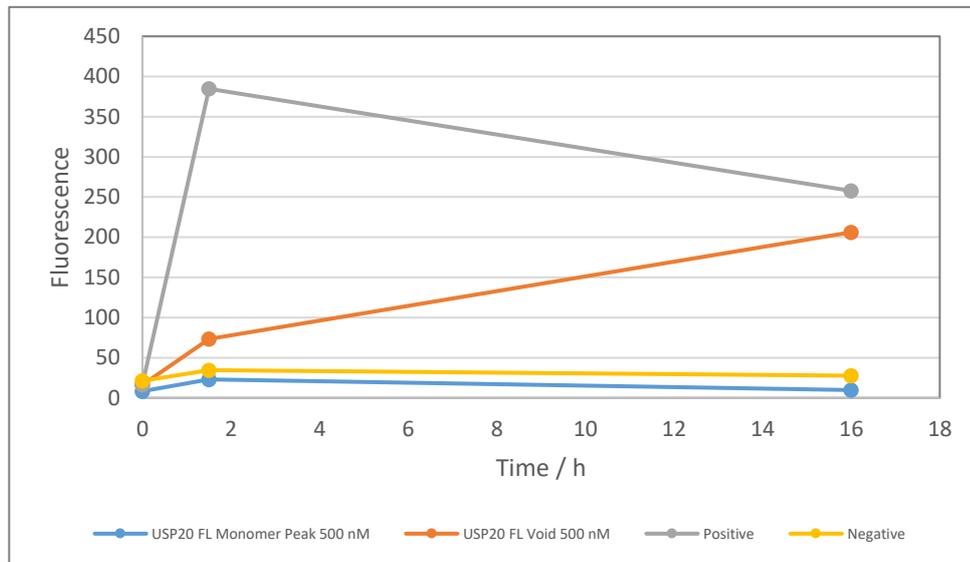


Figure 4.22. USP20FL Ub-AMC assay. The graph shows an increase in fluorescence over time for the positive control and the void volume only. The peak at 75.4 ml from the gel filtration column shows no increase in fluorescence as does the negative control. The positive control achieves the maximum fluorescence after 1 hour whereas it takes the void volume 16 hours to achieve a similar level of fluorescence.

As the pCOLD1 tag is not designed for increasing solubility, other pCOLD vectors can be used to produce fusions with solubility tags. pCOLD-GST was avoided because its use in the expression of USP33 did not produce active, soluble, monomeric protein (data not shown, work performed by Ibrahim Nur). Instead, pCOLD-TF was used, which produces a fusion with *E. coli* trigger factor (TF); a ribosome-associated molecular chaperone that not only acts a solubility tag, it also aids in protein folding.

Following sonication, a band at the correct molecular weight was visible in the lysate supernatant, unbound and pellet (full length protein is 154.1 kDa). This suggests that there is likely some misfolding of the protein, which possibly forms

aggregates and inclusion bodies. This is likely due to the high expression levels caused by the TF, which has been observed with other proteins in the laboratory. However, a large amount of protein is in the supernatant, but the SDS-PAGE shows that there are a series of bands that may correspond to breakdown products or contaminant proteins. There is a particularly intense set of breakdown bands at 55-66 kDa. The nickel column elution profile shows a double peak where SDS-PAGE analysis shows that the lower percentage B peak corresponds to the 55-66 kDa breakdown products (24.7% buffer B; 138.6 mM imidazole) and the later peak corresponds to the full length protein (33.6% buffer B; 181.3 mM imidazole). Due to the intensity of these bands, and the fact that the N-terminus is His-tagged, these are most likely breakdown products that include the TF protein and a small portion of USP20 (TF is 48 kDa), which is again routinely observed with TF fusions in the laboratory.

Anion exchange was used to separate the protein in the pooled fractions spanning 80-110 ml, which achieved separation of a large amount of the 55-66 kDa breakdown products. A large peak was obtained at 88.5% buffer B (442.5 mM NaCl), which appears to contain the largest quantity of full length TF-USP20 according to SDS-PAGE analysis. This peak was concentrated (80-110 ml) and run on a gel filtration Superdex 16/60 200 prep grade column to separate the remaining proteins. The chromatogram shows three discreet peaks; one for the full length, one for the remaining 55-66 kDa breakdown products and one for a smaller 27 kDa product. Interestingly, compared to the pCOLD1 expressed USP20

FL, no void peak was observed, which may be due to the chaperone function of the trigger factor aiding the folding of USP20.

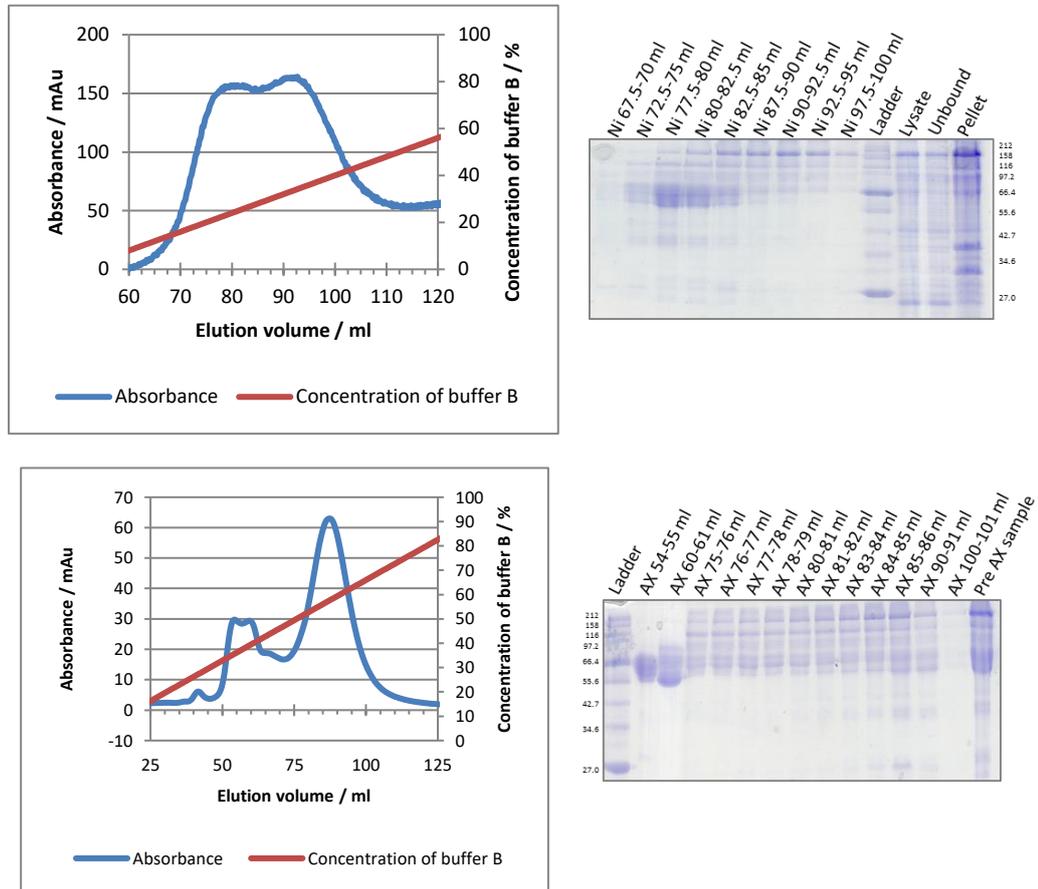


Figure 4.23. Nickel column and anion exchange of TF-USP20FL. (A) Nickel column chromatogram. Curves of UV absorbance and the concentration of buffer B are shown. A double peak is observed. The full length protein is found in the 33.6% buffer B peak. **(B)** SDS-PAGE of the nickel column. TF-USP20FL and the breakdown products can be observed in the overlapping peaks. TF-USP20FL is observed at 154 kDa and the later eluting breakdown products at 55-66 kDa. **(C)** Anion exchange of the pooled nickel fractions. Curves of UV absorbance and the concentration of buffer B are shown. Multiple peaks are observed from the gradient elution **(D)** SDS-PAGE shows that most of the full length protein is found in the later peak of anion exchange. However, there are many lower bands that could either be breakdown products or *E. coli* contaminants.

The large broad peak at 53.6 ml contains the full length protein. This elution volume corresponds to a size of 514.7 kDa and suggests some form of oligomerisation, micro-aggregation or complex formation. This is 3.3 times the size of the expected protein, which should elute at ~66.8 ml. The SDS-PAGE shows a series of lower molecular weight bands in lanes containing the full length protein. Again, these could be the folded TF-USP20 fusion with cleaved loops, or they could be contaminant proteins from *E. coli*. The issue is also confounded by the fact that TF can form homodimers and could lead to dimerization of the USP20 fusion. A 308.2 kDa dimer would not be far from the elution volume of the product observed on gel filtration.

The 55-66 kDa break down products elute at ~80.1 ml, which corresponds to a predicted molecular weight of around 54 kDa. The 27 kDa product elutes at 90 ml, with a predicted weight of 23.2 kDa. Therefore, these breakdown products are monomeric in solution, and are most likely products containing the N-terminus of the TF fusion. If the 55-66 kDa proteins are N-terminal cleavage products then cleavage would be occurring within the Znf-UBP domain, where the cleavage products would be between 52.8 and 63.3 kDa. This could suggest that the presence of TF is affecting folding of the Znf-UBP domain.

The 52.5-55 ml fraction was used for Ub-AMC activity assays. The assays were performed at 75 nM enzyme concentration. After 1 hour, both the positive control and TF-USP20FL showed moderate activity. After 16 hours, both were well above the background level of the negative control. This indicates that the USP20 TF fusion has activity. Fractions from 47.5-57.5 ml were pooled and concentrated as

these showed a consistent set of bands on SDS-PAGE, indicating that they were homogeneous. Later fractions containing the break down products of varying size were avoided. Chromatograms, SDS-PAGE gels and activity assays from the purification are shown in Figures 4.23 and 4.24.

Although it was unlikely that this protein preparation would form crystals, three crystal trials (JCSG+, PACT suite and Morpheus) were attempted with this protein at 4 and 2 mg/ml. Initially, around 50% of the drops showed precipitate. However, after one week, nearly all drops showed heavy precipitate. No crystals were obtained from these trials.

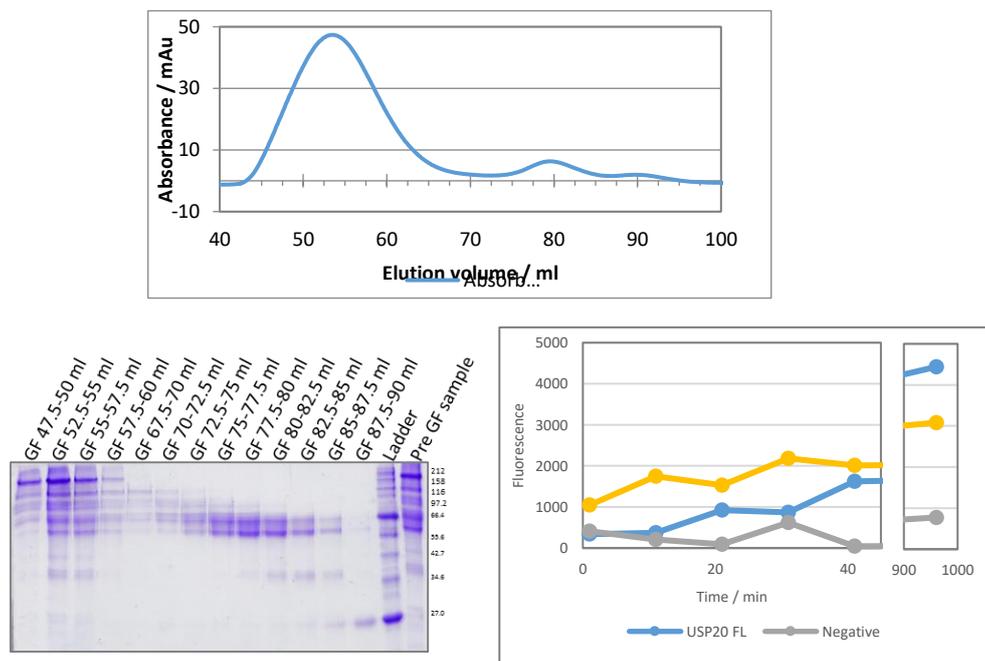


Figure 4.24. TF-USP20FL gel filtration and Ub-AMC assays. (A) Gel filtration chromatogram showing three discrete peaks. **(B)** SDS-PAGE of the gel filtration fractions. Clearly TF-USP20FL is found in the earlier eluting peak, with breakdown products of decreasing size following. **(C)** Ub-AMC assay of TF-USP20FL. A slow increase in fluorescence is observed over time. After 16 hours the positive control and the full length protein shows that cleavage of Ub-AMC occurred.

4.3.2.2 **None FL catalytic domain-containing constructs**

Seven non-FL constructs were made that included the catalytic domain. Three contained the full catalytic domain; USP20 Δ DUSPs, USP20Catalytic and USP20 Δ Znf-UBP. Four constructs that contained a catalytic domain with a spliced 182 residue disordered, intervening sequence were made: USP20FL Δ insert, USP20 Δ DUSPs Δ insert, USP20Catalytic Δ insert and USP20 Δ Znf-UBP Δ insert. The purpose of this was to produce a construct that expressed well and had ubiquitin C-terminal hydrolase activity so that enzyme kinetics and/or the structure of the catalytic and accessory domains could be obtained. All of these constructs were coexpressed with ubiquitin as this has been observed to increase the stability of other full length USP enzymes in our laboratory.

Small scale expressions showed that soluble protein was present in the supernatant following sonication of cell pellets from all construct expressions. However, the amount varies greatly. The best expressing construct in these small scale experiments is the full length USP20 (Figure 4.25). However, this construct contained the most visible breakdown products in the soluble fraction. The other constructs seemed to produce less soluble protein, but much cleaner Western blots. In all cases there is a lot of protein observed in the pellet by SDS-PAGE and Western blot. With all constructs producing soluble protein on small scale expression (200 ml), all were expressed as full scale expressions (5 L) for crystallisation and activity assays.

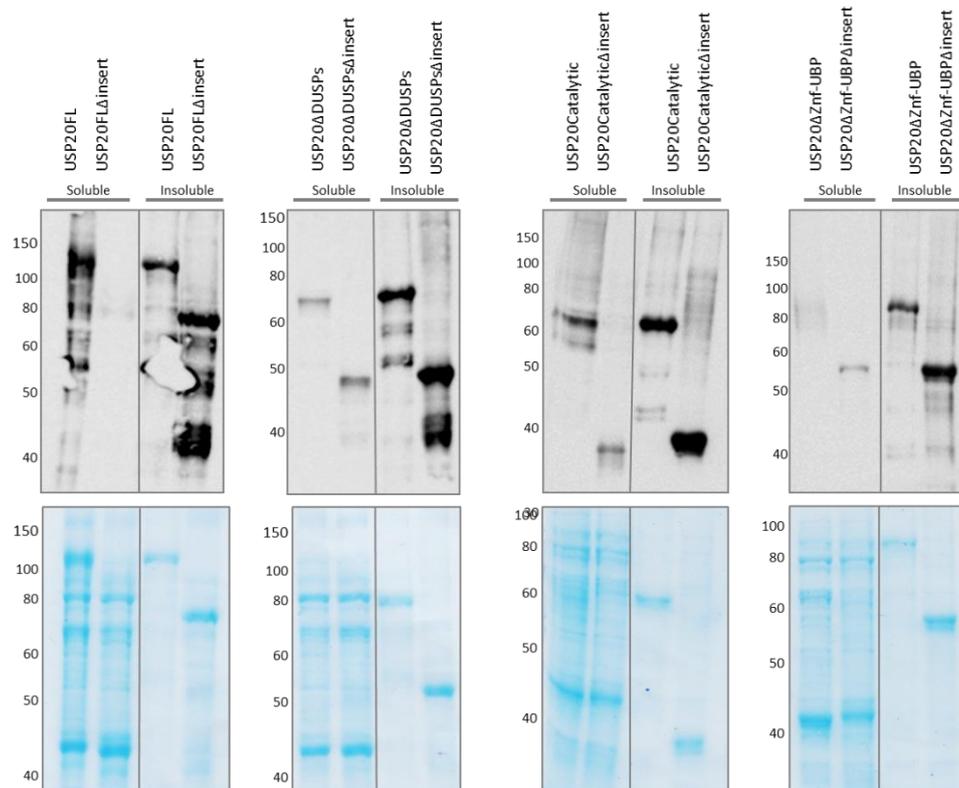


Figure 4.25. Small scale catalytic domain-containing construct expressions. Western blots (top) and SDS-PAGE (bottom) are shown for small scale expressions of all eight catalytic domain-containing USP20 constructs. For each construct, the soluble (sonication supernatant) and insoluble (sonication pellet) are shown.

5 litre cultures of 2YT were used for the 72 hour pCOLD1 expressions of all constructs. As DTT produced a peak for the monomeric protein for USP20FL, β -mercaptoethanol was used in the nickel column purification buffers for all other catalytic constructs. Although in most cases reasonable expression of the full length construct was obtained, in all cases, except the USP20 Δ Znf-UBP, protein was observed in the void volume on gel filtration. A large quantity of protein in the pellet is observed in all cases, indicating difficulty in expression of the USP20 catalytic domain in *E. coli*.

An example purification of the FL Δ insert is shown in Figure 4.26. This protein expressed well but all protein eluted in the void during gel filtration. Ub-AMC assays were performed on these void fractions to analyse whether the protein had any activity as USP20FL did, but no deubiquitinating activity was observed. The exact yield is difficult to determine, due to the fact that pure mono-disperse protein is unobtainable. However, following elution from the nickel column, the total mass of protein is approximately 4 mg: 0.5 mg/L of culture.

The USP20 Δ DUSPs, USP20 Δ DUSPs Δ insert, USP20Catalytic, USP20Catalytic Δ insert and USP20 Δ Znf-UBP Δ insert all expressed and purified in a very similar manner to that of the full length. For brevity, the chromatograms and SDS-PAGE are not shown. In all of them, protein was obtained from the nickel column at similar imidazole concentrations. They were all run on gel filtration, and all showed elution in the void volume, with no peak for the monomeric protein.

The USP20 Δ Znf-UBP construct was very different to the rest of the catalytic construct purifications. There was very little protein in the void. However, the major peak eluted at 88.7 ml, which corresponds to a band at approximately 27 kDa on SDS-PAGE. There is also a band at 74 ml, which should be the correct peak for the construct (actual MW, 88 kDa; predicted MW, 89 kDa). As the breakdown eluted from the nickel column, it is again likely to be an N-terminal breakdown, rather than a DUSP domains breakdown product. In this case it is likely that the cleavage product is the N-terminal portion of the catalytic domain, and the cleavage has occurred somewhere within the large insert. Breakdown in this region could produce a product between 13.9 and 33.4 kDa, based on protein size

fragments from the N-terminus to the start and end of the insert region. However, the activity assays showed no activity for the full construct peak, or the breakdown peak. Again due to the lack of activity, the USP20 Δ Znf-UBP peak was not used for crystallisation trials. Chromatograms and SDS-PAGE gels from the USP20 Δ Znf-UBP purification are shown in Figure 4.15.

A summary of the expression of each catalytic domain-containing construct is shown in Table 3.5.

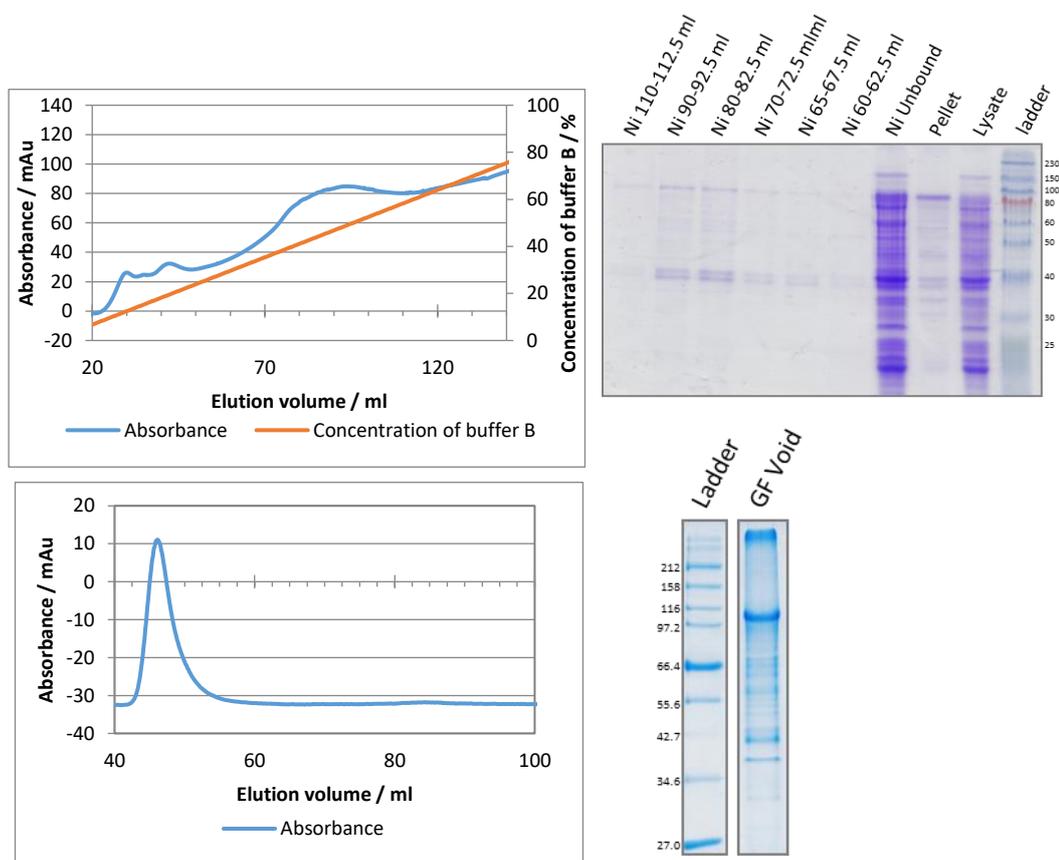


Figure 4.26. USP20FL Δ insert purification. (A) Nickel chromatogram and SDS-PAGE of nickel column fractions. The full length protein elutes as an elongated peak at 48% B. **(B)** SDS-PAGE of the nickel fractions. USP20FL Δ insert can be seen in the fractions of the peak. Also, a 40 kDa doublet band is observed. **(C)** The Gel filtration shows a large void peak, with no clear peak at the predicted elution volume for USP20FL Δ insert. **(D)** The SDS-PAGE of the void fractions shows that the USP20FL Δ insert protein is present.

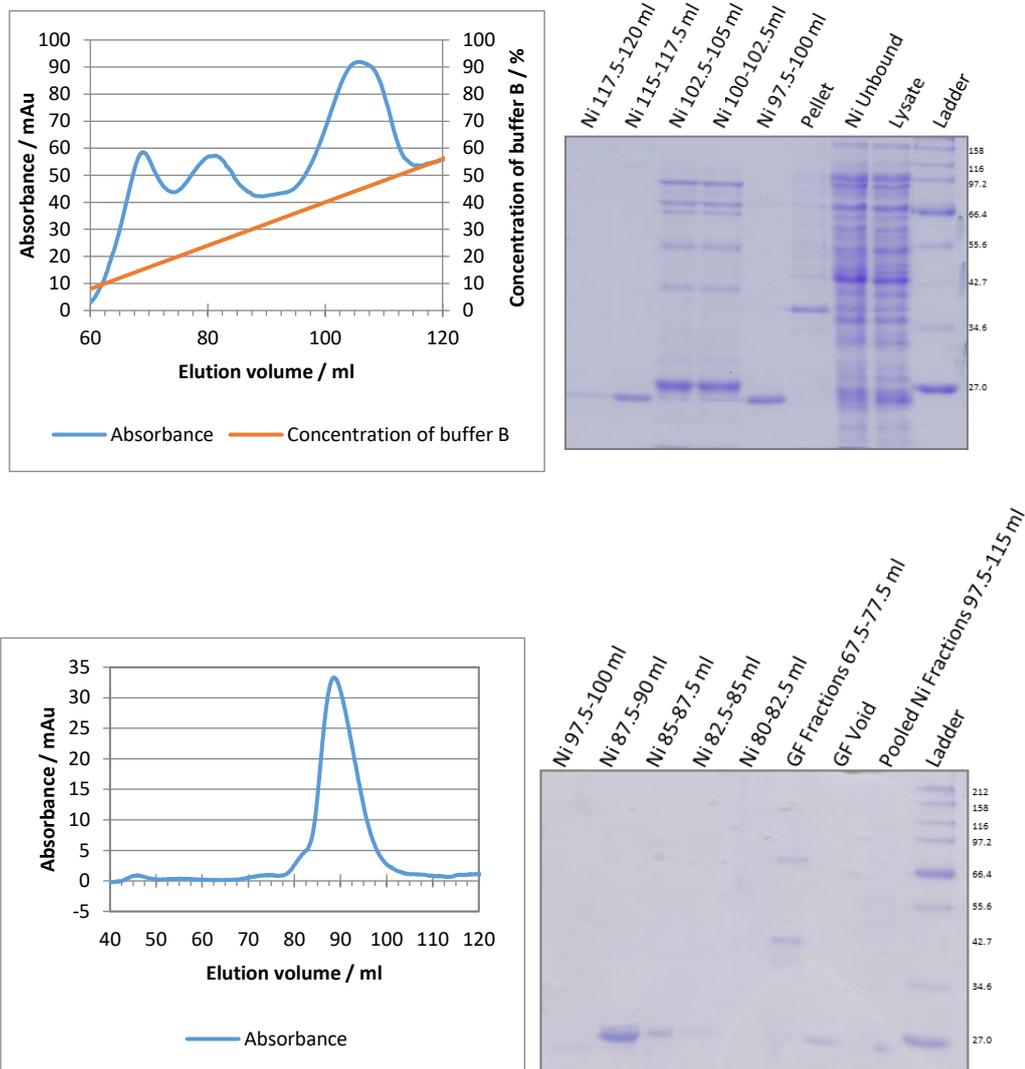


Figure 4.27. Purification of USP20 Δ Znf-UBP. (A) Nickel chromatogram and SDS-PAGE of nickel column fractions. The SDS-PAGE shows massive breakdown of the Cat-DUSPs construct. It is not obvious whether the full length protein is present. **(B)** Gel filtration chromatogram and SDS-PAGE. There is a peak which appears to contain a band that could be the full length protein; however, it also contains breakdown products. The major peak of the gel filtration is produced by a 27 kDa protein, which is most likely a His-tagged N-terminal breakdown product due to its elution from the nickel column.

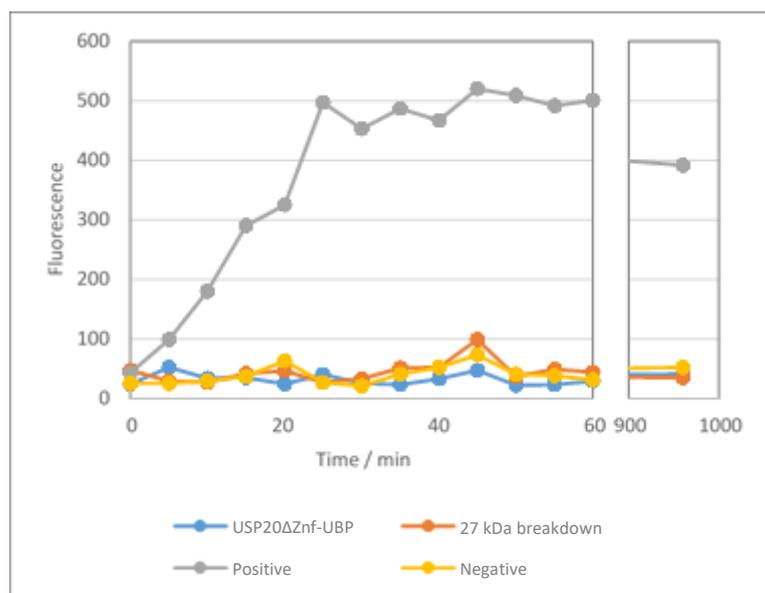


Figure 4.28. USP20ΔZnf-UBP Ub-AMC assay. The graph shows an increase in fluorescence over time for the positive control only. Neither the fractions containing the possible full length or the 27 kDa breakdown product show activity, even after 16 hours.

4.3.2.3 Summary of catalytic domain-containing constructs

A summary table of observed characteristics from all catalytic domain-containing constructs is given in Table 4.5. The approximate mass of full length protein is shown as determined by visual inspection of the SDS-PAGE of the nickel fractions. It also describes the main characteristics of the gel filtration and what information the SDS-PAGE gave about the gel filtration peaks. A summary of all the activity assays is given with a tick if activity was observed and a cross when no activity was seen.

Table 4.5. Summary of catalytic domain-containing constructs

Construct	Approximate mass of FL from nickel column	Gel filtration	SDS-PAGE	Activity
USP20 FL	+++	Monomer	Impure	Void✓ Monomer X
TF-USP20 FL	++++	Possible Dimer	Possible breakdown	Dimer✓
Znf-UBP+Cat	+++	Void	Impure	X
Cat	++	Void	Impure	X
Cat+DUSPs	+	Monomer	Monomer shows breakdown	X
FL USP20Δinsert	++	Void	Impure	X
Znf-UBP+Cat Δinsert	++	Void	Impure	X
CatΔinsert	++	Void	Impure	X
Cat+DUSPsΔinsert	++	Void	Impure	X

The number of '+' in the mass column indicates the relative amount of protein obtained from the nickel column. A tick or a cross indicates whether activity was found in following gel filtration.

4.3.3 DUSP domains

Two His-tagged constructs of the USP20 DUSP domains were expressed. These are shown in Figure 4.29.

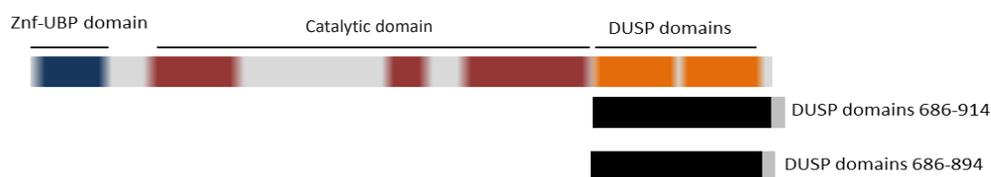


Figure 4.29. USP20 DUSP domains constructs. The regions of USP20 expressed are shown in black. Extra residues (His-tag and restriction site residues) are shown in grey.

4.3.3.1 USP20 DUSP domains 686-914 and 686-894

Two C-terminally His-tagged USP20 DUSP domains constructs were cloned: USP20 686-894 and 686-914. Both showed good expression, producing a good initial yield following nickel column purification. However, aggregation of both constructs was observed during concentration using centrifugal spin concentrators. The aggregation formed in the concentrator as a visible, fibrous solid, and the maximum concentration of the constructs in a typical nickel column or gel filtration buffer (150-300 mM NaCl, Tris pH 7.5) was approximately 1 mg/ml. At this limit, continual protein aggregation was observed and no further increase in concentration could be obtained. This caused major issues for any down-stream requirement of the protein, especially for crystallisation.

The expression and purification of the longer 686-914 construct is shown in Figure 4.30. Its elution profile from the nickel column (50.5%; 262.3 mM imidazole) showed a shouldered peak. 262 mM imidazole is a relatively high concentration for many proteins to elute at. As the histidine content of USP20 is not noticeably elevated, and there are no runs of consecutive histidines within the DUSP domains, it must mean that the His-tag is well exposed for binding to the nickel column. 40-57.5 ml fractions were pooled and concentrated before running on the gel filtration. The gel filtration showed a large amount of protein in the void volume compared to the peak of the monomer (68 ml elution = 24.9 kDa compared to 27.2 kDa actual MW). The protein in the void is possibly from the

shoulder peak, as is often the case with nickel columns producing a double peak; the early eluted peak elutes at the desired elution volume, whereas the later peak elutes in the void. Without the addition of DTT, the void peak was higher than the monomer peak (data not shown), so β -mercaptoethanol was used in the nickel column buffers of both DUSP domains constructs following this. Fractions corresponding to 65-80 ml were pooled and the resulting protein is very pure.

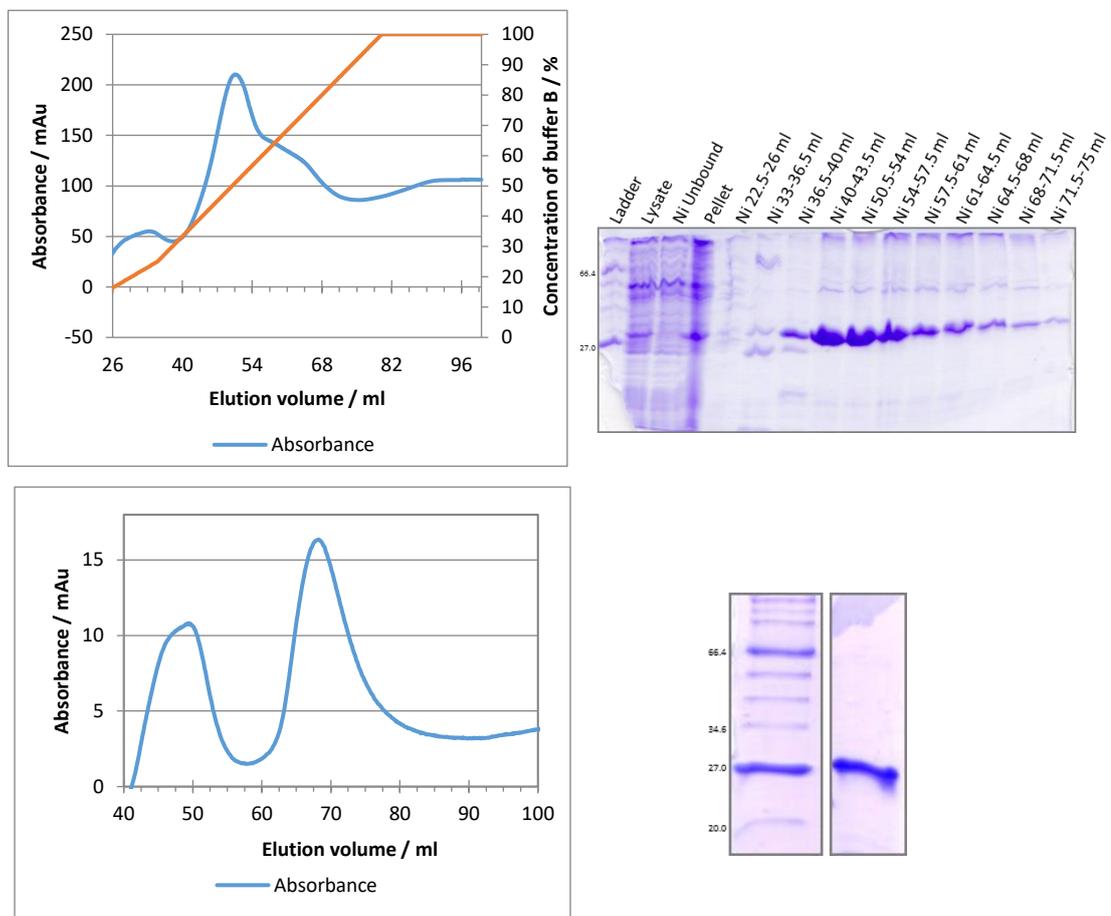


Figure 4.30. Purification of USP20 686-914. (A) Nickel chromatogram and SDS-PAGE of nickel column fractions. The chromatogram shows a shouldered peak. SDS-PAGE shows that the whole peak contains the USP20 686-914 construct. **(B)** The gel filtration shows a void and a single peak at 68.2 ml. SDS-PAGE shows that pooled fractions from this peak provide very pure protein with a single band corresponding to the 27.2 kDa protein.

In comparison, the shorter construct eluted as a single peak from the nickel column (52.5%; 271.9 mM imidazole), and as a tailed peak from the gel filtration (67.5 ml = 26.1 kDa compared to 25.0 kDa actual MW) (Figure 4.31). Again the elution imidazole concentration is high, suggesting a similar His-tag configuration to the longer DUSPs construct. The trailed peak was observed with all optimisations of the purification (explained below). As no smaller protein products are observed on SDS-PAGE across the tail of the peak, it suggests that the DUSPs could be interacting with the gel filtration column [315], which may cause the peak to trail from its expected elution volume. In some cases, a double band was observed on SDS-PAGE for the final purified product of the 686-894 construct. This band did not appear to affect the solubility of the protein.

Using the calibration of the Superdex 75 column, the sizes of the two DUSP domains constructs appear to be reversed on the gel filtration; the longer construct elutes later, suggesting a smaller hydrodynamic radius. This could be a true observation, or may be due to slight differences in the FPLC machine setup or sample (for example, extra tubing to connect the gel filtration column, or a larger sample volume for the smaller construct). It is most likely a non-significant variation produced by the procedure.

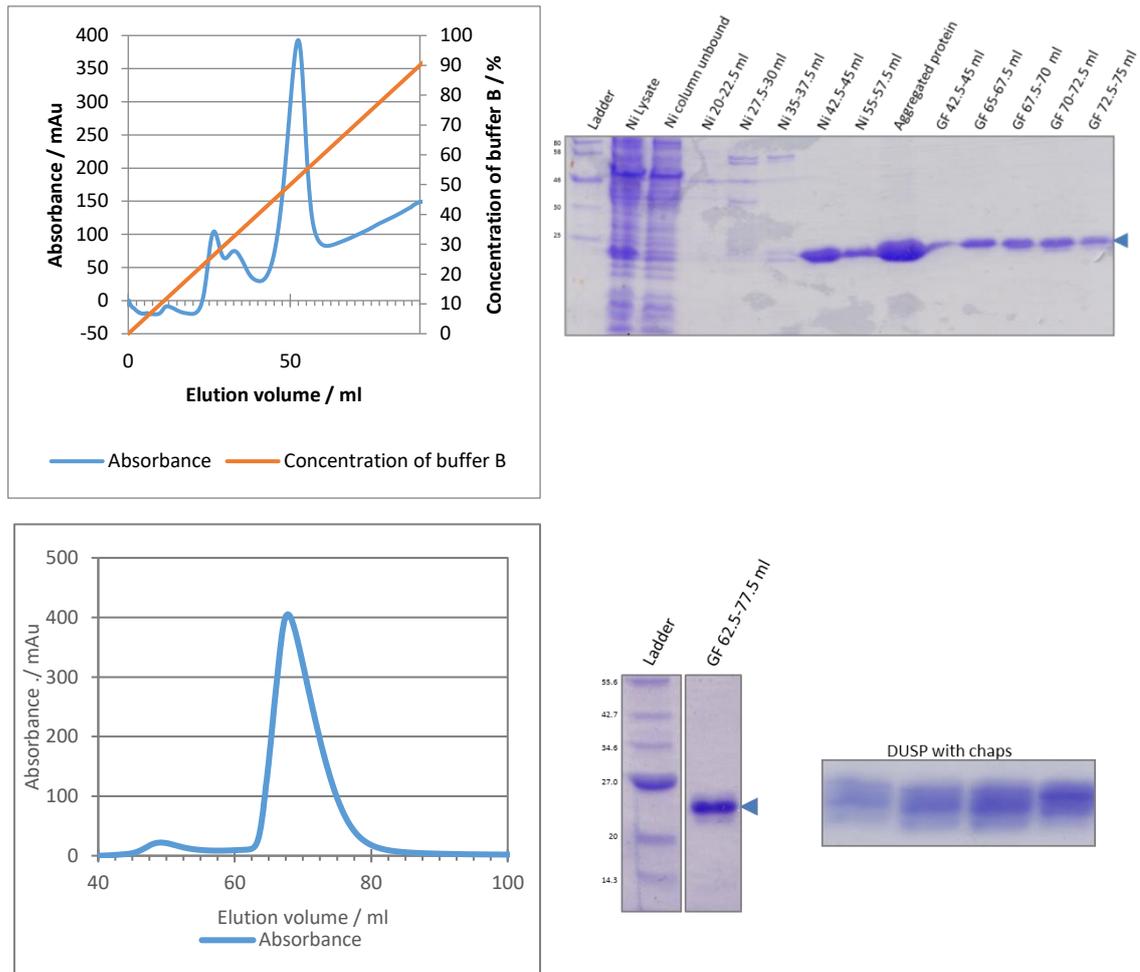


Figure 4.31. Purification of USP20 686-894. (A) Nickel chromatogram and SDS-PAGE of nickel column fractions. The chromatogram shows a single peak containing the 686-894 construct at 52.4% buffer B. SDS-PAGE shows that even after the single purification step, the protein is very pure. The aggregated protein in solution was also run on SDS-PAGE, and shows that it is primarily formed from the DUSPs protein. The gel filtration fractions show that the purified protein is very pure. A weak band is observed under the major band on the SDS-PAGE, indicating possible proteolysis at the terminus. However, the difference is very small. **(B)** The gel filtration shows a very small void peak with a large trailing peak at 67.5 ml. SDS-PAGE shows that pooled fractions from this peak provide very pure protein with a single band corresponding to the 25.0 kDa protein. Additionally, the buffer optimisation using CHAPS produced a series of bands around the typical height of the DUSP domains on SDS-PAGE. The fractions taken across the peak of the gel filtration are shown, and at least three distinct bands are present.

Although the solubility limits between the two DUSP domains constructs were similar, the purifications showed that the shorter construct produced mainly monomeric protein. In addition, the shorter construct, without the disorder at the C-terminus, should be a better candidate for crystallisation. Therefore, the shorter USP20 686-894 construct was chosen for attempting to increase its stability. The first alteration made to most buffers is to increase the salt concentration. The solubility limit was no different when using 150 or 500 mM NaCl, nor was it different with 1 or 5% glycerol in the purification buffers. The next step was to screen for additives that would increase the protein's stability and solubility.

To increase the stability of the DUSP domains, a native-PAGE screen was performed (Figure 4.32). Here, the DUSP domains protein was added to test buffers prior to running on the gel. The screens consisted of both chaotropes and kosmotropes, as well as some additives that don't fit well into either of these groups, but may still affect the protein. It was predicted that if an additive stabilised the protein, the native-PAGE would show a discreet band of protein, showing stabilisation of a single species, and therefore monodispersity in solution. Only one test buffer showed an obvious band on the gel: ethylene glycol.

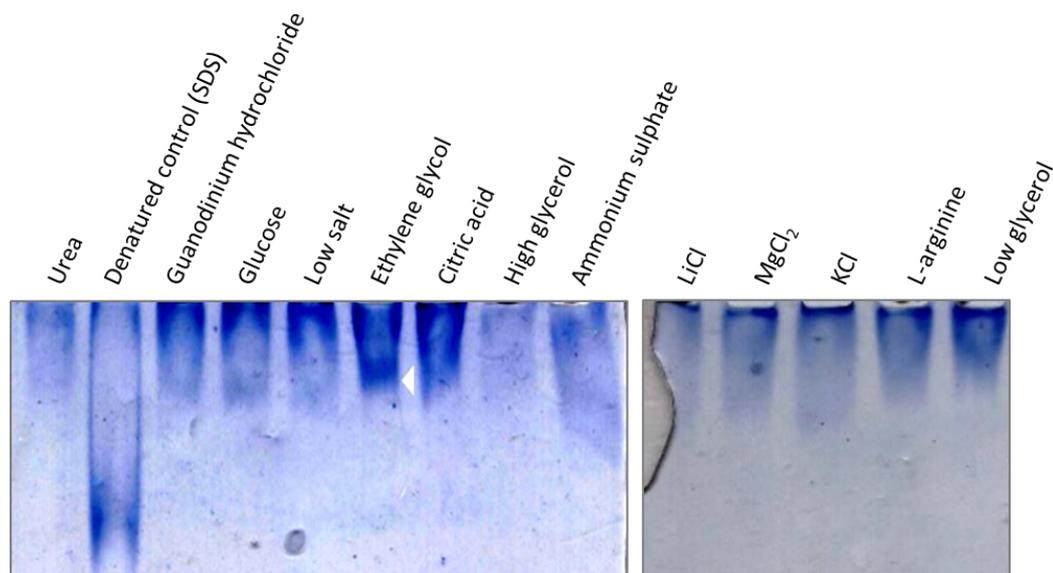


Figure 4.32. Native PAGE additive screen. Native PAGE shows that USP20 686-894 produces a smear in most cases. Ethylene glycol seemed to show a more pronounced band (white arrow) than any other additive. The denatured control has migrated further probably due to the charge difference caused by the presence of SDS.

Ethylene glycol was also observed as a hit in the Thermofluor assay as a thermostabilising additive for the DUSP domains (see section 3.4. Thermofluor). These data suggest that adding ethylene glycol to the DUSP domains' purification buffers will stabilise the protein, and could prevent aggregation in solution during the concentration steps. This additive allowed the protein to be concentrated to approximately 3 mg/ml before visible aggregation occurred. This is a three-fold increase in the solubility limit of the protein compared to the buffer containing 500 mM NaCl, 20/50 mM Tris pH 8.0, 1% glycerol and 1 mM β -mercaptoethanol.

In addition, a screen based on the method by Lindwall *et al.* [306] was performed (Figure 4.33). This is a relatively crude method where the bacteria are lysed in

different buffers and the band intensity is observed. Here it showed that pH 8.5 produced a comparatively high mass of protein, a result verified by the Thermofluor data (shown later). pH 8.0 had already been tested but whether a further increase to pH 8.5 would make a difference was tested. Increasing net negative charge does increase the solubility [316] so it could be that the higher the pH, the more soluble the protein becomes. Alone, this pH alteration appeared to

have no effect on the solubility limit, but all purification buffers were kept at pH 8.0 because of the evidence arising from the solubility assays.

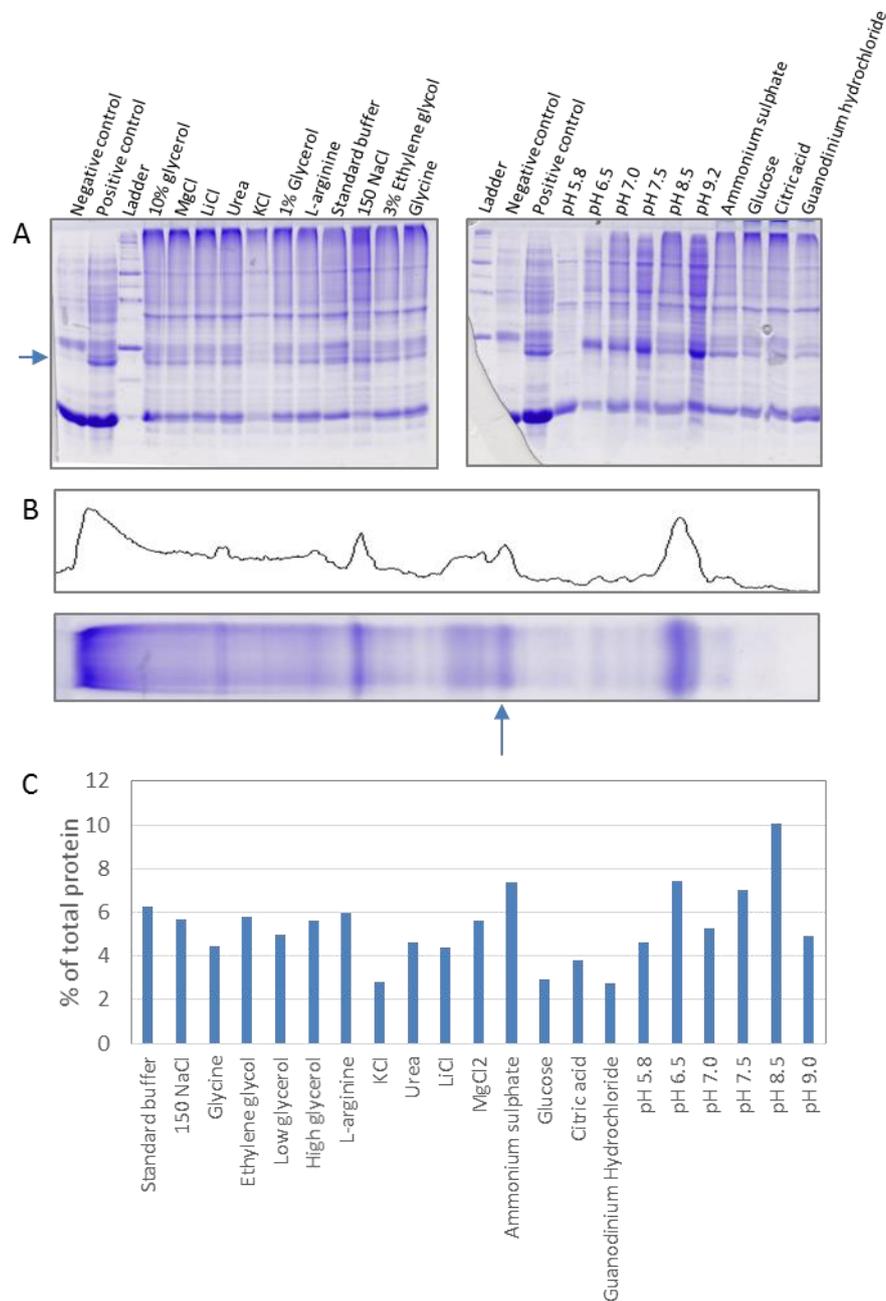


Figure 4.33. Stability screening USP20 686-894. (A) SDS-PAGE of the lysate from the stability screening. **(B)** An example of the ImageJ densitometry that was performed on each lane. The urea lane and its corresponding densitometry are shown. **(C)** The percentage of DUSP domains within the total lysate were calculated for each condition. The graph shows that pH 8.5 shows a distinctly higher percentage of DUSPs.

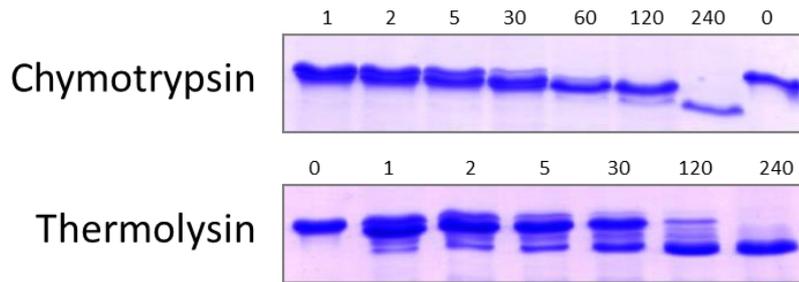
Membrane proteins are highly hydrophobic and detergents are used in their purification as they aid in protein solubility [317]. Although USP20 is not a membrane protein, if the aggregation of the DUSP domains is caused by hydrophobic interactions, then detergents could alleviate the extent of aggregation seen during the purification. 0.25% CHAPS, a mild detergent, was used during a purification to see if this could increase the protein's solubility. This did not occur, and SDS-PAGE analysis of the gel filtration curve shows that the protein seems to have degraded; observed by the presence of 3-4 discrete bands. This could be because of partial unfolding and enzymatic cleavage, mediated by the detergent. For this reason, detergents were not used in buffers for DUSP domains purifications.

4.3.3.2 **Limited proteolysis of USP20 686-914**

Limited proteolysis can be used to remove disordered regions of a protein, or cut down a protein's termini, in order to obtain a more stable construct in solution. In many cases this can be achieved while keeping the remaining protein correctly folded. This can aid in not only more stable protein, but increase the likelihood of obtaining crystals due to the lower disorder in the cleaved protein. USP20 686-914 was digested with chymotrypsin and thermolysin and incubated at 25 °C. The longer DUSP domains construct was used to identify whether using the Q894 terminus was a premature end to the domains. Chymotrypsin and thermolysin are non-specific with respect to their cleavage sites; chymotrypsin will cleave after most bulky hydrophobic side chains unless proline is in the P1' position.

Thermolysin also preferentially cleaves after bulky aromatics, unless there is a proline in the P2' position [318, 319]. Boiled samples from different time-points were run on SDS-PAGE to observe if cleavage of the protein occurred.

In both digestions, but particularly in the thermolysin digestion, a stable product is formed (Figure 4.34). These bands were excised and sent for tryptic digestion to analyse what part of the protein could be identified. In the control band, peptides close the N- and C- termini were observed. For the two digestions, these peptides were not detected. It could be the case that there was cleavage at both ends of the construct. There were no peptides following Q894, which suggests that this is a suitable protein terminus for the domains, and that the secondary structure may finish in this region. Cleavage was not observed centrally in the DUSP domains, as there was no dramatic drop in size of the bands on SDS-PAGE; all the bands remained well above the 20 kDa band on the ladder.



Undigested USP20 686-914

```
MSEEAMRERQQVVSLAAMREPSLLRFYVSREWLNKFNTFAEPGPITNQTFLC
SHGGIPPHKYHYIDDLVVILPQNVWEHLYNRFGGGPAVNHLVCSICQVEIE
ALAKRRRIEIDTFIKLNKAFQAEESPGVIYICISMQWFRWEAFVKGDNEPP
GPIDNSRIAQVKGSGHVQLKQGADYGQISEETWTYLNLSLYGGGPEIAIRQSV
AQPLGPENLHGEQKIEAETRAV
```

Chymotrypsin

```
MSEEAMRERQQVVSLAAMREPSLLRFYVSREWLNKFNTFAEPGPITNQTFLC
SHGGIPPHKYHYIDDLVVILPQNVWEHLYNRFGGGPAVNHLVCSICQVEIE
ALAKRRRIEIDTFIKLNKAFQAEESPGVIYICISMQWFRWEAFVKGDNEPP
GPIDNSRIAQVKGSGHVQLKQGADYGQISEETWTYLNLSLYGGGPEIAIRQSV
AQPLGPENLHGEQKIEAETRAV
```

Thermolysin

```
MSEEAMRERQQVVSLAAMREPSLLRFYVSREWLNKFNTFAEPGPITNQTFLC
SHGGIPPHKYHYIDDLVVILPQNVWEHLYNRFGGGPAVNHLVCSICQVEIE
ALAKRRRIEIDTFIKLNKAFQAEESPGVIYICISMQWFRWEAFVKGDNEPP
GPIDNSRIAQVKGSGHVQLKQGADYGQISEETWTYLNLSLYGGGPEIAIRQSV
AQPLGPENLHGEQKIEAETRAV
```

Figure 4.34. Limited proteolysis of USP20 686-914. (A) SDS-PAGE time-course of the chymotrypsin and thermolysin showing the time in minutes of each sample. **(B)** The mass spectroscopy results for the control band and the 240 minute band from each digestion. Observed peptides are in green and residues not seen in the assay are in black.

Multiple N-terminal peptides that were observed in the control sample were not observed in either cleavage sample (Figure 4.34). This suggests that the N-terminus is more susceptible to cleavage, and therefore possibly more disordered. This provides evidence that the cause of aggregation during the purification could

be breathing of the first helix of the DUSP domains (homology models predict the first helix to start on residue A689). The homology models also suggest that if this helix were removed, then the hydrophobic core residues of from the two other helices would be exposed. The resulting construct would most likely suffer solubility issues, meaning that shortening the N-terminus of the DUSP domains may not be a good solution.

4.4 Thermal shift assays for USP20 domain stabilisation

Screening proteins with the Thermofluor assay can identify buffers in which the protein is more stable; indicated by an increase in the protein's melting temperature (T_m). This is useful because studies have shown that increasing protein's stability, as measured by Thermofluor, increases crystallisation success of the protein [300, 320]. From the Thermofluor data, it is essential that graphs are interpreted with care. Amplification plots are obtained from the qPCR machine and these curves can be highly variable. Some of the data are shown (Figure 4.35) to show which curves were acceptable, and which curves are rejected for Boltzmann curve analysis. The curves should have similar profiles, with sharp increases in fluorescence. In some cases, the curve is prolonged, as in the DUSP domains with $ZnCl_2$ additive below, and cannot be interpreted because they would incorrectly appear as a large increase in T_m .

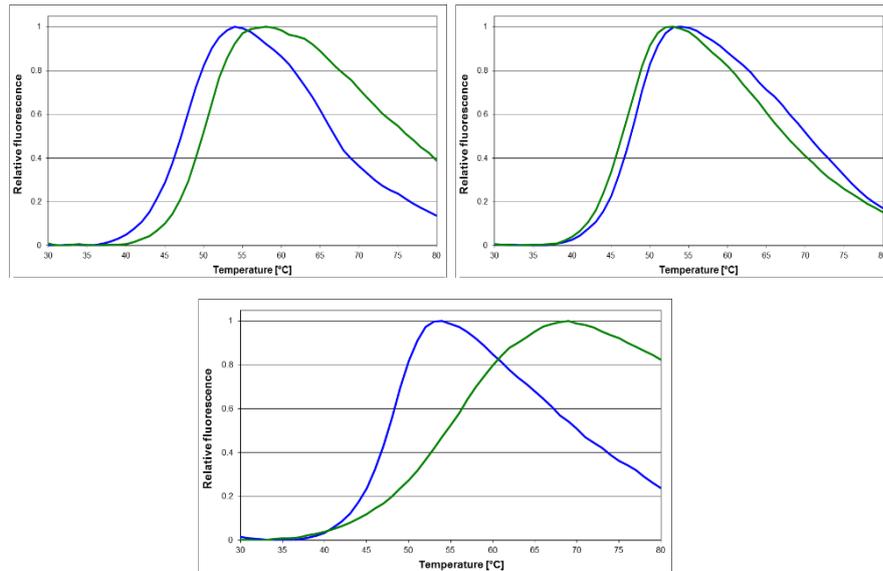


Figure 4.35. Thermofluor plots. Examples of the raw data received from the Thermofluor data are shown. Each graph shows two curves, a blue control curve and a normalised red assay curve. The relative fluorescence between the two curves are on the y axis and the temperature of the PCR tube is on the x axis. The top two graphs show acceptable curves that would be accepted in the data analysis. The left curve shows a large increase in melting temperature (T_m), and the right shows a small decrease. The bottom graph shows data that would be excluded from the analysis. The sigmoidal part of the graph does not match that of the control so assessing the T_m would produce a false positive increase. The melting temperature is taken from the middle of the sigmoidal curve of these graphs

Comparison of both proteins in a typical buffer of 150 mM NaCl and 50 mM Tris pH 7.5 (with 0.02% NaN_3) shows that the Znf-UBP is a highly stable protein with a T_m of 72.24°C. The DUSP domains, however, are far less stable with a low T_m of 47.32°C. Indeed, from the data above, it can be seen that the DUSP domains start the unfolding process around physiological temperatures (37°C). These T_m s could explain what is seen during purifications; the Znf-UBP domain is stable and far more soluble, whereas the DUSP domains aggregate. This aggregation may be caused by the low T_m of the DUSP domains, where regions of the protein partially unfold and interact with other partially unfolded proteins. This could lead to

conglomeration of many protein molecules that form solid, fibrous precipitates. This further supports the data obtained from the limited proteolysis, that the N-terminus is partially unfolding.

The NaCl concentration screen ranges from 50 to 750 mM. T_m s of both proteins correlate well with salt concentration (Figure 4.36). The data from the Znf-UBP Thermofluor show that the T_m increases linearly with the Log of the NaCl concentration ($R^2 = 0.9917$, $p < 0.0001$). At 50 mM NaCl the melting temperature of the Znf-UBP is 70.25°C and increases to 74.86°C at 750 mM NaCl. The DUSP domains' T_m increases linearly with concentration of NaCl ($R^2 = 0.9094$, $p = 0.0002$) and ranges from 46.61°C to 49.72°C at 50 and 750 mM NaCl concentrations.

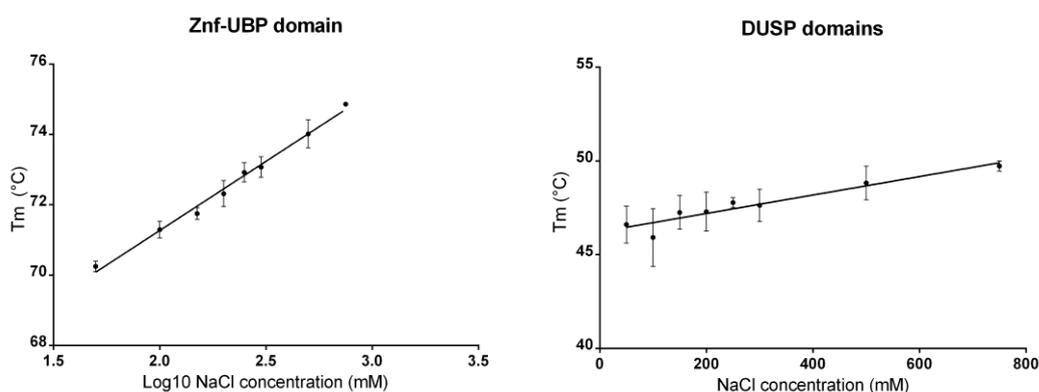


Figure 4.36. Effect of salt concentration on T_m of Znf-UBP domain and DUSP domains. Means of triplicate repeats are plotted for Thermofluor analysis of NaCl concentrations. Both proteins show the highest T_m with highest NaCl concentrations, indicating that buffers with high salt concentrations will stabilise them. The Znf-UBP domain shows a non-linear increase of T_m with NaCl concentration. At each subsequent increase in NaCl concentration, a lesser improvement of thermostability is achieved. The DUSP domains show a linear association with NaCl concentration.

Both domains were also tested with an additive screen (Figure 4.37). This screen consisted of a control buffer (150 mM NaCl and 50 mM Tris pH 7.5) and the same buffer with the addition of a series of additives. Only two additives showed a positive increase in T_m for the Znf-UBP domain: L-arginine and $MgCl_2$. The L-arginine assays showed a 3.2 °C increase in T_m compared to the control ($p=0.0126$). $MgCl_2$ assays show a 1.1 °C increase compared to the control ($p=0.0263$). The DUSP domains also only showed an increased T_m with two additives: glycerol and ethylene glycol. Glycerol produces a 3.4 °C increase ($p=0.0209$) and ethylene glycol produces a 3.3 °C increase ($p=0.0004$). Very interestingly, the Znf-UBP shows a massive destabilisation from the addition of EDTA and $ZnCl_2$. Destabilisation by EDTA is most likely due to the chelation of the three Zn^{2+} ions that the Znf-UBP domain is predicted to coordinate. These ions are probably required for structural integrity of the fold, and removing the ions causes serious destabilisation of the protein. This solves the uncertainty of whether the bacterially-expressed Znf-UBP domain contains the Zn^{2+} ions that it should have, which was the reason why an attempt at adding $ZnCl_2$ into the purification buffers was attempted. If it did not have the ions bound, then it would be unlikely that EDTA would destabilise the protein at all. The destabilisation caused by addition of $ZnCl_2$ explains the disruption of the monodispersity of the gel filtration trace when $ZnCl_2$ was added to the purification buffers. It supports the idea that the presence of free Zn^{2+} ions disrupts the ions that are present and coordinated in the structure.

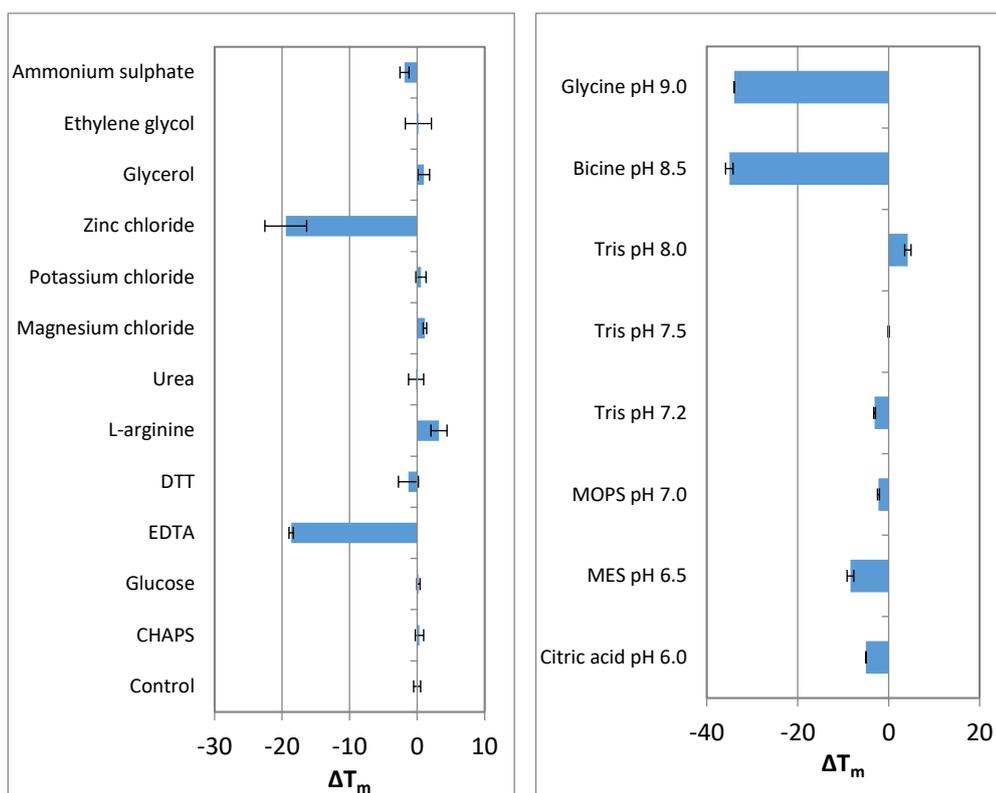


Figure 4.37. Effect of additives and buffers on T_m of the Znf-UBP domain. The ΔT_m , the difference between the control T_m and the sample T_m , are shown for the additive screen (left) and buffer/pH screen (right).

From the Znf-UBP domain buffer/pH screen, only pH 8.0 showed an increase in T_m . An increase of 4.22 °C was observed ($p=0.0005$). No significant increases in T_m were observed for the DUSP domain buffer/pH screen (Figure 4.38).

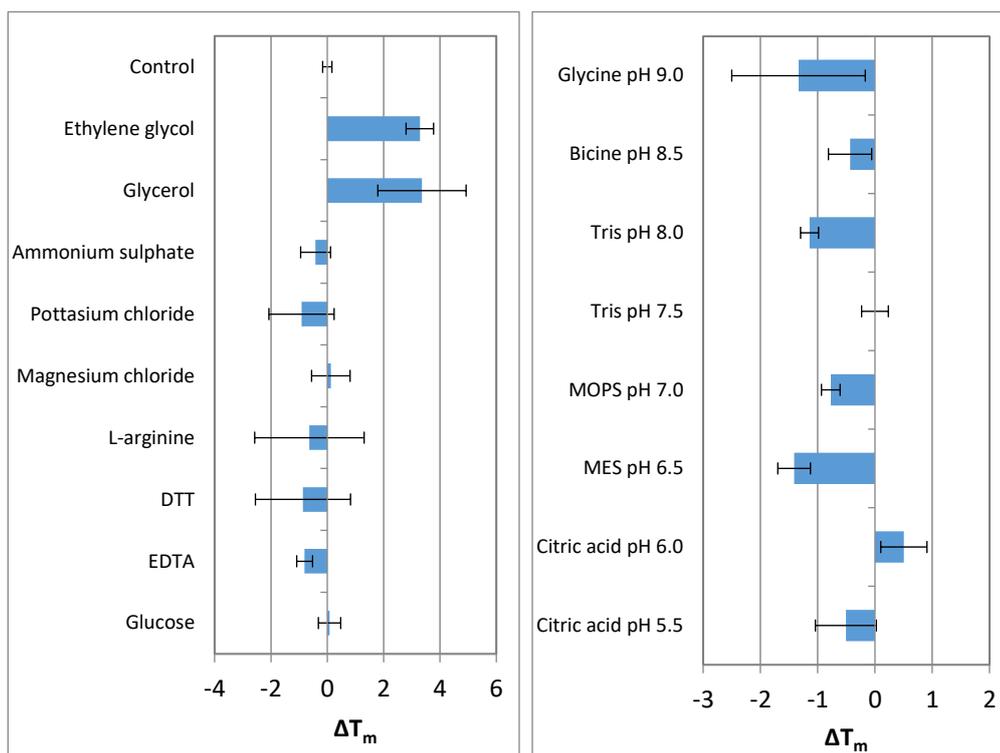


Figure 4.38. Effect of additives and buffers on T_m of the DUSP domains. The ΔT_m , the difference between the control T_m and the sample T_m , are shown for the additive screen (left) and buffer/pH screen (right).

The identification of high NaCl concentrations, Tris pH 8.0, L-arginine and $MgCl_2$ as thermostabilising conditions required testing as to whether they could influence the ability of the Znf-UBP domain to crystallise. The Znf-UBP domain was purified with a 500 mM NaCl buffer, Tris pH 8.0. Following gel filtration, the sample was split into three aliquots. One aliquot was put down in crystal trials without any additive. The other two samples were buffer exchanged into the same buffer but with 10 mM L-arginine or 100 mM $MgCl_2$ using a PD-10 column. These samples were then used in crystal trials, but no protein crystal grew in any of these conditions.

4.5 Novel solubility tag design

To increase solubility and to aid in crystallisation of the DUSP domains, they were expressed as fusions to solubility tags. Firstly, maltose binding protein (MBP) was used as this is a commonly used fusion tag for increasing protein solubility, and can lead to crystallisation of proteins that do not typically crystallise on their own [229]. However, although MBP can drive the crystallisation of proteins, its large size can still impede the crystallisation of the fusions [321]. To solve this problem novel tags were designed that would utilise a two-pronged approach to increase the probability of protein crystallisation. Firstly, tags would be chosen specifically to increase the solubility of the overall protein construct, thus acting as solubility tags. Also, the proteins chosen would have characteristics that should increase the likelihood of crystallising the fusion, and therefore act as a vehicle for carrier-driven crystallisation.

4.5.1 Identifying PDB proteins

To produce crystallisable solubility tags, the PDB database [232] was screened for proteins with properties that should maximise crystallisation efficiency. Criteria included: (1) proteins less than 400 amino acids. This reduces the chance that the protein contains flexible regions that may negatively impact on crystallisation. (2) No disulphide bonds. This is to ensure that protein expressions can be performed in reducing conditions, such as in *E. coli* cytoplasm. (3) No ligands in the structure. Firstly, proteins may alter in conformation when ligands bind. This may mean there would be inherent flexibility in the protein, which is undesirable. Secondly,

it could also mean that presence of the ligand would be required to ensure monodispersity in solution. (4) $<1.6\text{\AA}$ X-ray diffraction resolution. Higher resolution structures could be an indication that a protein is structurally very stable, and that there is very little flexibility. It furthermore indicates that the crystal lattice is very stable as the protein contributes many crystal contacts. These features would be desirable to promote the formation of quality crystals of the fusion protein. (5) *E. coli* expression. As this is the cheapest and quickest method of producing large quantities of recombinant protein for crystallisation, the expression system for most fusion proteins will be in bacteria. Also, as many insoluble proteins may be recombinant proteins from non-*E. coli* organisms, it is preferential to enhance the solubility of these proteins in this organism.

These criteria produced 447 entries when only searching proteins with 90% sequence homology. Candidates were shortlisted using Pymol and a literature search to assess factors such as solubility (concentration at which crystallisation trials were performed), globularity, charge distribution, B-factors, and C-terminal secondary structure. Solubility is an essential characteristic of a solubility tag; to increase the solubility of its fusion partner it must be soluble itself. Globularity is preferred for crystallisation as non-globular proteins tend to be less soluble and have more conformational flexibility, sometimes having no order at all [322]. As mentioned previously well structured, rigid proteins should be better for crystallisation, which are characteristics of more globular proteins. Charge distribution should be reasonably even, with no particularly charged or uncharged regions. Regions with high positive or negative electrostatic potential can lead to

aggregation, as can large regions of low potential (hydrophobic patches) [323-325]. B-factors are parameters describing the displacement of an atom about its average position [326]. The higher the B-factors, the more movement there is. Again, flexibility is unfavourable for crystallisation so proteins with regions of high B-factors should be avoided, unless they are at the termini and can be removed from the construct. A C-terminal helix is desirable because the secondary structure here should increase the rigidity where the fusion protein starts. This mimics the strategy used by the fixed-arm MBP, which was successful in producing diffraction-quality crystals [229]. The theory is that it reduces the overall flexibility of the protein, which should increase chances of crystallisation. The published purification method should not require any special additives as the fusion should be purified in a typical manner. For the purposes of this thesis, the proteins selected were also examined to ensure they do not contribute to cellular ubiquitination or deubiquitination pathways.

Finally, two of the PDB entries were chosen to use as solubility tags; the calponin homology domain from human β -spectrin (referred to as the PDB code: 1BKR [327]) and receiver domain from *Myxococcus xanthus* social motility protein frzS (PDB code: 2GKG [328]) (Figure 4.39).

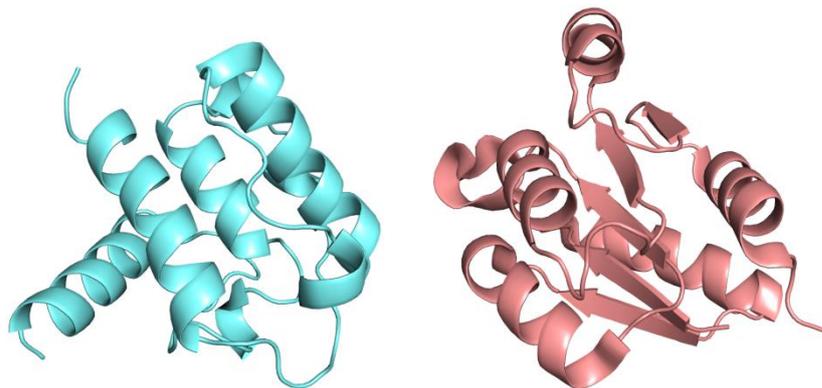


Figure 4.39. 1BKR and 2GKG. The structures of the two shortlisted proteins for crystallisable solubility tags are 1BKR (blue) and 2GKG (salmon).

1BKR is a completely alpha-helical protein whereas 2GKG has a central beta-sheet encapsulated in helices. Both proteins form compact globular structures with minimal flexibility as shown by visual representation of B-factors in Pymol (Figure 4.40A). Extreme flexibility occurs in the two terminal residues of both proteins. 1BKR's terminal Ser, Lys and Met residues have alpha carbon B-factors of 17.81, 34.79, and 54.18 Å², respectively. 2GKG's terminal Phe, Pro and Glu residues have alpha carbon B-factors of 13.58, 18.96 and 41.26 Å², respectively. Both proteins have acceptable surface charge distributions with no large, highly charged or hydrophobic patches (Figure 4.40B). The purification procedures for both proteins were standard and the stated concentrations for crystallisation were 10 mg/ml and 20 mg/ml for 1BKR and 2GKG, respectively [327-329].

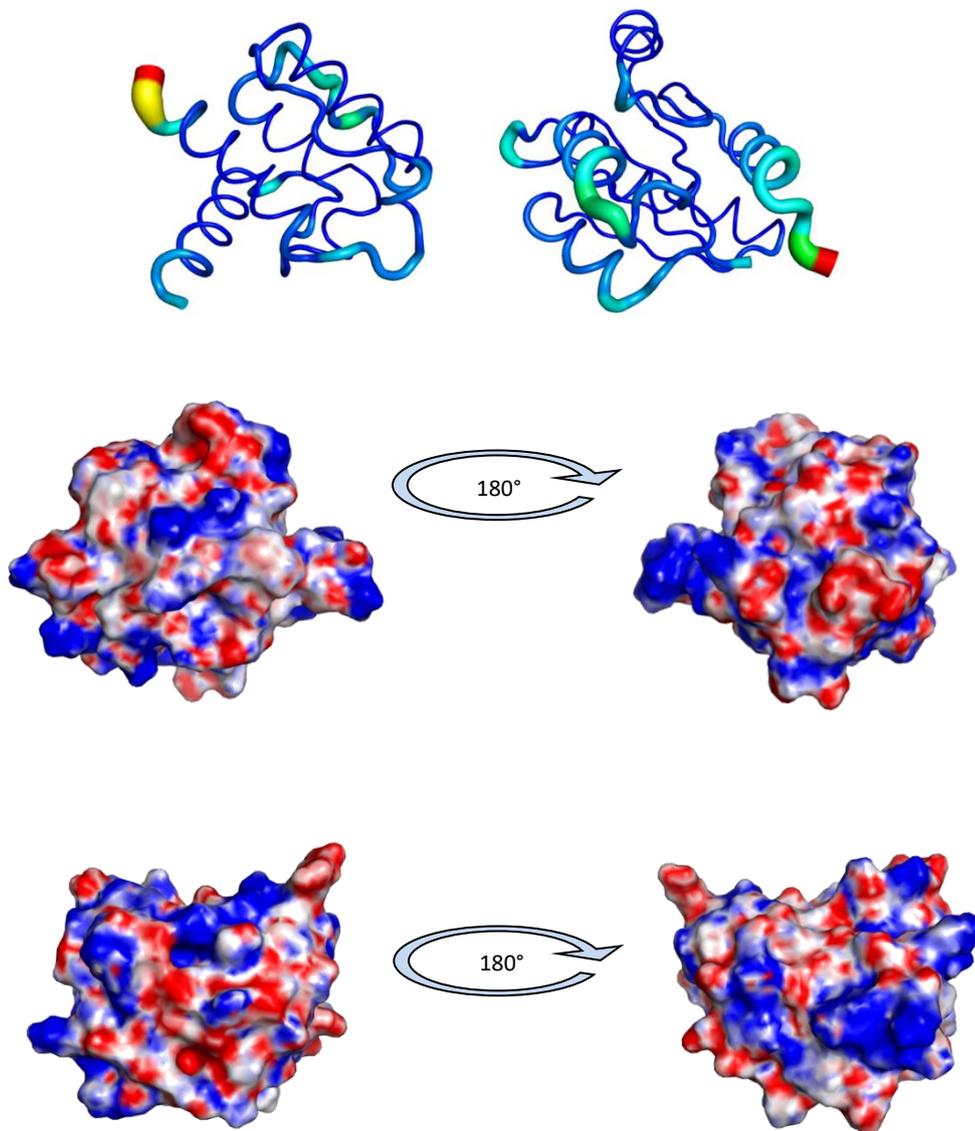


Figure 4.40. 1BKR and 2GKG B-factors and electrostatics. (A) B-factor putty representation of the proteins' B-factors. B-factors are represented by thicker tubing and colour coordinated with a blue-yellow-red spectrum. 1BKR is shown on the left and 2GKG is shown on the right. **(B)** 180° rotations of the surface vacuum electrostatics of 1BKR. **(C)** 180° rotations of the surface vacuum electrostatics of 2GKG.

4.5.2 Modifying the proteins

4.5.2.1 SERp

Surface entropy reduction was performed using the SERp server (Table 3.6). The server identifies amino acids with high entropy (lysine and glutamine) and assesses their suitability for mutation to reduce the surface entropy of the protein. This has been shown to increase the efficiency for crystallisation in a number of proteins. The server aims to produce a stretch of solvent exposed amino acids with low surface entropy by mutation of the high entropy residues to low entropy residues, such as alanine or serine. The server identified three clusters for each protein. For each case the solved structure was visualised to ensure the mutations would not impact the protein's stability.

Table 4.6. SERp cluster suggestions.

Protein	Cluster	Score	Residue(s)	Patch length
1BKR	1	5.45	Lys54 Lys55	2
	2	3.24	Glu91 Lys92	2
	3	2.56	Lys1	1
2GKG	1	4.17	Lys68 Lys69	2
	2	4.10	Lys1 Lys2	2
	3	3.21	Lys92 Lys94	4

Pymol was used to assess whether the suggested mutations from SERp may damage the stability or fold of the protein, shown in Figures 4.41 and 4.42.

The highest scoring cluster for 1BKR suggests mutating Lys54 and Lys55. Lys55 would be a suitable candidate for mutation because its side chain does not interact with any other amino acid in 1BKR; however, Lys54 seems to interact with an Asn20 side chain oxygen. As this interaction may contribute to the stability of 1BKR, and a whole cluster must be mutated in order for enhanced crystallisation, neither of these lysines were mutated. Similar to Cluster 1, Lys92 in Cluster 2 does not make interactions, whereas Glu91 forms a salt bridge with Lys4; therefore this cluster was also not chosen. Cluster 3 only consisted of the first lysine in the structure, which made no side chain interactions. Therefore, this residue was mutated to a serine.

The highest scoring cluster for 2GKG included Lys68 and Lys69, both of which made interactions to Asp residues in the structure. Cluster 2 included Lys1, which makes polar interactions with backbone atoms of Ile118, and Lys2, which forms a salt bridge with Asp43. Cluster 3 comprised Lys92 and Lys94. Lys92 made interactions with a nearby alanine backbone oxygen, and Lys 94 made no side chain interactions. It is likely that the Lys92 interaction is not required for structural integrity. Lys92 was therefore mutated to an alanine and Lys94 was mutated to a serine.

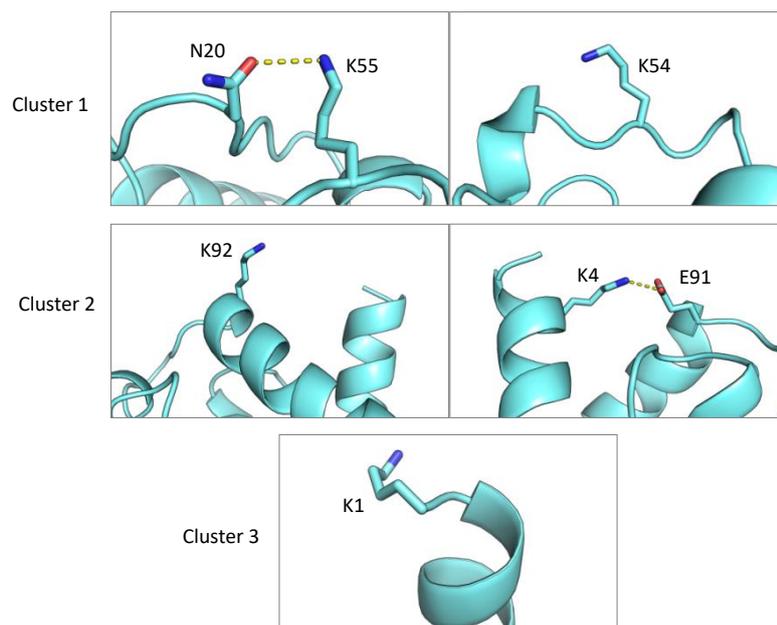


Figure 4.41. Evaluation of 1BKR SERp clusters. Each cluster from the SERp prediction is shown. Clusters 1 and 2 make potentially valuable interactions with other charged or polar residues. Cluster 3 is a reasonable target for alteration.

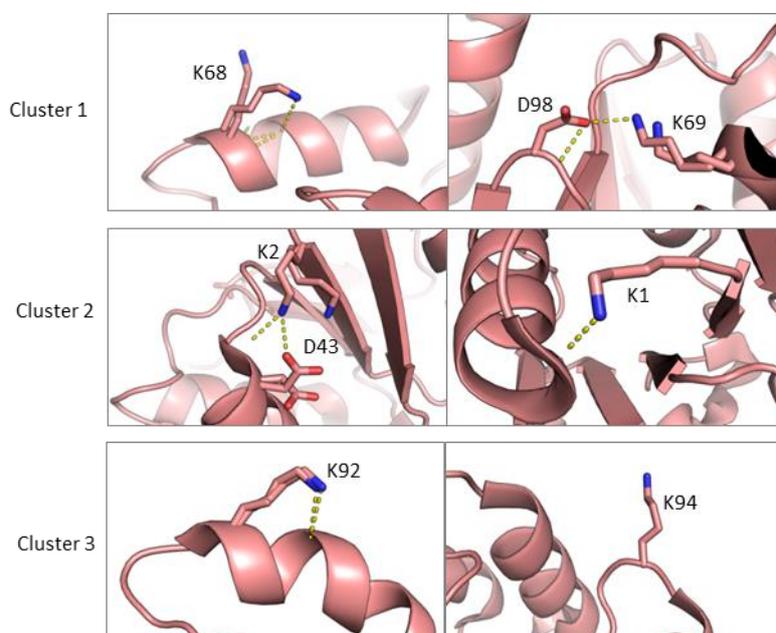


Figure 4.42. Evaluation of 2GKG SERp clusters. Each cluster from the SERp prediction is shown. Clusters 1 and 2 make potentially valuable interactions with other charged or polar residues. Cluster 3 is a reasonable target for alteration.

4.5.2.2 **Modification of the C-terminus**

In both structures of the proteins, the C-terminal residues displayed high flexibility, as indicated by the high B-factors. Although it was decided that a small amount of flexibility may be useful when there is a protein fused to the C-terminus of the tag (to avoid steric issues), too much flexibility will allow excessive movement between the tag and the fusion partner and could prevent crystallisation. It may also expose the linking region of the fusion and tag to proteolytic enzymes, which could lead to breakdown of the fusion in the cell or during purification. Therefore, in both cases, C-terminal residues were removed; 1BKR had its terminal lysine and methionine residues removed and 2GKG had its terminal glutamic acid residue removed.

4.5.2.3 **Codon optimisation**

The proteins were further modified to increase expression by codon conversion. As *E. coli* has a dissimilar codon bias to human genes [330], and it regulates its own gene expression by incorporating rare codons [331, 332], optimising the solubility tag genes to use the most common *E. coli* codons could increase their expression [333]. During gene optimisation by the sequence manipulation suite, a BamHI restriction site was incorporated into the 1BKR sequence. As this restriction site was to be used in the multiple cloning site (MCS), these codons were manually altered to the second most frequent codons in *E. coli*.

4.5.2.4 **Multiple cloning site**

A multiple cloning site was designed to follow the gene for 1BKR and 2GKG. The MCS restriction site consists of eight in-frame restriction sites; 5'-3' NotI, NheI, KpnI, BamHI, EcoRI, SacI, HindIII and XhoI. The first two sites are based on the MCS from short linker pMAL fusion vectors [321]. Cloning into the first two NotI or NheI produces a fusion with a linker of either three alanines or four alanines and a serine, respectively. This fixed-arm linker revolutionised crystallisation of fusions with MBP, and thus it was hoped to be as effective in the 1BKR and 2GKG fusions. The final vector consisted of 1BKR or 2GKG followed by an MCS inserted into a PRSF vector (RSF origin and kanamycin resistance gene). Full sequences of the DNA used to create the 1BKR and 2GKG vectors can be found in the appendix.

4.5.2.5 **2GKG-His mutant**

In addition to these two vectors, a 2GKG vector was produced without a His-tag. Instead, a gene was synthesised with R46H, K71H, D73H, D74H, K76H, N77H, K94H, and K96H mutations. This was done in order to increase the rate of crystallisation by decreasing the disorder usually associated with the presence of a His-tag. The residues that were mutated to histidine should bind to a nickel column to allow IMAC purification without the presence of a His-tag. However, these radical changes to the primary sequence could have a negative impact on the protein's ability to fold properly.

4.5.2.6 Final solubility tag vectors

The final sequences up- and downstream of the tag DNA sequences are shown in Figure 4.43. The sequences of the 1BKR tag and 2GKG are identical, whereas, due to the removal of the 5' His-tag sequence, 2GKG-His is different. The 5' MCS is identical for all tags. Vector maps of all three proteins are given in Figure 4.44.



Figure 4.43. Upstream and downstream sequences of the solubility tag vectors. The DNA sequences, restriction sites and corresponding in-frame amino acids from the initiator methionine in the final vector. The sequences are the same upstream and downstream of 1BKR and 2GKG. The sequence is different upstream if the 2GKG-His vector due to the lack of a 5' His-tag sequence.

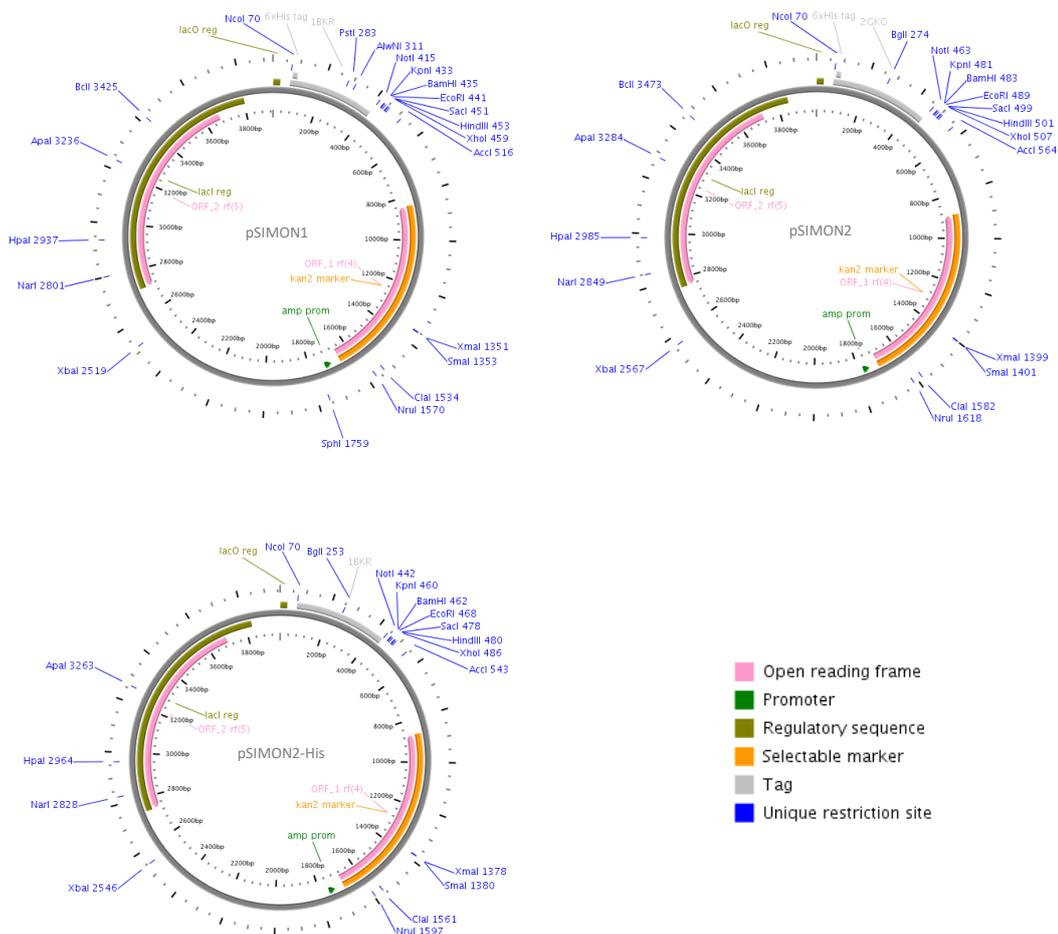


Figure 4.44. Maps of the solubility tag vectors. PlasMapper [334] maps of the solubility tag vectors, pSIMON1 (containing 1BKR), pSIMON2 (containing 2GKG) and pSIMON2-His (containing 2GKG-His).

4.5.3 Tag purifications

To ensure that the tags still fold and to assess the relative solubility and stability of 1BKR and 2GKG, both constructs were expressed alone without fusion to another protein. This was achieved by using site-directed mutagenesis to alter the first codon of the linker (alanine) into a stop codon. Both purifications were generally unremarkable; obtaining a single peak on nickel-affinity purification (not shown) and a monodisperse peak on gel filtration (Figure 4.45). One issue with the

2GKG tag is that it elutes at a low concentration of imidazole: 15.9% buffer B (96.3 mM imidazole). This may be due to partial burial of the his-tag, which, although does not cause an issue with purification of the DUSP fusion used in this thesis, has caused problems with purification of other fusions in the lab (data not shown). In addition, this prevents its detection by ELISA and Western blotting by the anti-tetra histidine antibody used in the lab. Notwithstanding 2GKG's low imidazole concentration elution, both 1BKR and 2GKG protein samples were very pure following two purification steps (Figure 4.45).

The molecular masses of 1BKR and 2GKG are 13.3 and 14.0 kDa, respectively; and their elution volumes from the HiLoad 16/60 Superdex 75 prep grade gel filtration column were 84.24 and 81.53 ml. These elution volumes give respective predicted masses of 9.2 and 10.9 kDa; lower than their actual mass, but in agreement with their highly globular, compact structures. By far, 2GKG was more soluble than 1BKR. At approximately 10 mg/ml 1BKR was showing signs of fibrous aggregation during concentration. Although quantification of 2GKG's concentration is more difficult because it lacks tryptophan residues, this protein reached approximately 100 mg/ml without signs of aggregation, and after overnight room temperature incubation showed no precipitation or breakdown. Overall, the total yields of 1BKR and 2GKG are approximately 5 and 20 mg/L of culture, respectively. Following crystallisation trials, crystals formed in five conditions in the 1BKR trials, and

diffraction was observed from multiple crystals. However, 2GKG only formed crystals in one condition, and diffraction did not occur.

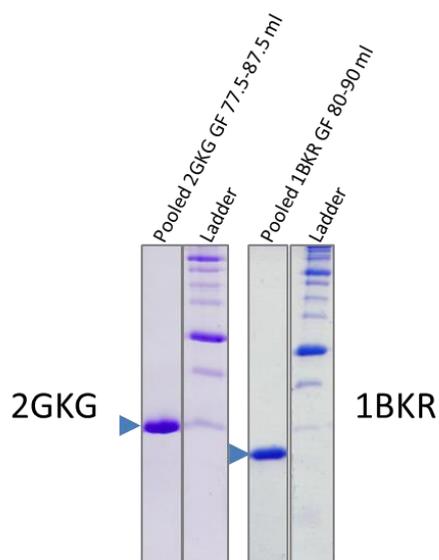
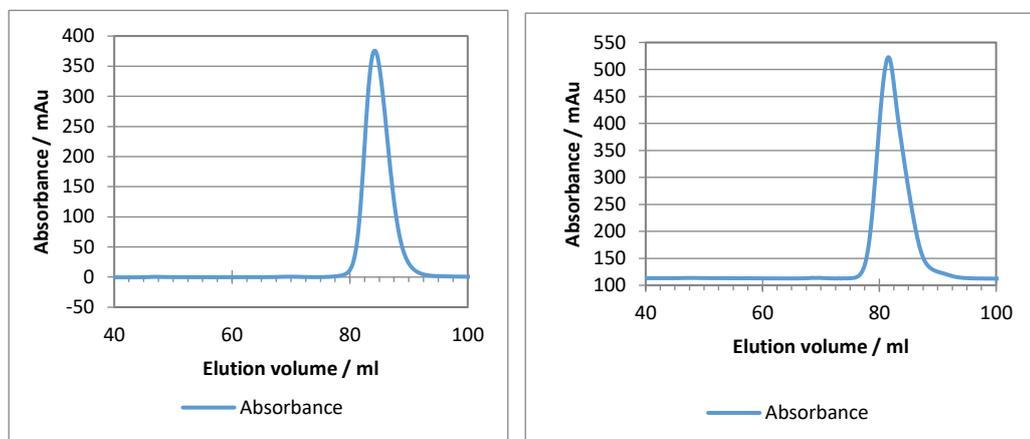


Figure 4.45. 1BKR and 2GKG gel filtration. (A) Gel filtrations of 1BKR (left) and 2GKG (right) show single monodisperse peaks at elution volumes that predict monomeric protein. SDS-PAGE shows that pooling of all fractions from the peak produces very pure protein, which runs at the correct size.

These data suggest that alterations made during production of the fusion vectors have not negatively impacted the proteins' characteristics. Also, as single tags, 1BKR is more likely to form crystals at these concentrations, as more crystals formed in the crystallisation trials. This may suggest that 1BKR may be more likely to drive crystallisation when bound to a fusion partner. However, it may just be that 2GKG alone is so soluble that much higher concentrations of protein are required for the same efficiency of crystallisation. The data also suggest that 2GKG may be the more superior tag for increasing solubility of the construct as it is far more soluble than 1BKR alone. A fusion of 1BKR and 2GKG with the DUSP domains of USP20 will give some evidence to these statements.

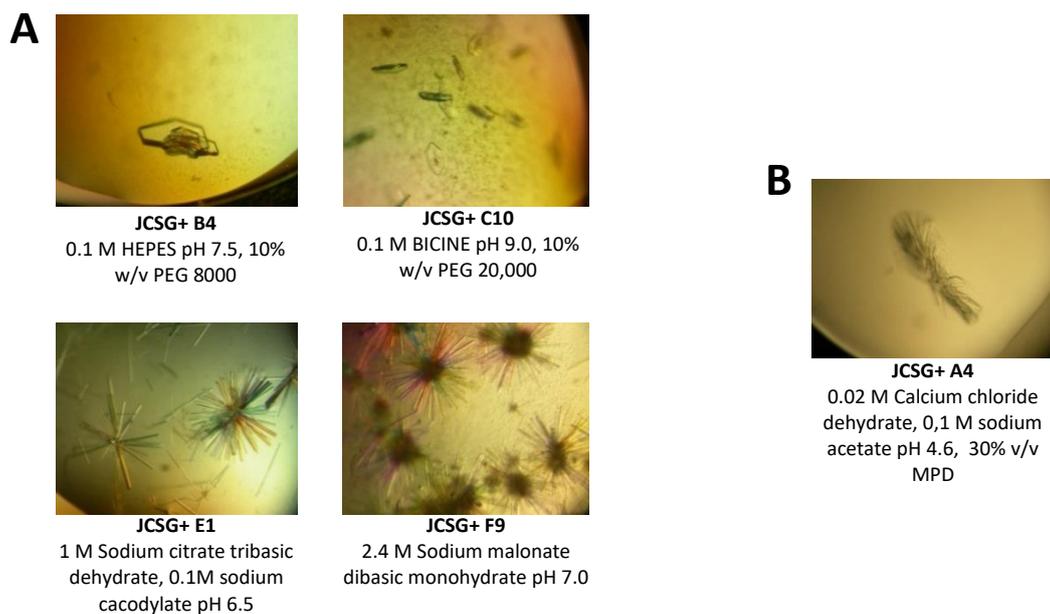


Figure 4.46. Crystals of the solubility tags. (A) Crystals obtained from the 1BKR crystallisation trials. **(B)** The crystal cluster obtained from the 2GKG crystallisation trial.

4.5.4 DUSP domains-solubility tag fusions

Three solubility-tagged constructs of the USP20 DUSP domains were expressed.

These are shown in Figure 4.47.

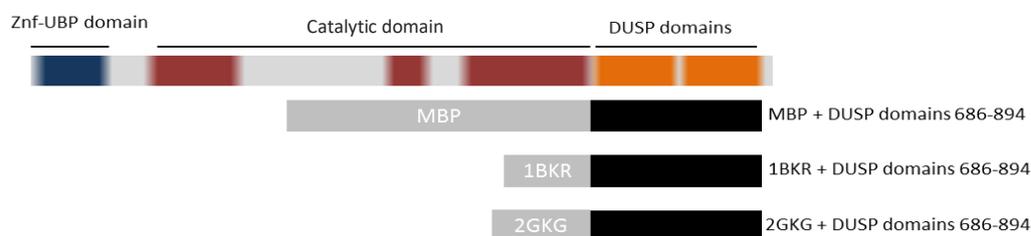


Figure 4.47. USP20 DUSP domains fusion constructs. The regions of USP20 expressed are shown in black. Extra residues (His-tag, solubility tag and restriction site residues) are shown in grey.

4.5.4.1 MBP-DUSP domains purification

To compare the new tags with a commonly used tag, the DUSP domains were cloned into pMALX(E); a vector that produces an MBP fusion with a short, fixed-arm linker. The *E. coli* MBP sequence in this vector also has mutations to decrease its surface entropy compared to wild type MBP, and increase the rates of crystallisation of the fusion. In addition, it is also a periplasmic form of the MBP protein; meaning the fusion protein is translocated to the periplasmic space following translation.

The MBP-DUSPs fusion (USP20 686-894) massively increased the expression levels and solubility of the DUSPs. This protein was very soluble; reaching concentrations of 15 mg/ml without visible precipitation or aggregation.

The total yield of protein following step elution from the MBP-trap column is around 50 mg. When this protein is separated by gel filtration on the HiLoad 16/60 Superdex 200 prep grade column, four peaks are observed: void, oligomer, monomer and MBP/breakdown (Figure 4.48). The majority of the protein elutes in the oligomer peak, which is a broad peak that spans from void to the monomer peak. The elution volume of the oligomer peak is 53.7 ml; giving it a predicted molecular weight of 509.8 kDa. The monomer elutes at 74.5 ml, predicting a protein mass of 87.0 kDa. The actual mass of the fusion protein is 63.9 kDa, suggesting that the protein is in an extended conformation, increasing its hydrodynamic radius, and therefore presenting an elution volume lower than expected. The oligomer peak is broad with a predicted mass eight times greater than the actual protein mass, which may suggest that the protein is forming soluble micro-aggregates, or that the proteins are forming dynamic micelle-like structures with an equilibrium of eight units.

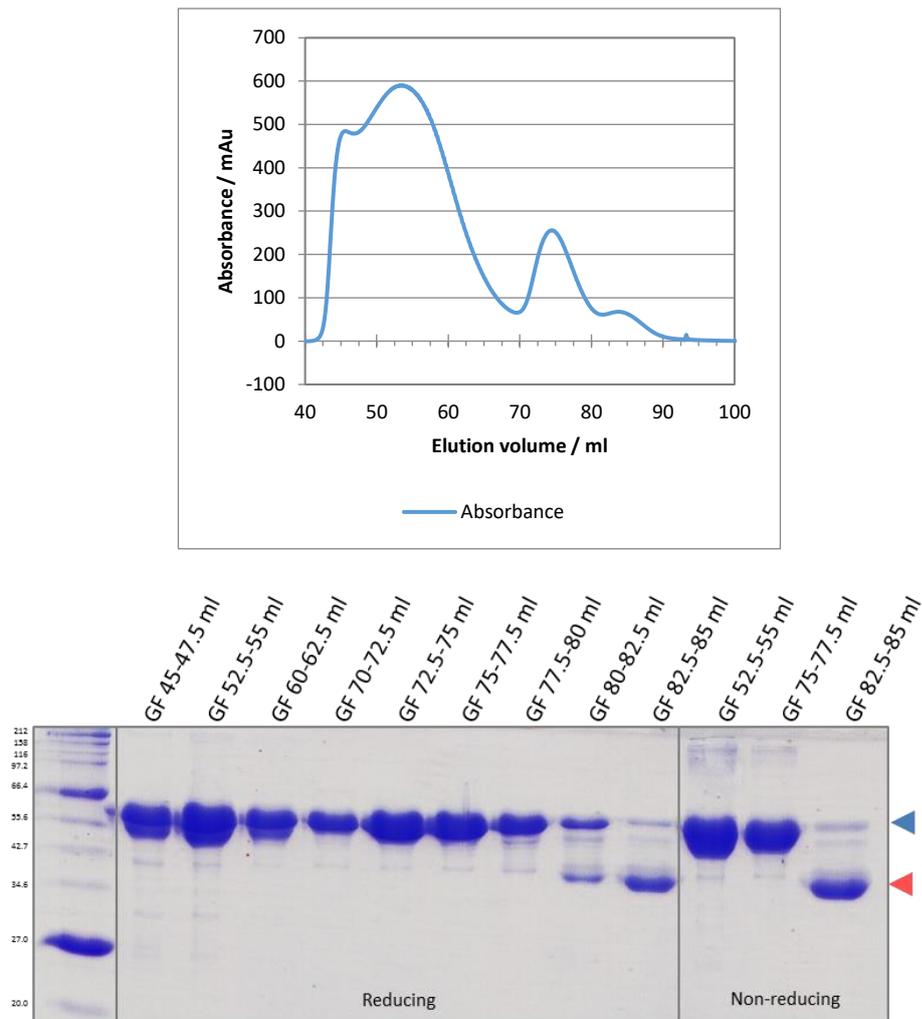


Figure 4.48. MBP-DUSPs gel filtration. (A) The gel filtrations chromatogram of MBP-DUSPs shows four obvious peaks; a void peak, high molecular weight peak, monomer peak and breakdown peak. **(B)** SDS-PAGE shows that all peaks, except the 84 ml breakdown peak, contain the full length fusion protein (blue arrow). The major breakdown product is the correct size for MBP (red arrow). Also, the non-reducing SDS-PAGE samples show that only a small fraction of the proteins are forming cysteine-mediated oligomers (as observed by the high molecular weight bands in these lanes).

SDS-PAGE was performed to ensure that the oligomers were not caused by the presence of disulphide crosslinking between the proteins (Figure 4.48). This should not occur because the proteins are expressed in the reducing environment of the *E. coli* cytoplasm, and DTT was used prior to gel filtration. Nonetheless, if the

cysteines formed disulphides and were not solvent accessible, then the cysteines may not be reduced by the DTT. Comparisons between reducing and non-reducing samples on SDS-PAGE show that a minor component of the oligomer peak is formed by the presence of these bonds, however the majority of the protein is still a single band of correct molecular weight, indicating that they are not disulphide-linked. This suggests that either polar/electrostatic or hydrophobic interactions mediate the oligomer formation. However, the use of buffers with high salt (500 mM NaCl), low salt (150 mM NaCl) or detergent (0.25% tween) did not reduce the oligomeric fraction (data not shown).

When the MBP column elution sample is subject to anion exchange, multiple peaks are obtained (Figure 4.49); two of which are evidently the MBP-DUSP domains fusion as assessed by SDS-PAGE. The first is a small peak that elutes at 36.9% of buffer B (197.4 mM NaCl) and the second is a larger, broad peak at 53.5% of buffer B (276.8 mM NaCl). Analytical gel filtration shows that the later eluting peak consists primarily of high molecular weight species, whereas the earlier peak contains proportionally more of the monomeric species. Additionally, the analytical gel filtration also indicates other species in the sample, indicated by multiple discreet peaks on the chromatogram. By combining the fractions composing the smaller peak on anion exchange, a single (shouldered) peak is obtained. SDS-PAGE shows that the shoulder contains a cleavage product consisting of either MBP or MBP and a small part of the DUSP domains. At this stage in the purification, this additional protein was still a considerable

contaminant to the sample of full length MBP-DUSPs fusion. Therefore, the sample was further anion exchanged to produce relatively pure protein.

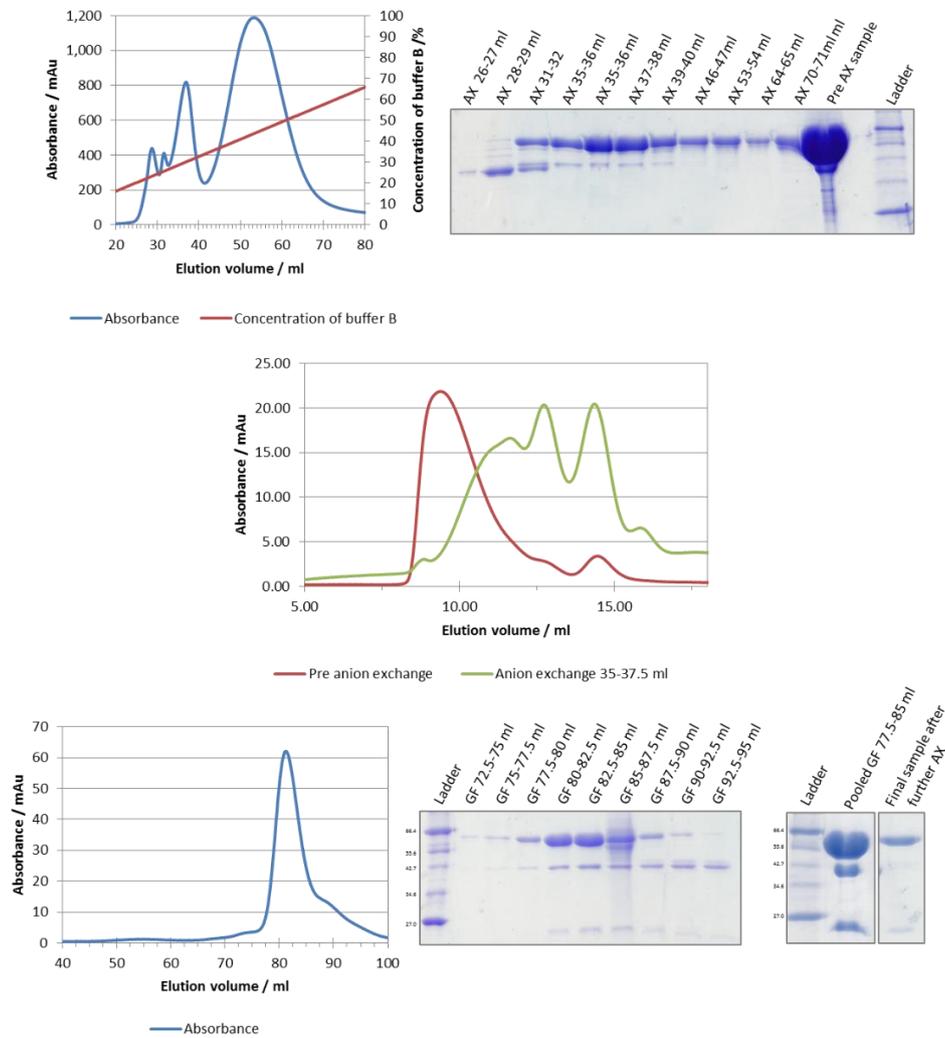


Figure 4.49. MBP-DUSPs extended purification (A) The anion exchange (right chromatogram) shows four peaks. The two smaller peaks correspond to the major MBP breakdown product and a slightly larger fragment. The two largest peaks appear to contain the full length MBP-DUSP fusion. **(B)** Analytical gel filtration of the pre-anion exchange and the smaller fusion-containing peak show that the large aggregates are removed from this sample. **(C)** Gel filtration using the superdex 200 column show that the pooled anion exchange peak produces a good peak on gel filtration, although it is slightly contaminated with the MBP and DUSP breakdown products. The final SDS-PAGE (right) show that a final repeat anion exchange after gel filtration can increase the purity of the sample by removing most of the MBP breakdown product. However, there is still some contamination from both breakdown products in the sample.

4.5.4.2 **1BKR-DUSPs purification**

Protein expression of this fusion was about equal to that of the His-tagged construct. In comparison to the MBP-DUSP domains construct, 1BKR-DUSP domains was far less soluble. However, it still had superior solubility to the His-tagged construct. The maximum solubility of this construct was approximately 3 mg/ml in a typical gel filtration buffer (150 mM NaCl, 20 mM Tris pH 7.5, 1% glycerol), equalling that of the His-tagged construct in the modified ethylene glycol buffer. Increasing the NaCl concentration to 500 mM and adding ethylene glycol to the purification buffer of 1BKR-DUSP domains did not alter the solubility limit of this protein. An example purification of 1BKR-USP20 is given in Figure 4.50.

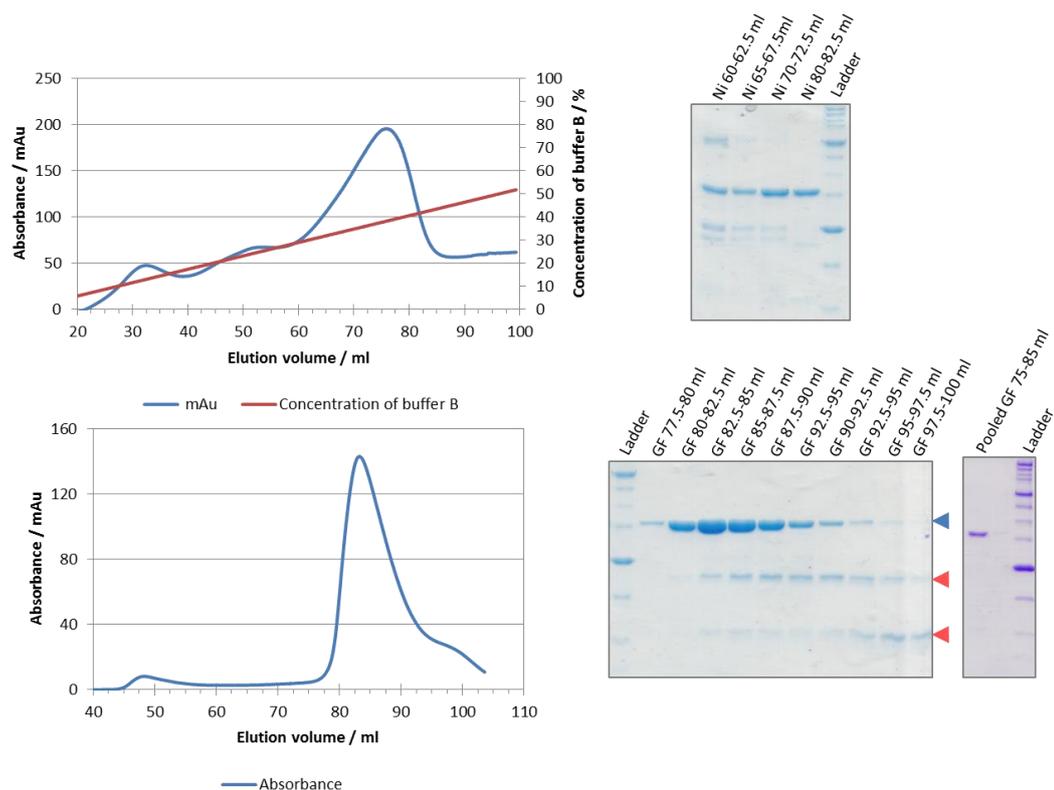


Figure 4.50. 1BKR-DUSPs purification. (A) Nickel column of 1BKR-DUSPs. The SDS-PAGE shows relatively pure protein in the peak at 75.1 ml. The later sample on SDS-PAGE shows some breakdown into the two components of the fusion. **(B)** The protein elutes at a monomeric elution volume for the protein. However, the gel filtration seems to show increased amounts of breakdown, even though there were protease inhibitors in the sonication buffer. The first four fractions of the peak were taken to avoid the breakdown products, and the final sample (right SDS-PAGE) shows relatively pure protein, with only small amounts of the breakdown products.

The solubility of the fusion and the profiles obtained from the chromatograms did not seem to alter when using either 150 mM NaCl, 20 mM Tris pH 7.5, 1% glycerol or 500 mM NaCl, 20 mM Tris pH 7.5, 5% ethylene glycol in the gel filtration buffer. Approximately 3.3 mg of protein was obtained from the nickel column for this fusion. The protein eluted at 83.5 ml from the HiLoad 16/60 Superdex 200 prep grade gel filtration column, giving it a predicted molecular weight of 40.5 kDa;

again larger than its actual molecular weight of 37.3 kDa, suggesting extended tag-protein structure. Due to aggregation during the concentration steps following the nickel column and gel filtration, and avoidance of breakdown-contaminated fractions, a total mass of 1.8 mg was obtained at a concentration of 3 mg/ml. The total yield of this protein is around 0.6 mg/L of culture. This sample contained a small amount of the breakdown products of the fusion (red arrows Figure 4.50). Crystallisation trials were performed but no crystals were observed; possibly due to the low concentration of protein in the drops.

4.5.4.3 **2GKG-DUSPs purification**

The 2GKG-DUSPs fusion expresses extremely well; almost comparable to the MBP fusion. The solubility of this construct was also similar; reaching approximately 10 mg/ml without visible aggregation. However, a peculiar trait of this fusion is that above approximately 8 mg/ml at 4 °C, the protein solution forms a semi-solid gel that returns to liquid upon warming. This process seems to have little effect on downstream properties of the purification as diluted nickel column fractions that are prevented from gelification gave the same gel filtration profiles and solubility as their gelified counterparts. This also led to a slow purification process because during concentration, the sample viscosity increased, slowing the rate of concentration itself. An example purification of 2GKG-DUSPs is given in Figure 4.51.

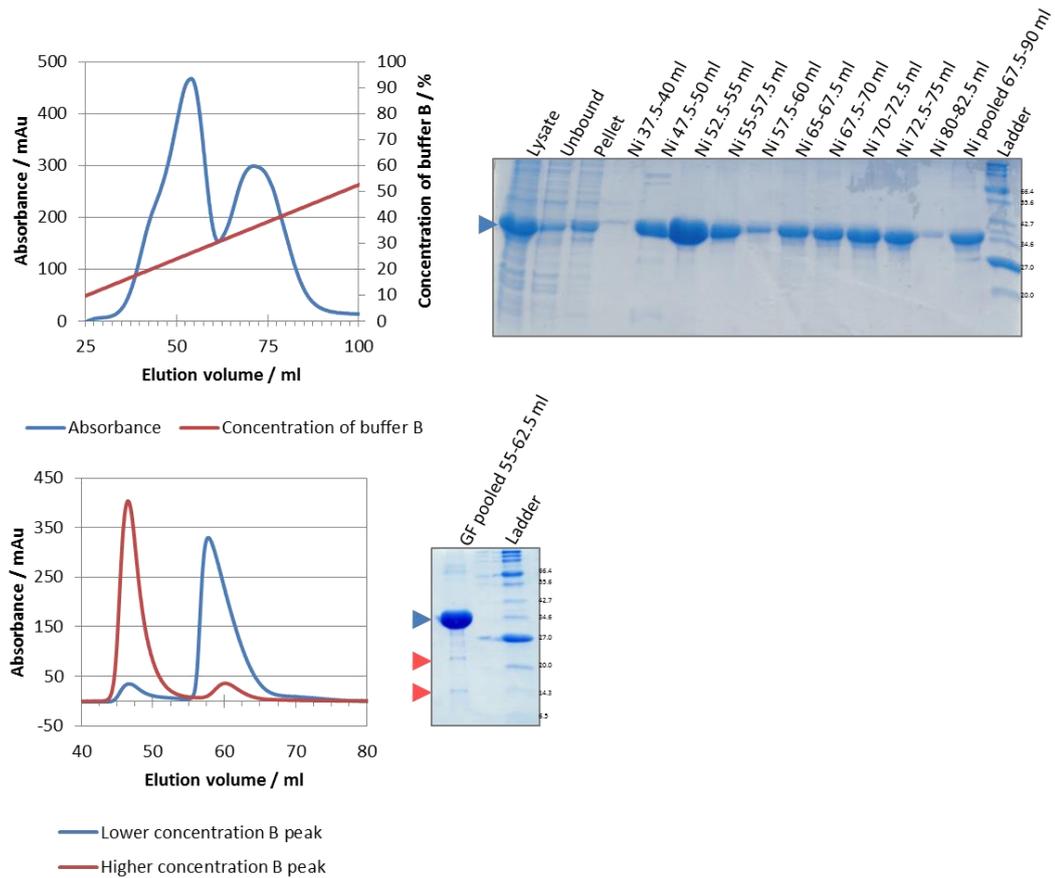


Figure 4.51. 2GKG-DUSPs purification. (A) Nickel column of 2GKG-DUSPs. The double peak of the nickel column chromatogram contains the fusion throughout. **(B)** The peak at 54.2 elutes at a monomeric volume on gel filtration (blue) whereas the 71.3 ml nickel peak elutes primarily in the void (red). The SDS-PAGE of the pooled gel filtration fractions show that the final sample contains the DUSP- and 2GKG-sized breakdown products, as do all the previous solubility tag purifications.

Like the 1BKR fusion, the solubility and the profiles obtained from the chromatograms of 2GKG-DUSPs did not seem to alter when using either 150 mM NaCl, 20 mM Tris pH 7.5, 1% glycerol or 500 mM NaCl, 20 mM Tris pH 7.5, 5% ethylene glycol in the gel filtration buffer. Total protein mass following the nickel column was approximately 60 mg. However, this protein elutes in two roughly equal peaks on the nickel column. One peak elutes from the column at 26.3% buffer B (146.2 mM imidazole) and the other elutes at 36.1% (193.3 mM

imidazole). To resolve these peaks enough for adequate gel filtration, a lower gradient of buffer B concentration must be used during the nickel column elution. Each peak was gel filtered separately and it can be clearly seen that one peak elutes primarily in the void and the other peak elutes as a monomer. The monomer protein elutes at 57.9 ml, giving it a predicted mass of 47.3 kDa; again higher than its actual mass of 38.0 kDa. The peak is relatively monodisperse, but similar to the His-tagged DUSPs and 1BKR-DUSPs, it trails to the right of the peak. Again, the final protein sample is contaminated with a small amount of lower molecular weight proteins, which are most likely breakdown products (red arrows Figure 4.51). Again by selecting the purest fractions of the gel filtration peak, protein was lost as consequence. The total mass of protein from fractions 57.5-62.5 ml was 2.25 mg at a concentration 2.74 mg/ml. Crystal trials were performed with this sample and crystals had formed after 9 days (Figure 4.52).

4.5.4.4 **2GKG-DUSP crystals**

The crystals for the 2GKG-DUSPs trials were cryo-cooled in their mother liquor with 20% glycerol prior to subjecting them to X-rays. The diffraction from these crystals was poor: they all had a similar trait of overlapping spots down one axis (Figure 4.53). Down the other axes, spots were better resolved but even in some case split spots and spots with poor profiles were observed. It was most likely that the spots were due to multiple mis-aligned crystal lattices due to the plate morphology of the crystal. Indexing software (XDS and iMosflm) could not accurately index these poor diffraction images and therefore the structure could not be solved.

These crystals were not reproducible using fresh protein purifications, nor were crystals formed by seeding with these crystals. In addition, SDS-PAGE analysis of other drops of similar conditions in the crystallisation trials that had no crystals showed the majority of the protein in the drop had broken down. This leads to speculation that the crystals formed were not of the fusion protein, but were possible just the tag alone. It is less likely to be the DUSPs because previous trials with concentrations lower than used in these trials produced only precipitate. In addition, a UV microscope was used to image a drop with some crystal fragments in. The image using bright-field microscopy showed two visible crystal fragments. However, the UV image showed no fluorescence of the fragments. This confirms that there are no tryptophans in the protein forming the crystal. The fusion protein and the DUSP domains have tryptophans, so should be observed by the microscope. However, it is likely that as no tryptophans are present in the 2GKG protein, then the crystal are formed from 2GKG itself.

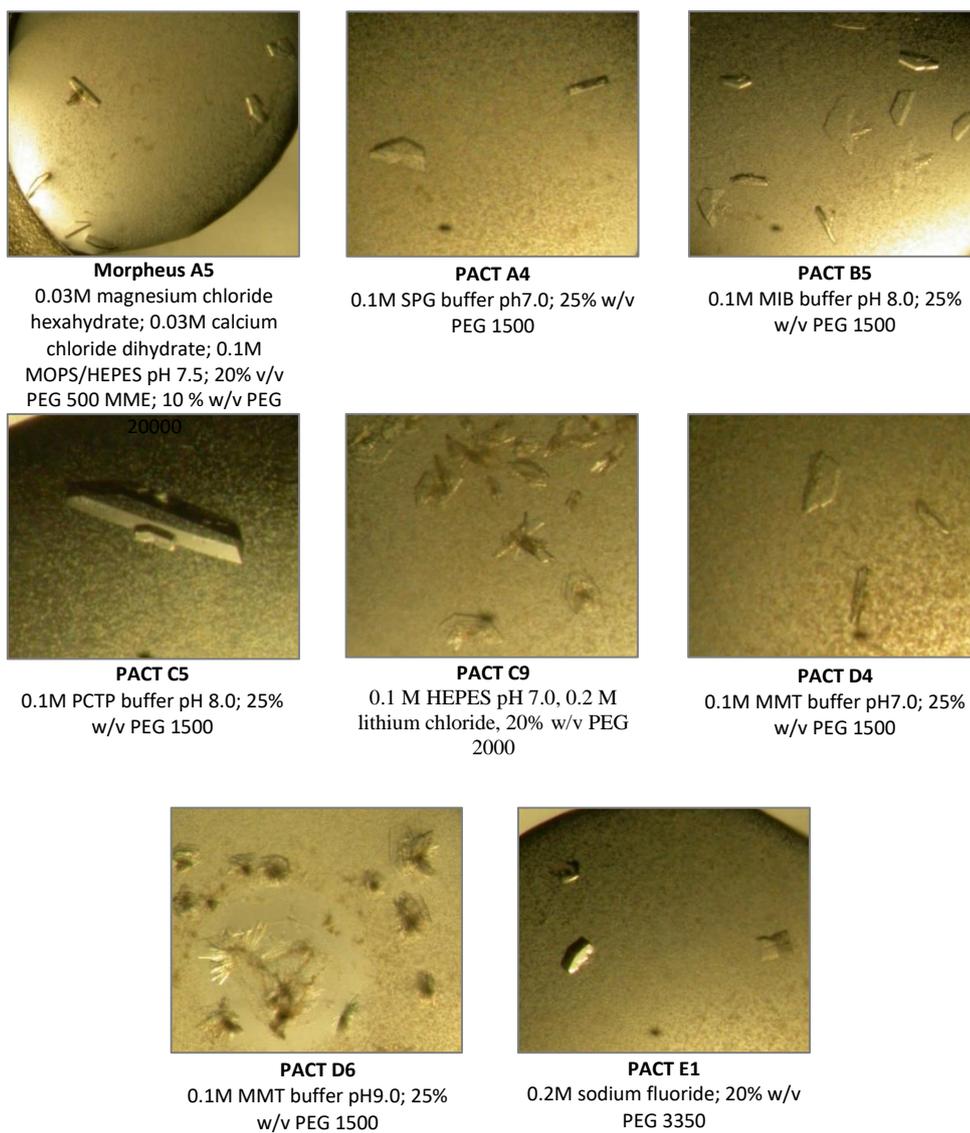


Figure 4.52. Crystals from 2GKG-DUSP crystallisation trials. The crystals that were obtained from the 2GKG-DUSP screens are shown with their screen and condition.

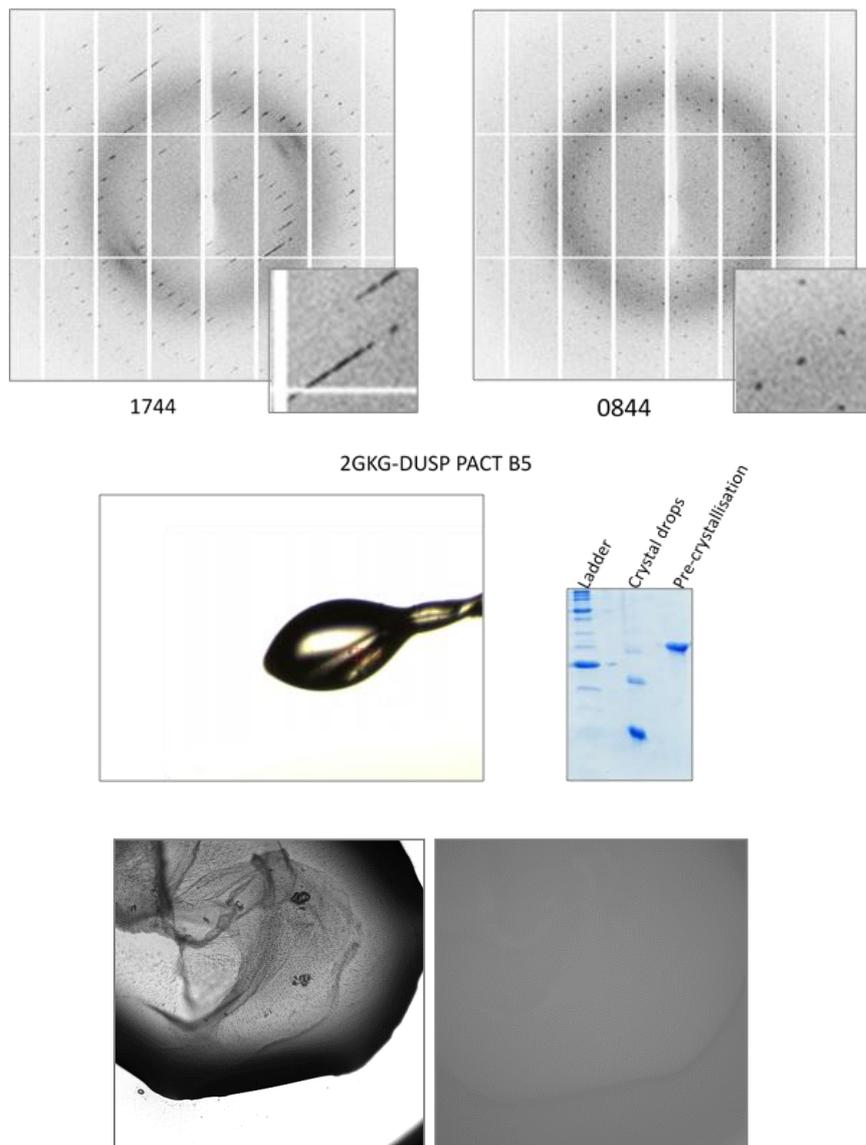


Figure 4.53. Diffraction and assessment of 2GKG-DUSP crystals. (A) The diffraction images of the 2GKG-DUSP crystal from condition PACT B5. The images were taken with 0.1° oscillation at a detector distance of 267.02 mm. The left image shows multiple overlapping and smeared spots (image 1744; Φ angle=174.4°). The diffraction from the image at 90° rotation of the crystal shows spots with acceptable profiles (image 844; Φ angle=84.4°). (B) Crystal loop photo prior to X-ray exposure. (C) The SDS-PAGE of similar conditions in the crystallisation drops show breakdown producing proteins of equal size to the DUSP domains and the tag alone. (D) A comparison of the light field image (left) and UV image (right) shows no fluorescence of the crystal fragments (white arrows in light field image) in the UV image. As 2GKG has no tryptophans, and the DUSP domains do, the crystal fragments should fluoresce if they are formed from the fusion or DUSP domains breakdown.

4.5.5 Comparison of novel tags to other tags

In comparison to other tags, these two novel tags have their advantages and disadvantages. As both proteins (but especially the 2GKG tag) were soluble alone, they showed promise that they could work as effective solubility tags. Indeed, the novel tags worked well in regards to increasing the solubility of the double DUSP construct. The 2GKG tag was especially effective at increasing solubility, producing a protein almost as soluble as that of MBP. The 1BKR tag only marginally increased the solubility of the DUSPs to that of what was achieved with buffer modification. Further analysis would be required to assess whether the fusions were more soluble than other tags, such as GST or SUMO.

As yet, the tags themselves have not been utilised for affinity purification, so the cloning vector provides an N-terminal hexahistidine tag. MBP and GST are routinely used as both solubility and affinity tags, which gives multiple functions for their use as fusion partners. It is possible that the tags could be utilised in this manner. As the calponin homology domain of beta spectrin (1BKR) binds to EPB41 (also known as 4.1R), an erythrocytic cytoskeletal protein [335], affinity columns could be produced using this protein and a method to purify 1BKR-tagged proteins could be created. However, whether this would be more cost-effective and useful in comparison to using the C-terminal his tag would require further investigation. FrzS of *Myxococcus xanthus* has, as yet, no known ligand [336] so it is unclear whether it could be used as an affinity tag in the future. A common problem with MBP is that it forms oligomers, and thus can compromise the purification of monodisperse protein (observed routinely in the lab of the author). This was seen in the purification with the MBP-DUSP domains fusion, but not with the 1BKR or 2GKG fusions. This meant that the purification of the novel tag DUSP domain fusions was much simpler than the MBP-fusion. Whether this observation would be seen in other 2GKG and

1BKR fusion partners would also require further investigation. Using the current method of metal affinity chromatography, the purification protocol is cheap and effective.

Like most other tags, other than SUMO and ubiquitin, cleavage can only be obtained by using a specific sequence between the tag and protein of interest. However, as the premise of these novel tags was to produce a crystallisable solubility tag, this attribute is not a major issue. With regards to the crystallisability of the new tags compared to other tags, it is still unclear. Only the DUSP domains protein was attempted in crystal trials and crystals were not produced with any of the MBP or novel tag fusions. This limited dataset cannot be used to define whether enhanced crystallisation is observed with the novel fusion proteins. Ongoing work is being performed to investigate the crystallisability of the fusions with other proteins.

Large protein tags have been useful for phasing of the crystal diffraction data using molecular replacement. This has been achieved with both GST [337] and MBP [338] fusion structures. The novel tags were designed to be small to aid crystallisation. When solving a crystal structure, a technique known as molecular replacement can be used to locate proteins within the asymmetric unit. Using molecular replacement to find the tag first, and then to use this to solve the structure of the fusion partner is possible with large tags. As the tags are relatively small compared to other tags, it would require diffraction data from fusion-containing crystals to assess whether this would be possible.

4.6 Summary of crystallisation trials

Given below are the summaries of the crystallisation trials for all USP20 constructs with relatively small tags (His-tags and cleavage tags; Table 4.7) and those expressed as fusions with solubility tags (Table 4.8) As of writing this thesis, no protein crystals containing the USP20 domains have been obtained. However, crystals may take up to a year to appear in some cases. Many of the crystal trials below have well exceed one year, but many are not so old, so they may yield crystals in the future.

Table 4.7. Summary of crystallisation trials with His-tagged USP20 constructs.

Construct	Buffer	Components added prior to crystal trial	Screen	Concentration (mg/ml)	Temp (°C)
Znf-UBP 1-101	150 mM NaCl, 20 mM Tris pH 7.5, 1% glycerol	0.5 mM TCEP	J+, PA, PE	8 + 4	20
	150 mM NaCl, 20 mM Tris pH 7.5, 1% glycerol, 1mM ZnCl ₂	0.5 mM TCEP	J+, PA, PE	8 + 4	20
	500 mM NaCl, 20 mM Tris pH 7.5, 1% glycerol	0.5 mM TCEP	CS, IN, MI	8 + 4	20
	500 mM NaCl, 20 mM Tris pH 8.0, 1% glycerol	0.5 mM TCEP	J+, PA, MO	8 + 4	20
	500 mM NaCl, 20 mM Tris pH 8.0, 1% glycerol	0.5 mM TCEP 10 mM L-arginine	J+, PA, MO	8 + 4	20
	500 mM NaCl, 20 mM Tris pH 8.0, 1% glycerol	0.5 mM TCEP, 100 mM MgCl ₂	J+, PA, MO	8 + 4	20
	500 mM NaCl, 20 mM Tris pH 8.0, 1% glycerol	0.5 mM TCEP	J+, PA, MO CS, IN, MI	7.5 + 3.25	10
	150 mM NaCl, 20 mM Tris pH 7.5, 1% glycerol	0.5 mM TCEP	J+, PA, MO	10 + 5	10
Znf-UBP 1-108 un-cleaved	300 mM NaCl, 20 mM Tris pH 8.0, 2% glycerol	0.5 mM TCEP	J+, PA, MO	8 + 4	20
Znf-UBP 1-108 cleaved	500 mM NaCl, 20 mM Tris pH 8.0, 2% glycerol	0.5 mM TCEP	J+, PA, MO, CS, MI, PE	8 + 4	20
	500 mM NaCl, 20 mM Tris pH 8.0, 2% glycerol	0.5 mM TCEP, 1:1000 thermolysin	J+, PA, MO	8 + 4	20
DUSPs 686-894	500 mM NaCl, 50 mM Tris pH 8.0, 1% glycerol, 5% ethylene glycol, 1 mM β-mercaptoethanol, ~260 mM imidazole pH 8	0.5 mM TCEP	J+, PA, MO	3 + 1.5	10
	500 mM NaCl, 20 mM Tris pH 8.0, 1% glycerol, 5% ethylene glycol	0.5 mM TCEP	J+, PA, MO, CS, MI, PE	1 + 0.5	10
	500 mM NaCl, 20 mM Tris pH 8.0, 1% glycerol, 5% ethylene glycol	0.5 mM TCEP	J+, PA, MO, CS, MI, PE	1 + 0.5	20

J=JCSG+; PA= PACT suite; MO=Morpheus screen; PE=Pegs II suite; CS=Compass suite; IN=index; MI=MIDAS

Table 4.8. Summary of crystallisation trials with solubility-tagged USP20 constructs.

Construct	Buffer	Components added prior to crystal trial	Screen	Concentration (mg/ml)	Temp (°C)
TF-USP20FL	500 mM NaCl, 20 mM Tis pH 7.5, 2% glycerol	0.5 mM TCEP	J+, PA, MO	3 + 1.5	20
MBP-DUSPs (MBP→GF)	150 mM NaCl, 20 mM Tis pH 7.5, 1% glycerol, 40 mM maltose	0.5 mM TCEP	J+, PA, MO	10 + 5	10
	500 mM NaCl, 20 mM Tis pH 7.5, 1% glycerol, 40 mM maltose	0.5 mM TCEP	J+, PA, MO, MG, PC, MI	8 + 4	10
	500 mM NaCl, 20 mM Tis pH 7.5, 1% glycerol, 40 mM maltose	0.5 mM TCEP	J+, PA, PE	8 + 4	4
MBP-DUSPs (MBP→AX→GF→AX)	~150 mM NaCl, 20 mM Tis pH 7.5, 1% glycerol	0.5 mM TCEP	J+, PA, PE, MG, AS	8 + 4	10
1BKR-DUSPs	150 mM NaCl, 20 mM Tis pH 7.5, 1% glycerol	0.5 mM TCEP	J+, PA, MO	5 + 2.5	10
	500 mM NaCl, 20 mM Tis pH 7.5, 5% ethylene glycol	0.5 mM TCEP	J+, PA, PE	5 + 2.5	10
2GKG-DUSPs	150 mM NaCl, 20 mM Tis pH 7.5, 1% glycerol	0.5 mM TCEP	J+, PA, MO	10 + 5	10
	300 mM NaCl, 20 mM Tis pH 7.5, 1% glycerol, 5% ethylene glycol	0.5 mM TCEP	J+, PA, MO	8 + 4	10
	500 mM NaCl, 20 mM Tis pH 7.5, 1% glycerol	0.5 mM TCEP	PA, MO, PE, MG, PC	8 + 4	10
	500 mM NaCl, 20 mM Tis pH 7.5, 1% glycerol	0.5 mM TCEP, protease inhibitors	J+, PA, MO	8 + 4	10
	500 mM NaCl, 20 mM Tis pH 7.5, 1% glycerol	0.5 mM TCEP, seeding crystals	J+, PA, MO	8 + 4	10

J=JCSG+; PA= PACT suite; MO=Morpheus screen; PE=Pegs II suite; MI=MIDAS; MG=MemGold; PC=Procomplex; Ammonium sulphate

5 Results: Interaction studies

Mapping USP20 protein interactions

5.1 Investigating USP20 interactions with yeast two hybrid

In order to dissect interactions of known proteins with USP20, a yeast two hybrid (Y2H) system was used. The system utilises LexA-bait and VP16-prey fusions to assess interaction. USP20 was divided into the Znf-UBP domain (residues 1-101/1-108), catalytic domain (residues 147-685) and DUSP domains (residues 686-894); based on the domain analysis for bacterial expression. These domains were cloned into both pBTM116mod and pASV3mod vectors to produce LexA and VP16 fusions, respectively. When the USP20 domains were expressed as LexA fusions, the catalytic domain did not express (assessed by Western blot; Figure 5.14), and the DUSP domains caused transcription of the β -galactosidase gene, irrespective of whether the VP16-fusion was present. This meant that only the Znf-UBP domain-LexA fusion was viable for use in the assay. As this was noticed in the initial USP20- β -arrestin-1 assays, these baits were not used in further Y2H experiments. Thus, all further interaction assays were conducted with USP20 domains in the VP16 prey vector and other proteins in the LexA bait vector.

5.1.1.1 DUSP domains as a transcriptional activator

Activation of transcription, as seen by the LexA-DUSP domains fusion, is a common issue with the Y2H system [339, 340]. However, this autoactivation raises an important question: is its ability to activate transcription physiologically relevant?

Ptashne and colleagues identified numerous proteins that acted as transcriptional activators [341-343]. A common occurrence in these proteins is multiple repeats of acidic residues that increase the negative charge in a particular region of the protein. Indeed, the amino acid sequence of the DUSP domains shows these acidic repeats (Figure 5.1). Also, these residues are all present in the DUSP domains of USP20 from all reviewed mammalian USP20 sequences except for *Ovis aries* (missing one Glu residue), and also in the earlier divergent *Danio rerio*. Using JPred 4 to predict secondary structure and solvent exposure, the DUSP domains prediction shows multiple solvent exposed doublets of acidic residues. It is possible that these residues form an area of negative charge capable of activating transcription. To answer whether this is a physiological role of the DUSP domains *in vivo* would require further investigation, but it is conceivable when assessing the amino acid sequence and structural predictions alongside the Y2H data.

109-146 and 895-914; the disordered region between the Znf-UBP and catalytic domain, and the DUSP domains and the C-terminus, respectively. For this reason, lack of interaction seen in the assays would not necessarily indicate a lack of interaction between the two proteins. On the contrary, the proteins were chosen for yeast two hybrid because they have been shown to bind (as specified below) to USP20 or USP33. In addition, steric issues may arise because of the nature of the fusion proteins, which could also prevent true physiological interactions.



Figure 5.2. Review of USP20 domain architecture. The Znf-UBP domain is shown in blue, The USP domains in red and the DUSP domains in yellow

5.1.1.3 USP20 domains and β -arrestin-1

The USP33- β -arrestin-1 and USP33- β -arrestin-2 interactions were shown by Shenoy *et al.* [344] through Y2H and immunoprecipitation experiments. USP20 was also seen to bind β -arrestin-2 [97], also through immunoprecipitation. As USP20 and USP33 both interact with the β 2-adrenergic receptor (β 2-AR) [88], USP33 binds equally to both β -arrestin-1 and β -arrestin-2, it was expected that USP20 would also interact with the receptor's accessory protein, β -arrestin-1. To test this hypothesis using Y2H, full length β -arrestin-1 was used as bait and prey to individual USP20 domains, which were also used as bait and prey. As explained before, the Znf-UBP domain was the only viable assay of all three domains when USP20 domains were expressed as bait fusions. This assay showed no increase in signal when both bait Znf-UBP and prey β -arrestin-1 were expressed than when prey β -arrestin-1 was substituted by empty

pASV3mod control plasmid ($p=0.4578$) (Figure 5.3A). As the LexA- β -arrestin-1 fusion produced a low background signal in the control assay, it allowed testing of all three USP20 domains as VP16 fusions. Assays with each USP20 prey fusion showed no increase in signal when comparing expression of bait and prey fusions with bait and control (Figure 5.3B). Therefore, in this assay no interaction could be detected between USP20 domains and β -arrestin-1 (VP16-USP20-Znf-UBP and LexA- β -arrestin-1, $p = 0.0657$; VP16-USP20-Catalytic domain and LexA- β -arrestin-1, $p = 0.4720$; VP16-USP20-DUSPs and LexA- β -arrestin-1, $p = 0.4366$).

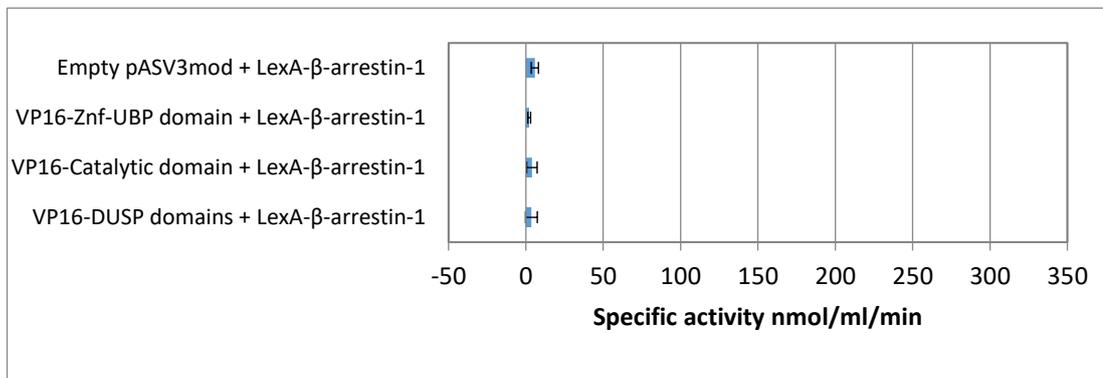
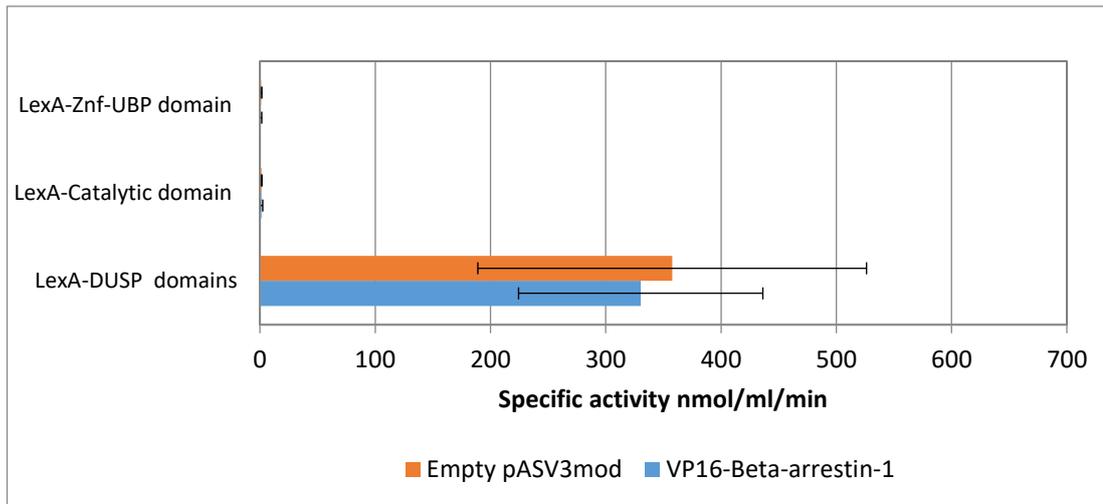


Figure 5.3. β-arrestin-1 yeast two-hybrid (A) With β-arrestin-1 in pASV3mod (VP16 fusion) and USP20 domains in pBTM116mod (LexA fusion), no binding is observed for the Znf-UBP domain or catalytic domain. Activity is observed for the DUSP domains, but is also observed when the DUSP domains and empty VP16, showing that the DUSP domains auto-activate transcription in the Y2H assays. **(B)** Upon reversal of the domains, no auto-activation is observed by β-arrestin-1, however no activity is observed for any of the assays. Therefore Y2H assays show no interaction between β-arrestin-1 and the USP20 domains.

5.1.1.4 USP20 domains and TRAF6

All following interaction studies with the Y2H use USP20 domains as VP16 fusions (pASV3mod) and test proteins as LexA fusions (pBTM116mod). TRAF6 has been shown to interact with USP20 using immunoprecipitation experiments [151, 345]. In these experiments only full length enzymes were used, so there is no information about domains or residues involved in the interaction. These Y2H experiments using full length TRAF6 bait and USP20 prey showed no interaction (Figure 5.4). None of the assays using individual domains and TRAF6 showed an increase in activity over that of the control with TRAF6 and LexA alone (VP16-USP20-Znf-UBP and LexA-TRAF6, $p = 0.1721$; VP16-USP20-Catalytic domain and LexA-TRAF6, $p = 0.2029$; VP16-USP20-DUSPs and LexA-TRAF6, $p = 0.3017$).

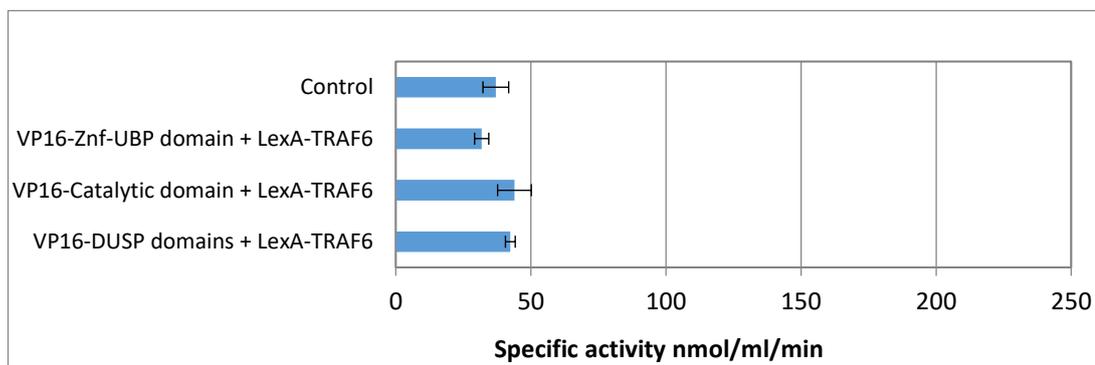


Figure 5.4. TRAF6 yeast two-hybrid. High background signals are observed for the TRAF6 assays. No interaction between TRAF6 and the USP20 domains is observed.

Interestingly, a paper was published on the binding of TRAF6 and USP20 after these assays were performed [97]. They found that the presence of β -arrestin-2 was required to act as a scaffold in order for USP20 to bind to TRAF6. This may be a reason for the lack of binding observed in this assay, as this scaffold is not present.

5.1.1.5 USP20 domains and RAD17

RAD17 has been shown to interact with USP20 by Shanmugam *et al.* [182] with immunoprecipitation assays. They showed that full length USP20 would pull down with RAD17, and also broke USP20 down into four sections and used those to map the interaction. Although the paper never specified the residues, it appears that USP20 was broken down into the following components: the Znf-UBP and disordered loop; the first half of the catalytic domain; the second half of the catalytic domain; and both DUSP domains. In their study, the Znf-UBP domain and loop clearly immunoprecipitated with RAD17, suggesting the majority of the interaction is mediated by this region. Both halves of the catalytic domain showed a weak interaction also, indicating that the core catalytic domain or disordered insert, also contributes to this association.

As some mapping of this interaction is published, Y2H was used to confirm that these interactions were binary and to characterise the interaction with truncation and mutational studies. No interaction could be seen using Y2H assays using only the Znf-UBP domain (1-108, excluding the following disordered loop) and whole catalytic domain (147-685), (VP16-USP20-Znf-UBP and LexA-RAD17, $p = 0.3735$; VP16-USP20-Catalytic domain and LexA-RAD17, $p = 0.4904$) (Figure 5.5). Also, no interaction was seen by the DUSP domains (a statistically significant decrease in activity was observed: VP16-USP20-DUSPs and LexA-RAD17, $p = 0.0294$). Therefore, this Y2H assay could not be used to investigate the interaction between RAD17 and USP20.

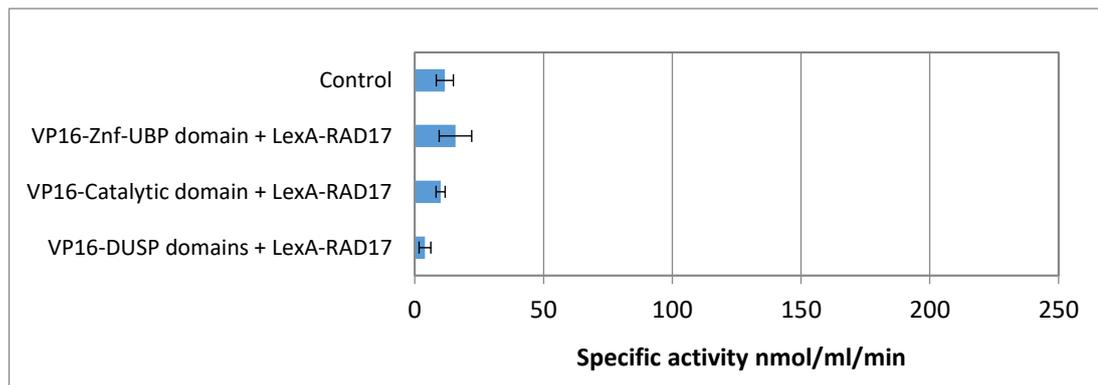


Figure 5.5. RAD17 yeast two-hybrid. No interaction between TRAF6 and the USP20 domains is observed.

5.1.1.6 USP20 domains and PLK1

USP20 and PLK1 were detected as interacting proteins by Sowa *et al* [172] in a global proteomic analysis identifying DUB interacting proteins. Shanmugam *et al* [182] also showed that depletion of USP20 led to a decrease in PLK1 levels. The exact mechanism of this effect was not explained, and although they stated that an interaction was observed between USP20 and PLK1 through immunoprecipitation experiments, the data were not shown. This means that yeast two hybrid would be useful in concluding that there is a true binary interaction, and can be used to characterise it. PLK1 comprises an N-terminal catalytic domain and two C-terminal polobox domains (Figure 5.6).

The interaction between USP20 prey domains and full length PLK1 as bait was investigated using Y2H (Figure 5.67). A positive interaction was seen by the DUSP domains in this assay, with a mean 5.3 times greater than the control prey plasmid (VP16-USP20-DUSPs and LexA-PLK1, $p = 0.0009$). The Znf-UBP domain and catalytic domain showed no interaction with PLK1 (VP16-USP20-Znf-UBP and LexA-PLK1, $p =$

0.8948; VP16-USP20-Catalytic domain and LexA-PLK1, $p = 0.6965$). This novel finding suggests that the DUSP domains of USP20 interact with PLK1.



Figure 5.6. PLK1 domain architecture. A schematic of the domain architecture of PLK1 is shown. The kinase domain is shown in green, the D-box in dark grey, the poloboxes in yellow, and non-domain regions in light grey.

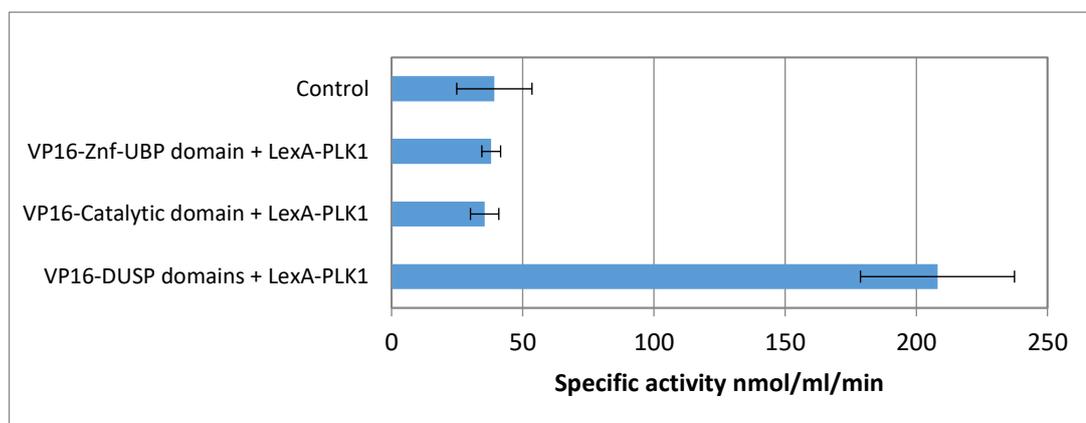


Figure 5.7. PLK1 yeast two-hybrid. High background signals are observed with PLK1. However, the DUSP domains assay shows significantly higher activity than the control assay with PLK1 alone. This suggests that there is an interaction between the DUSP domains and PLK1. No activity is observed in assays with other USP20 domains. Therefore, there is no interaction between the other domains and PLK1 in yeast cells.

5.1.1.7 The USP20-PLK1 interaction is mediated by the DUSP domains and polobox domains

To further analyse the interaction of USP20 DUSP domains and PLK1, the DUSP domains and PLK1 were dissected. The VP16-DUSP domains fusion contains both DUSPs, which should fold to produce two independent domains. It is unknown whether the interaction is mediated by one of the DUSP domains, or both are

required. Therefore, this construct was mutated so that the residue corresponding to USP20 780 coded for a stop (A780X). The remaining fusion construct included a VP16 fusion with USP20 residues 686-789; the single, N-terminal DUSP domain. The full length PLK1 can be divided into three independent domains: the catalytic domain and two poloboxes. To remove the C-terminal domains, PLK1 residues 408 and 508 were mutated to stop codons. This removed either just the terminal polobox (L508X; Figure 5.8) or both poloboxes (I408X). Yeast two-hybrid assays involving truncation mutants are shown in Figure 5.9.

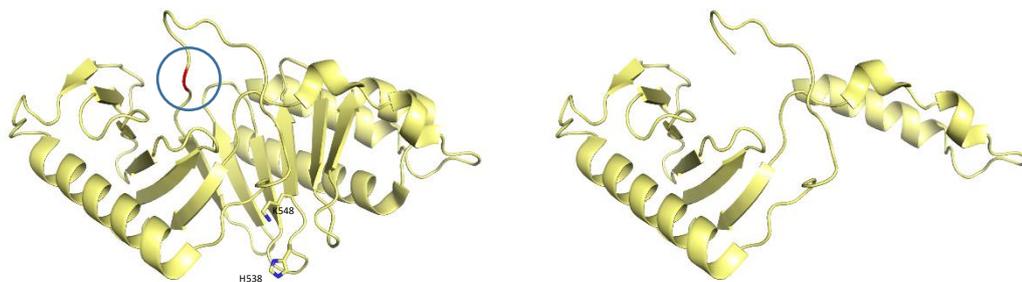


Figure 5.8. Polobox truncation. The structure of the poloboxes (PDB code 1UMW) is shown (left). The H538 and K540 pincer residues are shown. Circled is residue 508, coloured red, which was mutated to a stop codon. The right image shows the structure with the residues 509-603 removed.

Following removal of the terminal DUSP domain of USP20, the interaction with full length PLK1 is lost completely (VP16-USP20-DUSPs 686-894 and LexA-PLK1 vs VP16-USP20-DUSP 686-789 and LexA-PLK1, $p = 0.0004$); reducing the signal 17.4 times, down to background levels (empty pASV3mod and LexA-PLK1 vs VP16-USP20-DUSP 686-789 and LexA-PLK1, $p = 0.7425$). This suggests that the terminal DUSP or both DUSP domains are required for interaction with PLK1.

Removal of all poloboxes, or the terminal polobox, impeded the interaction completely, again returning the signal to background level compared to full length PLK1 (VP16-USP20-DUSPs 686-894 and LexA-PLK1 vs VP16-USP20-DUSPs 686-894 and LexA-PLK1 Δ polobox2, $p = 0.0003$; empty pASV3mod and LexA-PLK1 Δ polobox2 vs VP16-USP20-DUSPs 686-894 and LexA-PLK1 Δ polobox2, $p = 0.2642$; VP16-USP20-DUSPs 686-894 and LexA-PLK1 vs VP16-USP20-DUSPs 686-894 and LexA-PLK1 Δ poloboxes, $p = 0.0004$; empty pASV3mod and LexA-PLK1 Δ poloboxes vs VP16-USP20-DUSPs 686-894 and LexA-PLK1 Δ poloboxes, $p = 0.5785$). In addition, as expected from these results, no interaction was seen between truncated DUSPs and either of the PLK1 truncations.

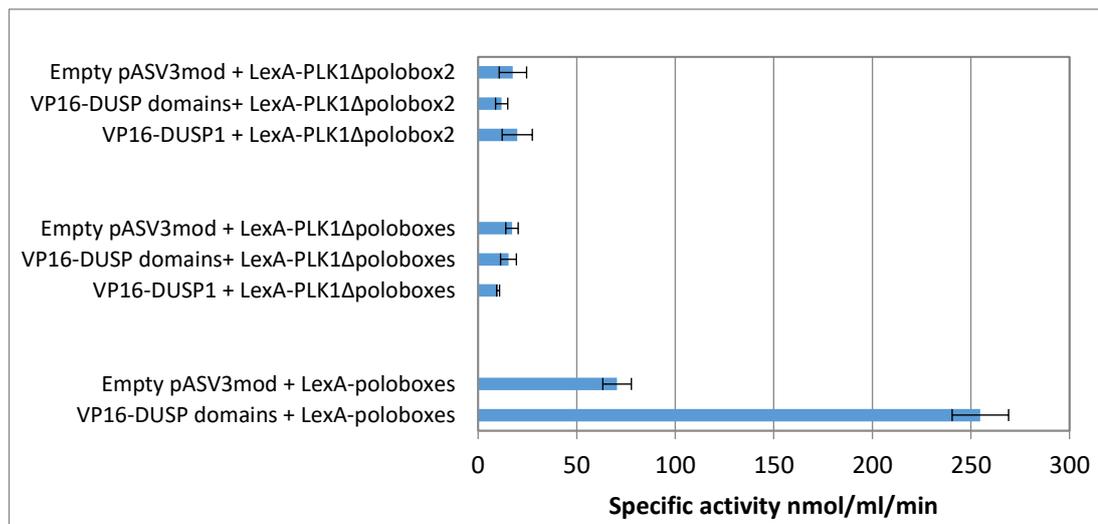


Figure 5.9. PLK1-USP20 yeast two-hybrid domain truncation. Removing either the one or both poloboxes from the C-terminus of PLK1 abolishes the interaction with the DUSP domains. Also, expressing the poloboxes as a VP16 fusion also shows interaction. This indicates that the poloboxes alone mediate the interaction with PLK1 and the DUSP domains of USP20 and that the interaction is dependent on the presence of polobox 2. Truncation of the DUSP domains was also investigated. Truncation of the 2nd DUSP domain also prevents interaction of the poloboxes.

To confirm that the interaction is mediated solely by the poloboxes, residues 367-603 of PLK1 (the poloboxes) were cloned into pBTM116mod and used as bait for the DUSP domains in the Y2H assay. A similar activity was seen to that of the full length PLK1, confirming that the minimal interacting portion of PLK1 lies with the poloboxes (VP16-USP20-DUSPs 686-894 and LexA-poloboxes vs empty pASV3mod and LexA-poloboxes, $p = 0.0001$).

Studies have shown that the polobox domains act as a protein-recruitment module for PLK1 [346]. The polobox β -sheet sandwich generates a phosphopeptide sensing pocket, utilising critical histidine and lysine residues to detect the phosphorylated serine/threonine of the binding protein. Mutations in these residues completely abrogate PLK1s function by preventing proper substrate binding. Much work has been done to elucidate the motif for the optimum phosphopeptide for the poloboxes: Ser-[pSer/pThr]-[Pro/X] [212, 213]. The binding of phosphopeptides containing this motif to the poloboxes is crucially dependent on the first serine. Without this residue, binding is almost completely abolished.

Using the motif S-[pS/pT]-[P/X], the Human Protein Reference Database PhosphoMotif Finder [347] identifies 10 sites in USP20 that contain these motifs. Only one of these motifs is found within the DUSP domains (685-687). However, in the yeast two-hybrid construct, the first serine is not present because the DUSP domains start on residue S686. Also, it would not be expected for the human proteins to be phosphorylated in the yeast system unless there are closely related yeast homologue kinases that could do this.

Table 5.1. Polobox binding motifs in USP20.

Position in protein	Sequence	Phosphorylated PBD motif observed?	References
60-64	HSTIH	No	
102-106	GSSSK	No	
287-291	DSSSD	No	
303-307	GSSQA	Yes	pS305 [348]
370-374	RSSSP	Yes	pS372 [349]; pS373 [348, 350-356]
387-391	RSSSR	Yes	pS390 [357]
405-409	LSSSP	Yes	pS407 [351, 358]; pS408 [348, 350, 352, 359-361]
465-469	VSTTV	No	
684-688	KSSEE	No	

Residues that are confirmed by experimental evidence to be phosphorylated are shown in blue.

Without the motif in the DUSP domain construct, Zdock was used to give preliminary data as to where the DUSPs might bind to the poloboxes (Figure 5.10). Docking of a PHYRE2 predicted homology model of each DUSP domain on a structure of the poloboxes (PDB code: 1UMW) provided multiple docking solutions for each DUSP domain. All solutions were observed to be binding in the known phosphopeptide binding cleft. Interestingly, the top docking solution of the second DUSP showed aspartic acid 841 interacting with the critical phospho-sensing histidine. It is possible to hypothesise that in the interaction between the DUSP domains and poloboxes, the negative charge on the aspartic acid could mimic the negative charge of a phospho-serine/threonine, if oriented correctly. To investigate whether the known phospho-sensing residues are involved in the interaction, the Y2H constructs were mutated using site-directed mutagenesis. In the LexA-poloboxes construct, H538 was mutated to an alanine and the partially buried K540 was mutated to a methionine. Methionine was chosen because its side chain is a linear chain, like that of lysine, but does not carry the positively charged head, which is critical for sensing the negative charge of the phosphate and forming hydrogen bonds with oxygen. In

addition, D841 in the VP16-DUSPs construct was mutated to arginine in order to reverse the charge on this residue. If it is mimicking the negative phosphate, then charge reversal should prevent binding.

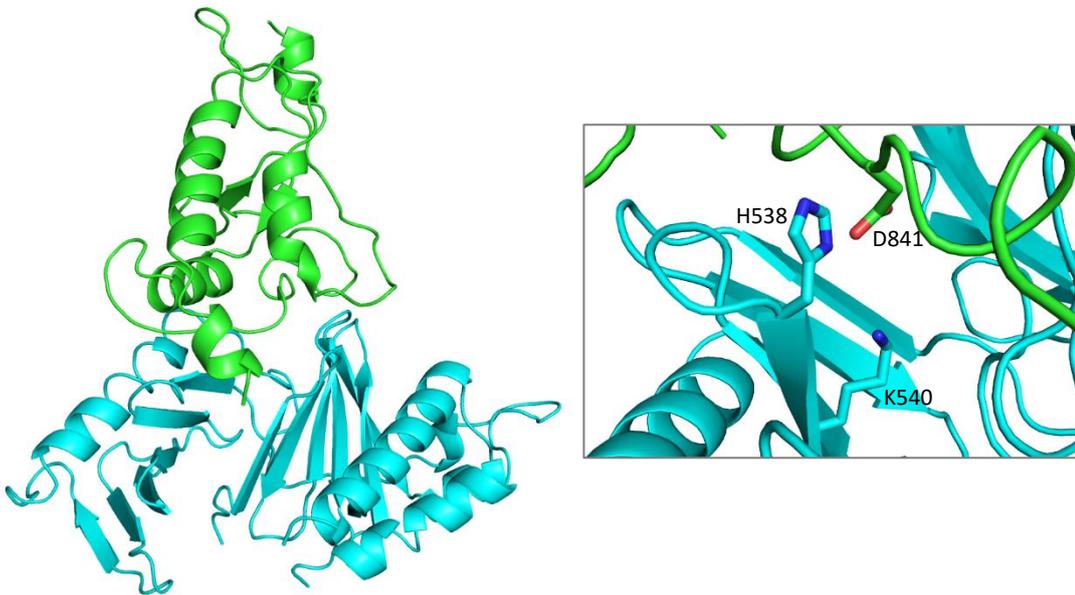


Figure 5.10. Docking studies of DUSP domains and PLK1 poloboxes. Zdock was used to dock homology DUSP domain models onto the poloboxes. All predictions docked the DUSP domains onto the cleft that is already known to bind phosphorylated peptides. The top prediction (**A**) showed that D841 may act as a phospho-mimic and could mediate the interaction without phosphorylation of the DUSP domains. USP20 is shown in green. The poloboxes of PLK1 are shown in blue. PLK1 H538 and K540, and USP20 D841 are shown in stick form in **B**.

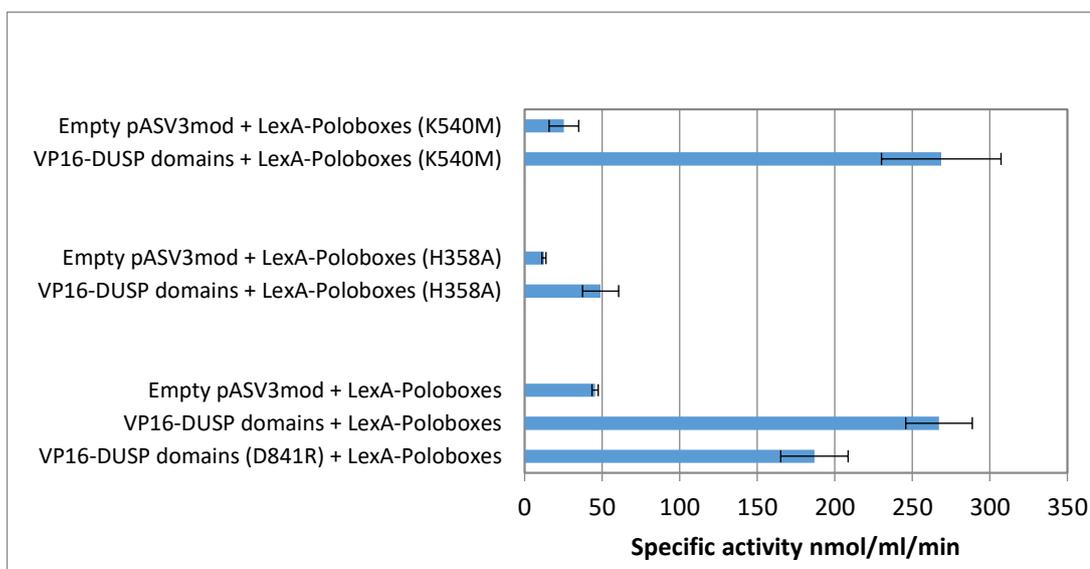


Figure 5.11. PLK1-USP0 yeast two-hybrid mutation assays. The K540M mutation had no significant effect on the interaction between the DUSP domains and the poloboxes. The H538A mutation massively reduced the specific activity in the assay. This indicates that this residue is integral to the interaction. However, some residual activity is observed, so other residues must be involved. The D841R mutant seems to reduce the signal slightly, but significantly. This suggests that the interfaces that the docking predicted maybe correct. However, D841 only contributes a minor component of the overall interaction, and most likely does not act as a phospho-mimic as K540 is not involved. This suggests that the phosphor-sensing cleft of the poloboxes can also detect other non-canonical and un-phosphorylated motifs.

Yeast two-hybrid assays involving PLK1 pincer residues H358 and K540, and USP20 D841 are shown in Figure 5.11. The Y2H assay shows that mutation of the lysine residue has no effect on binding of the poloboxes to the DUSPs (VP16-USP20-DUSPs 686-894 and LexA-poloboxes vs VP16-USP20-DUSPs 686-894 and LexA-poloboxesK540M, $p = 0.9578$). However, mutation of the histidine causes a 5.5 fold decrease in signal from the assay (VP16-USP20-DUSPs 686-894 and LexA-poloboxes vs VP16-USP20-DUSPs 686-894 and LexA-poloboxesH538A, $p = 0.0001$). This suggests a severe loss of binding between the DUSPs and the poloboxes. However,

the signal is still nearly four-fold higher than the background signal, indicating that binding is not 100% abolished (VP16-USP20-DUSPs 686-894 and LexA-poloboxesH538A vs empty pASV3mod and LexA-poloboxesH538A vs, $p = 0.0057$). Mutation of the aspartic acid residue in the DUSP domains reduces the signal by nearly 1.5 times (VP16-USP20-DUSPs 686-894D841R and LexA-poloboxes vs VP16-USP20-DUSPs 686-894 and LexA-poloboxes, $p = 0.0104$). This is a small decrease, but none-the-less, statistically significant. It is still well above the signal of the background inferring that binding between the domains is still occurring but slightly diminished (VP16-USP20-DUSPs 686-894D841R and LexA-poloboxes vs empty pASV3mod and LexA-poloboxes, $p = 0.0002$). The findings from the mutant constructs suggest that the region of PLK1's poloboxes that confers phosphopeptide specificity is involved in the interaction with the DUSP domains of USP20. It is likely that the aspartic acid is also involved, possibly contributing some minor interactions, but is certainly not a critical residue for binding.

A summary of the Y2H interactions with USP20 and PLK1 are shown in Figure 5.13.



Figure 5.12. PLK1-USP20 yeast two-hybrid summary. A summary of the PLK1-USP20 interactions is given. For each interaction studied by Y2H, schematics are given for the regions of the proteins present. Also, the activity produced by each assay is given in the centre of the diagram. ‘+++’ indicates maximum activity. ‘++’ and ‘+’ indicate relative activities to the maximum. ‘-’ indicates no observed activity.

5.1.1.8 Yeast two hybrid western blots

Western blots were performed on all yeast transformants to ensure expression of all constructs. Example blots are shown below for all constructs analysed. Although only single blots are shown for each construct, western blots were performed for both LexA and VP16 in all combinations used for assays. Expression was observed for all constructs except VP16- β -arrestin-1 and LexA-catalytic domain. Western blots showed that in most cases expression of the construct was good. Only the VP16-catalytic domain construct was difficult to detect, requiring more lysate loading to observe a band by ECL detection.

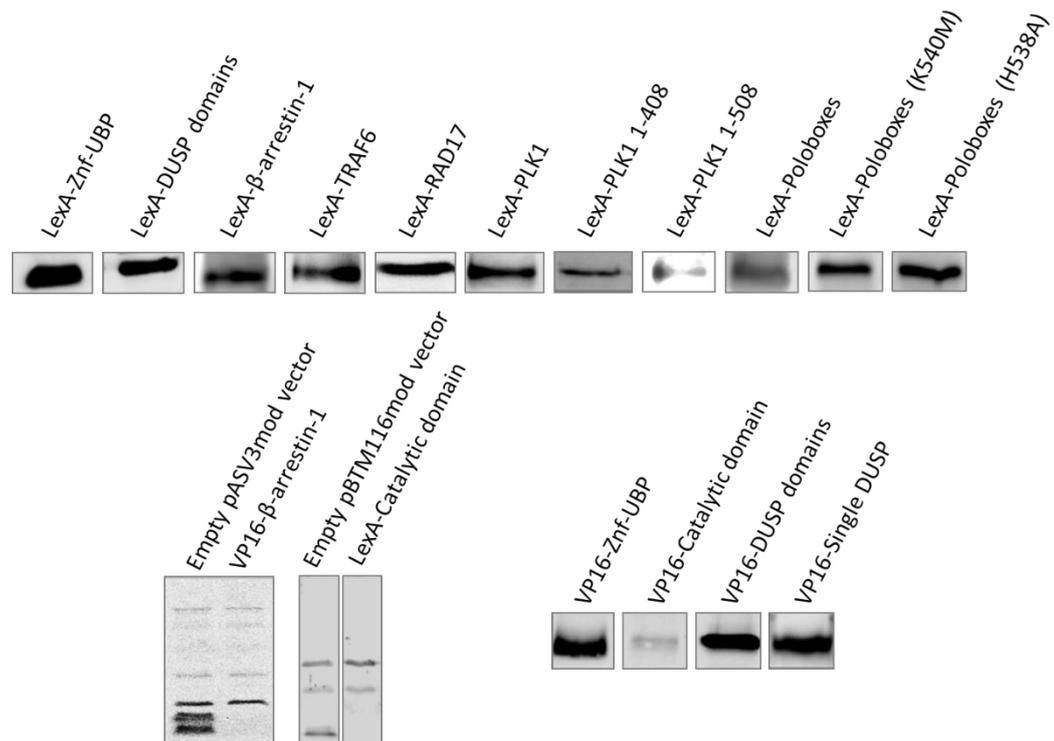


Figure 5.13. Yeast two-hybrid Western blots. (A) Western blots of the LexA fusion proteins in the Y2H assays. (B) Western blots of undetectable proteins. (C) Western blots of VP16 fusion proteins. Only the VP16- β -arrestin-1 and LexA-catalytic domain of USP20 fusions could not be detected by Western blot.

5.1.2 *In vitro* interaction assays

5.1.2.1 Expression and purification of poloboxes

From a five litre culture of bacteria containing MBP-Poloboxes, around 100 mg of total protein was obtained from the MBP-trap column using a step gradient of 0-10 mM maltose. The buffers were optimised due to aggregation and precipitation of protein through initial purifications at 150-300 mM NaCl and Tris as the buffer. Using buffers containing HEPES and salt concentrations of 500 mM, the protein was much more stable. The protein was first separated from aggregated protein using a Superdex 200 column. Here, there is a high peak at the void volume, indicating a large amount of protein aggregate; however, a larger proportion of protein appears

to be monomeric. This protein was cleaved with TEV and separated again with a nickel column (to remove TEV), a Superdex 75 column (to separate MBP and poloboxes) and finally a last MBP column to remove residual MBP. The final polobox product is very pure and reasonably stable. To ensure that there was no TEV contaminant in the sample (cleaved poloboxes are very similar in size to TEV), an uncleaved MBP-Polobox sample was incubated with TEV and the purified poloboxes (Figure 5.15). No cleavage was observed, indicating that the sample was pure poloboxes. The total yield of final poloboxes product is 2.6 mg/L of culture. An example poloboxes purification is shown in Figure 5.14.

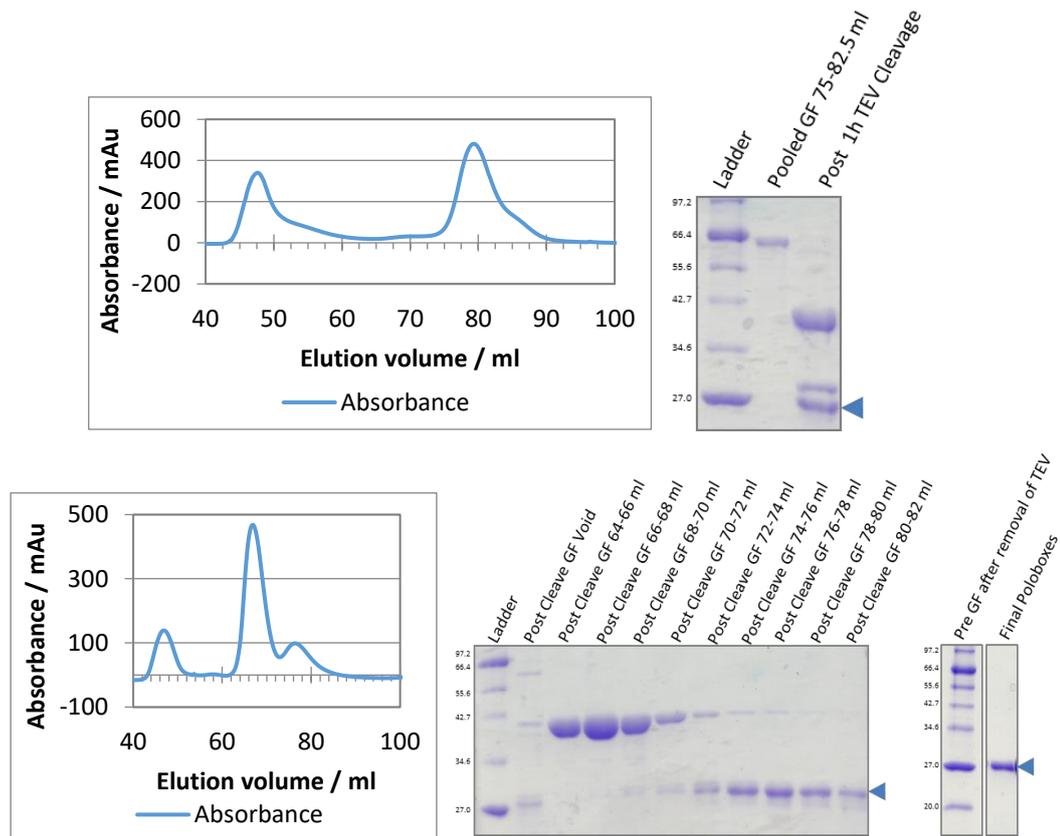


Figure 5.14. MBP-poloboxes gel filtration and cleavage. (A) The MBP-Poloboxes were eluted from the MBPTrap column and run directly on gel filtration Superdex 200. The trace shows a shouldered peak at a monomeric elution volume. The shoulder contained breakdown products that were avoided when pooling the fractions. The pooled protein was cleaved using TEV protease and was fully digested after 1 hour at room temperature. **(B)** The digested protein was passed through a nickel column and the flow-through was concentrated and run on a Superdex 75 column. The column doesn't base-line resolve the MBP and the poloboxes, so MBP protein is pooled with the poloboxes fractions. After passing the pooled fractions through an MBPTrap column, the resultant poloboxes are very pure.

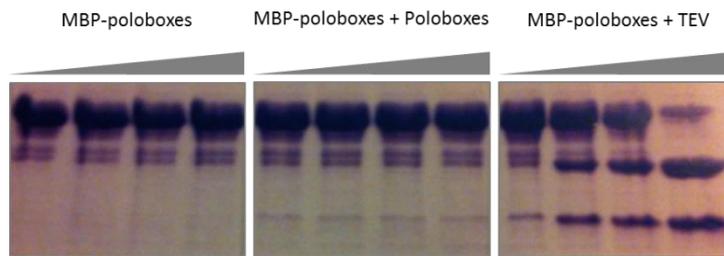


Figure 5.15. Poloboxes contamination test. To ensure that there was no TEV contamination of the poloboxes, 100 μg of the un-cleaved MBP-poloboxes was incubated alone (negative control; shown left), with 1 μg of the purified poloboxes (middle) or with 1 μg of TEV protease (positive control; right). The grey triangle indicates time. Cleavage was only observed in the TEV cleavage assay concluding that there was no TEV in the final poloboxes protein sample.

5.1.2.2 Purified DUSPs and poloboxes interact in a binary fashion.

I have shown that the USP20 DUSPs and PLK1 poloboxes interact using Y2H. Although conducting the hybrid experiment with recombinant proteins in the yeast strain *Saccharomyces cerevisiae* L40 should prevent interactions between the Y2H proteins and cellular proteins, it cannot be ruled out that these interactions could occur. In rare circumstances, this could lead to misinterpretation of the interaction as a direct binary interaction, whereas it is really being mediated by one or more additional proteins that may or may not be physiologically relevant. This is especially important with highly conserved proteins, such as PLK1; alignment of the poloboxes of human PLK1 and *S. cerevisiae* Cdc5 shows 24.4% sequence identity (not shown). Also, it has to be ruled out that the interaction is not in some way mediated by the presence of the extra residues in the Y2H fusions (the VP16 or LexA tag and linker regions). Therefore, it is essential to confirm that this interaction is binary using pull down assays or *in vitro* assays.

The MBP-poloboxes construct was used in pull down assays with the DUSP domains of USP20. Due to plasmid incompatibility, all pull downs were performed using either mixed pellets or reconstituting one expression lysate with purified protein. The MBP-poloboxes construct (uncleaved) was used to pull down the DUSPs using an MBP-trap. Cleaved poloboxes (by adding TEV to lysate) was also attempted, as the MBP could interfere sterically with the interaction. The DUSPs were also used to pull down both MBP-poloboxes and cleaved poloboxes using the Hi-Trap Chelating HP column. An example SDS-PAGE is shown where purified DUSPs were incubated with MBP-trap bound MBP-poloboxes. The sample was run on a nickel column, washed with 150 mM NaCl and eluted with an imidazole gradient. Fractions are shown where the DUSP domains are clearly visible, and a weak band at the correct height was observed for the MBP-poloboxes (Figure 5.16). In all cases, weak bands were seen at the correct heights, however it was not clear whether these were specific pull downs by the proteins. As a result of high concentrations of the proteins in the lysate the proteins may have bound non-specifically to the column due to the lower salt concentrations (150 mM NaCl).

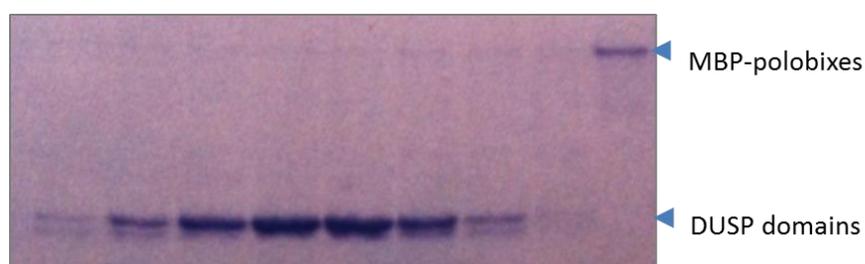


Figure 5.16. DUSPs-MBP-poloboxes pulldown. Pulldown assays using the DUSP domains show that a weak band at the height of the MBP-poloboxes are observed in fractions across the peak from the nickel column.

An ELISA using plate-bound poloboxes (un-tagged, cleaved PLK1 367-603) and His-tagged USP20 DUSPs (686-894) as a probe was performed (Figure 5.17). Multiple controls were used to ensure any signal was only due to specific interaction between the DUSPs and poloboxes. A His-tagged USP11DU probe was used to ensure that signal was not due to inadequate washing of the His-tagged protein, and that the poloboxes do not bind non-specifically to other proteins. A BSA probe was used to show that the purified polobox protein does not produce a signal (TEV protease is used in the purification and this could produce a signal if not completely removed). The DUSP was also used as a probe on BSA to show that the DUSP protein was not sticking to the plate non-specifically, and that it was being effectively washed away.

The data showed that the amount of DUSP domains that bound to the poloboxes increased linearly with increasing concentration of the probe, whereas increasing the concentration of all other probes does not increase signal. Linear regression of all assays by Prism GraphPad shows that only the regression of the DUSP and poloboxes assay shows a slope that is significantly non-zero (DUSP-poloboxes, $p = 0.0002$; DU15-poloboxes, $p = 0.3271$; BSA-poloboxes, $p = 0.8298$; DUSP-BSA, $p = 0.7971$). The OD_{450} observed for this ELISA is relatively low for the concentration of the probe. In addition, a linear increase in concentration of the DUSPs probe causes a linear increase in OD_{450} , therefore the poloboxes are not being saturated by the DUSPs at the concentrations used. The bacterially expressed and purified protein ensures that no mammalian proteins are mediating this interaction. Together, these findings

confirm the Y2H data: that the DUSP domains form a binary interaction with the poloboxes of PLK1, and that the interaction is probably of low affinity.

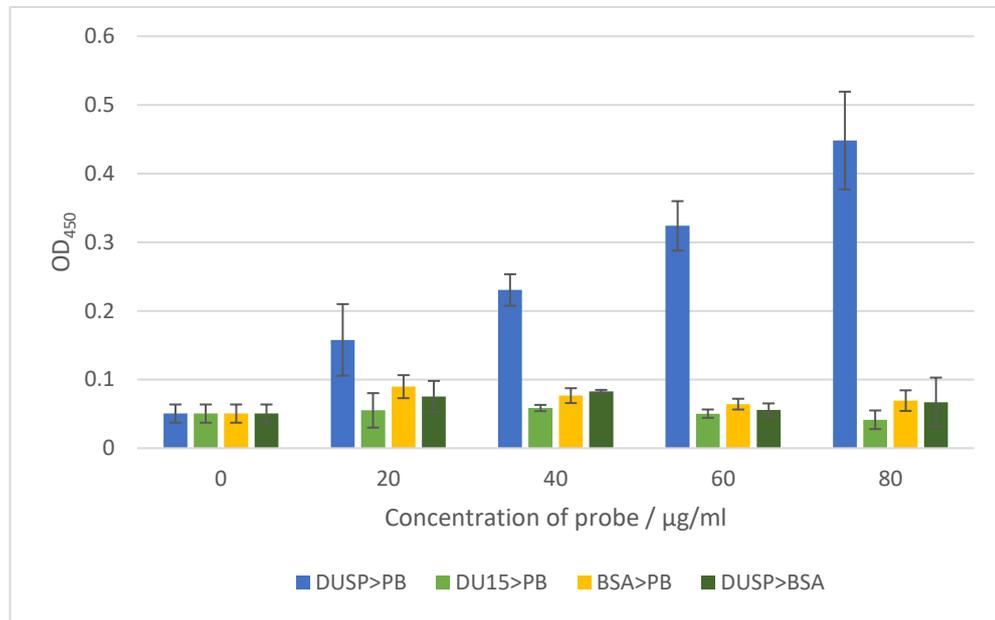


Figure 5.17. DUSP domains-poloboxes ELISA. Immobilised poloboxes were probed with increasing concentrations of DUSP domains. The signal from the ELISA increases with increasing DUSP domains probe concentrations. No significant increase was observed with the negative control, indicating a specific, binary interaction between the DUSP domains and the poloboxes of PLK1.

In an attempt to further confirm these findings, a far-Western blot was attempted (Figure 5.18). The principle is very similar to that of ELISA, but the poloboxes are bound to PVDF membrane rather than the ELISA plate. The poloboxes were spotted on the membrane from masses of 0.625 pmol to 40 pmol. The poloboxes were probed at 100 µg/ml (3.4 µM) of DUSP domains protein sample. Following detection by chemiluminescence, it may be possible that signals are observed at some of the spots when probing the poloboxes with the DUSPs. However, the data were inconclusive due to the poor signal obtained from the blot where the poloboxes were probed with the DUSP domains, and certainly does not allow a quantitative measurement of the assay. For this reason, the far-Western blot could not further support the data already obtained from the Y2H and ELISA.

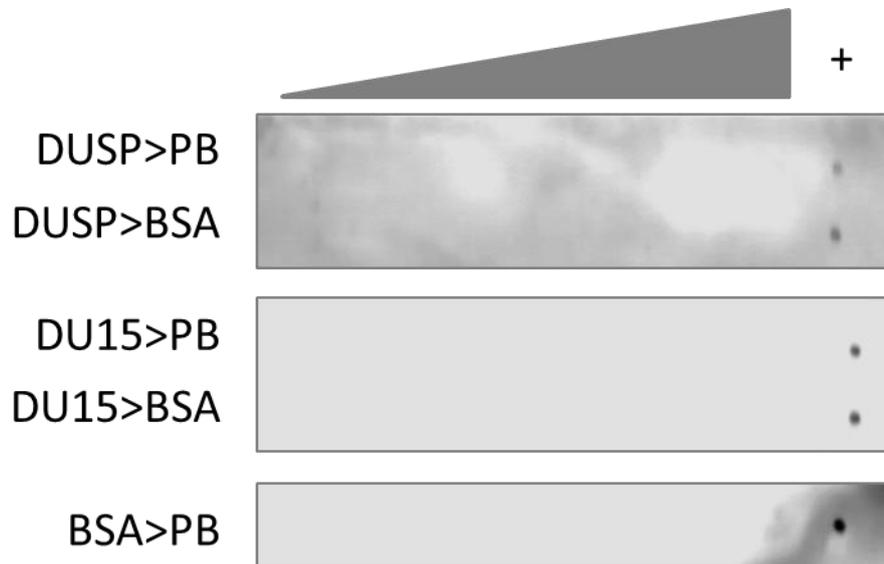


Figure 5.18. DUSP domains-poloboxes far-Western. Increasing masses of poloboxes and controls were spotted onto PVDF (indicated by the increasing grey marker at the top). The poloboxes and controls were probed with 100 µg/ml of DUSP domains. Signals from the control spots can be seen ('+' lane), but no signal is observed for the poloboxes probed with the DUSP domains.

5.1.2.3 USP20FL and PLK1FL show very weak interaction by ELISA

I previously showed that the DUSP domains definitely interact with full length PLK in Y2H and the poloboxes in Y2H and ELISA. However, it has not yet been shown that the full length USP20 binds to the full length PLK1. Full length TF-USP20 was used to probe full length PLK1 in ELISA assays (Figure 5.19). Again, multiple controls were used to exclude non-specific signals. USP20 was tested against PLK1 to check for specific binding and milk to ensure proper washing. Milk was used to probe PLK1 to assess background signal from the protein itself and USP20 to act as a positive control. The ELISA shows that there is a very small signal increase when probing PLK1 with USP20 compared with the washing and background control. Although this increase is small, it is statistically significantly different to the negative controls (USP20-PLK1 vs USP20-milk $p = 0.0001$; USP20-PLK1 vs milk-PLK1 $p = <0.0001$)

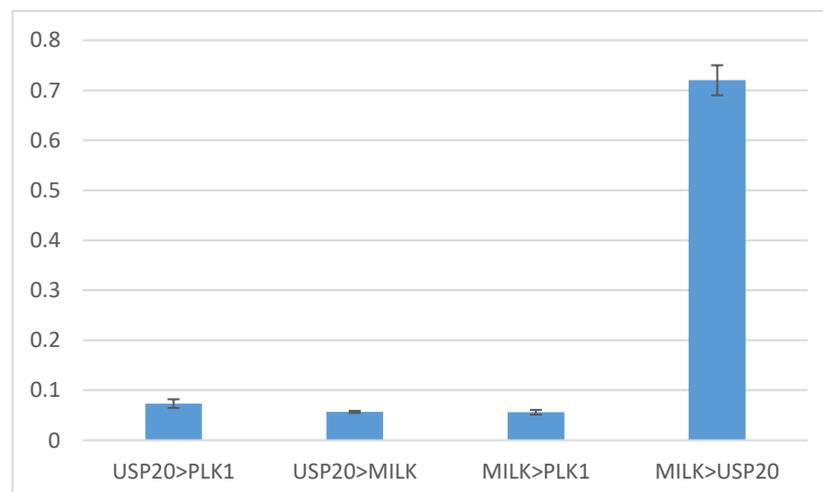


Figure 5.19. TF-USP20-PLK1 ELISA. Immobilised full length PLK1 and controls were probed with TF-USP20 or milk. The signal from probing PLK1 with TF-USP20 is very slightly, but significantly higher than the negative controls.

It is likely that the small increase shows a specific interaction between the two full length proteins. In addition to the predicted low affinity interaction between the DUSPs and PK1, another reason for the very low signal may be due to the protein purity of TF-USP20. The sample is far from being pure full length enzyme. This extra protein in the sample preparation would lower the amount of protein available to bind at any given concentration, and therefore would lower the apparent signal in ELISA assays. To fully confirm whether the two full length enzymes bind to each other, a better sample of full length enzyme should be used.

5.2 Pull down assays

Pull down assays were used to further dissect USP20's interactome (Figure 5.20). Because of protein mass requirements of 2 mg, only the Znf-UBP domain (USP20 1-101) and DUSP domains (686-894) were used. Data from MOPED [362], PaxDb [363] and MaxQB [364] show that USP20 and many of its currently known interactors are expressed in HEK-293 cells. Therefore, HEK-293 cells make a suitable choice when obtaining lysate for USP20 pull down assays. Whole, His-tagged Znf-UBP and DUSP domains were covalently bound to cyanogen bromide-activated beads and used to pull down proteins from the HEK-293 lysate. Ethanolamine-blocked bead controls with no bound protein showed very little protein being sequestered from the lysate. Also, an additional control of the DUSP-Ubl domain of USP11 was performed. This was done because minor contaminants from the purification procedure or proteins that non-specifically bind proteins or beads can be seen on the SDS-PAGE gel following the pull downs. These non-specific bands can be seen in all of the controls and pull downs. It was easiest to use the Znf-UBP and DUSP

domains pull downs to assess non-specific binding as these had the most obvious bands on SDS-PAGE, many of which were seen in both lanes. Although proteins could potentially bind both the Znf-UBP domain and DUSP domains, this was deemed to concern few interactions and therefore only bands individual to the pull down were considered as specific interactors. Therefore, only bands that appeared in either pull down, but not both, were excised and analysed with mass spectroscopy.

The Znf-UBP showed many individual bands compared to the controls and the DUSP domains pull down. Four sections of the lane were excised (numbered brackets next to lane) and included a total of eight distinct bands visible to the eye. Section one was medium-sized and contained two visible bands. Section two was the smallest taken, containing one band. Sections three and four were large sections, and contained three and two individual bands, respectively. The data obtained from the mass spectroscopy for all bands contains a great number of proteins that may have been bound to the Znf-UBP domain. In order to maximise the identification of true binders, rather than non-specific interactions with the protein or beads, information from the mass spectroscopy analysis, such as the number of peptides, score, protein cover, were used. In addition, CRAPome, an online server that contains mass spectroscopy data from negative mass spectroscopy experiments, and the blacklist proteins identified by Trinkle-Mulcahy *et al.* [365] was also employed. The CRAPome server shows how many times a specific protein has been observed in their 411 negative experiment datasets. Although this number can guide how likely it would be found even if it were not interacting with the Znf-UBP domain, it does not rule

out a positive interaction. Proteins found in more than 20% of the negative experiments (a score of more than 82) were removed from the list.

The blacklist includes groups, such as cytoskeletal proteins, motility proteins, DEAD box proteins, translation elongation and initiation factors, heatshock proteins, histones, hnRNP proteins and ribosomal proteins; and also an individual list of commonly found proteins from sepharose-bead pulldowns. Again, any of these proteins could be interacting, and many proteins were identified from all blacklist groups in the pull downs, but to reduce the chance of finding non-specific binding, these proteins were removed from the list.

Tables 5.2-5.5 show the top scoring genes encoding the proteins identified from sections 1-4 of the Znf-UBP pull downs. Table 5.6 shows top scoring hits from the DUSP domains pull down. The tables also show additional data obtained from the mass spectroscopy about the peptides and total cover of the proteins. The score is the sum of the ion scores for each distinct peptide match (not including peptides that match multiple proteins). The ion score is calculated from the probability that an individual peptide match is a random event.

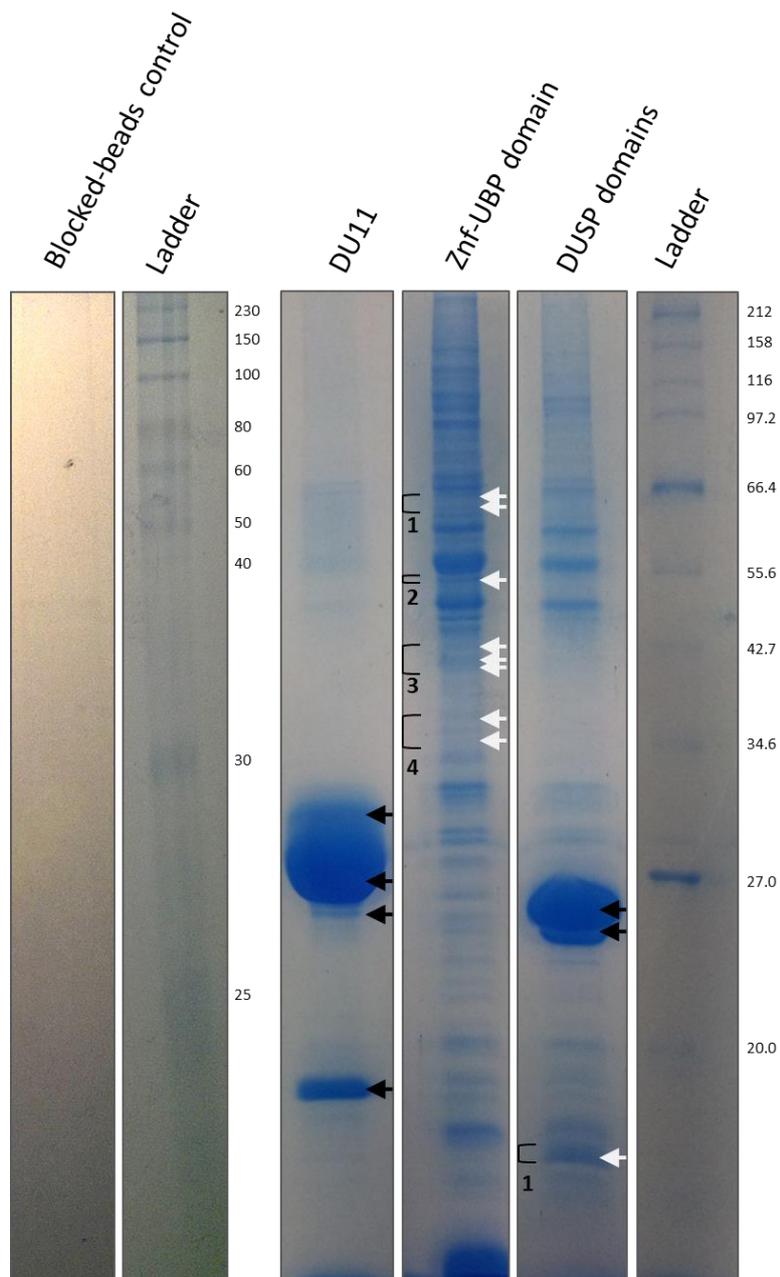


Figure 5.20. USP20 pulldown assays. The SDS-PAGE of pulldown samples are shown. The no-protein, blocked bead controls are shown in the very left lane. A small amount of protein can be seen bound to these beads, but the bands are very weak. The Znf-UBP domain and DUSP domains show many similar bands. These are assumed to be non-specific bead/protein interactors, but could be proteins that interact with both proteins. However, for increasing the chances of finding bona fide interactors, only bands that were not observed in either lane were chosen for excision and mass spectroscopy. Black arrows indicate the control or USP20 proteins that were observed in the lanes that were not exposed to HEK293 lysate (not shown). White arrows show lane specific bands. The numbered brackets indicate the excised regions that were sent for mass spectroscopy.

Table 5.2. Interactors from Znf-UBP domain band 1

Gene	Count	Score	% cover	CRAPome	Gene	Count	Score	% cover	CRAPome
PGM3	81	1717	58.5	5 / 411	PDPK1	35	878	50.2	0 / 411
IGF2BP2	78	1526	51.1	80 / 411	CDC7	35	748	36.1	1 / 411
VPS33B	76	1875	53.2	6 / 411	PIGT	35	636	35.4	2 / 411
CLPX	72	1657	50.4	23 / 411	ABCB7	31	920	31.2	7 / 411
GNL3	71	1468	52.7	80 / 411	LRRC40	30	772	32.6	9 / 411
STXBP2	66	1583	56.2	1 / 411	PRKAA2	30	684	28.4	5 / 411
EHD4	60	1564	63.8	3 / 411	NELFB	29	694	28.4	18 / 411
EDC3	58	1256	63.2	10 / 411	PICALM	28	876	28.4	18 / 411
ABCE1	57	1357	50.8	40 / 411	PTPN9	27	809	31.7	0 / 411
DNAJC11	55	1206	47.3	10 / 411	ANAPC7	27	702	32.7	9 / 411
hCG_31253	54	1405	58.2	N/A	hCG_19665	26	685	23.8	N/A
FARSB	54	1207	44.7	39 / 411	TMEM209	26	657	31.2	3 / 411
STXBP3	52	1160	45.3	5 / 411	PARP2	26	640	33.3	9 / 411
CTH	51	1215	52	2 / 411	MAP3K7IP1	25	793	44.8	45 / 411
MYEF2	47	1072	42.8	16 / 411	ATIC	24	606	29.1	36 / 411
MYEF2	47	1042	46.2	16 / 411	PAK4	24	597	23.4	29 / 411
ILVBL	45	1078	43.8	9 / 411	CDKAL1	23	678	29.9	12 / 411
RIC8A	42	918	40.9	1 / 411	DENND6A	23	527	25	0 / 411
THUMPD3	42	844	35.3	2 / 411	RBFOX2	22	642	31.1	14 / 411
SLC27A4	41	1239	41.1	2 / 411	CBS	22	634	35.8	25 / 411
PRKAA1	39	804	34.1	9 / 411	VPS33A	22	626	26.7	2 / 411
stxbp1	37	1026	43.8	2 / 411	CHEK2	22	557	27.5	0 / 411
PLK1	37	961	38	14 / 411	NUP58	22	555	30	3 / 411
VPS45	37	873	37.9	0 / 411	STT3A	22	473	17.3	23 / 411
PAPSS2	36	738	30.1	3 / 411	CDC45	21	639	25.6	5 / 411

Table 5.3. Interactors from Znf-UBP domain band 2

Gene	Count	Score	% cover	CRAPome	Gene	Count	Score	% cover	CRAPome
UMPS	113	2310	72.9	11 / 411	NDUFV1	36	815	37.6	6 / 411
ETF1	106	2238	74.4	56 / 411	BCCP	37	736	50.3	
hCG_2005638	98	1421	57		FLJ96593	24	729	31.8	
hCG_27698	55	1375	47.7		DBT	32	680	33.8	89 / 411
YARS	50	1277	50.4	71 / 411	ATP6V1H	24	674	36.6	18 / 411
POLD2	57	1270	53.6	8 / 411	LRRC14	29	665	34.7	0 / 411
VRK1	71	1268	65.2	5 / 411	GLUD1	28	662	32	90 / 411
RNMTL1	51	1209	53.3	15 / 411	NXN	27	598	35.4	5 / 411
CCZ1B	51	1153	48.1	1 / 411	STAMPB	29	563	42	0 / 411
AP2M1	50	1111	47.7	21 / 411	NARS2	22	563	27.5	5 / 411
RBM34	40	1094	47.4	32 / 411	NARS2	22	510	29.8	5 / 411
UBE1C	40	1088	54		CPVL	23	508	22.5	21 / 411
ARHGAP1	51	1085	51.9	4 / 411	DCP2	31	506	32.9	1 / 411
GSK3A	47	1077	52	6 / 411	RBM14	17	500	19	52 / 411
ILK	55	1025	52	1 / 411	ATG4B	16	486	20.2	0 / 411
PRIM1	49	999	52.9	2 / 411	SHC1	14	455	25.5	9 / 411
MAT2A	40	919	46.3	64 / 411	ALG2	17	444	26.9	0 / 411
PTPN1	30	913	45.1	21 / 411	RBM4B	14	443	18.7	18 / 411
ADSL	36	909	44.6	31 / 411	TRMT5	17	438	18.9	3 / 411
RCC1	32	901	56.1	61 / 411	CAMK2D	14	436	18.6	15 / 411
MINA	40	865	43	0 / 411	RRAGC	13	416	25.8	0 / 411
PSMD4	32	849	41.6	73 / 411	SLC30A9	12	407	15.5	1 / 411
STAU2	32	839	43	32 / 411	PARS2	12	401	22.1	4 / 411
MINA	38	832	42.9	0 / 411	RBM22	17	401	30.7	25/411
ULK3	29	827	38.6	0 / 411	POLR3D	14	398	26.1	0 / 411

Table 5.4. Interactors from Znf-UBP domain band 3

Gene	Count	Score	% cover	CRAPome	Gene	Count	Score	% cover	CRAPome
ACAT1	118	1798	78	63 / 411	CECR5	37	930	51.4	12 / 411
TRMT10C	77	1631	70	21 / 411	MRI1	45	926	50.1	5 / 411
PSMD6	72	1470	66.3	82 / 411	GNA13	48	911	48.8	22 / 411
STOML2	58	1371	52.2	32 / 411	RBM4	36	898	48.6	33 / 411
GTPBP10	57	1361	63.3	5 / 411	MPI	35	878	45.9	3 / 411
IDH3B	67	1331	61.5	17 / 411	ERAL1	35	874	44.4	10 / 411
ACAA1	54	1258	61.6	4 / 411	ECI2	36	861	56.9	16 / 411
DRG1	51	1200	57.2	53 / 411	TOMM40	42	859	47.4	36 / 411
GMDS	57	1183	70.7	3 / 411	TWF2	39	841	58.5	7 / 411
GALK1	49	1178	51.3	12 / 411	DNAJB6	33	837	41.1	25 / 411
ATP6V1C1	55	1157	62.3	0 / 411	TFB1M	36	836	52.6	6 / 411
ERLIN2	36	1148	46	45 / 411	UQCRC2	31	835	44.4	38 / 411
IDH2	50	1138	40.7	19 / 411	L2HGDH	41	828	39.5	6 / 411
ACOT7	47	1114	65.7	29 / 411	MAP2K4	41	824	47.6	2 / 411
DNAJB11	50	1104	45.5	52 / 411	sept2	34	821	54.5	88 / 411
SAE1	50	1094	61.3	43 / 411	RBM4B	30	788	41.5	18 / 411
ACTR2	52	1072	47	51 / 411	PCID2	34	788	47.4	19 / 411
TRUB1	44	1057	49.2	1 / 411	PRKACB	43	783	45	8 / 411
PRKACA	51	1031	53	11 / 411	HOMER2	29	783	45.5	1 / 411
ACADSB	38	1012	45.1	7 / 411	UBLCP1	41	774	40.9	0 / 411
PSMD13	41	1007	57.2	79 / 411	METTL15	33	755	32.2	1 / 411
DRG2	42	1002	54.9	6 / 411	GNAS	44	751	15.3	55 / 411
PHF6	54	997	69	42 / 411	PRPSAP1	34	749	55.1	59 / 411
CNP	37	960	51.9	26 / 411	DCAF7	29	741	48.8	33 / 411
RFC2	34	938	52.8	56 / 411	C4orf27	35	738	42.5	2 / 411

Table 5.5. Interactors from Znf-UBP domain band 4

Gene	Count	Score	% cover	CRAPome	Gene	Count	Score	% cover	CRAPome
CHORDC1	88	1313	80.1	21 / 411	MTHFD2	24	703	53.8	29 / 411
SRM	67	1269	77.8	17 / 411	HSD17B4	22	701	20.5	68 / 411
POLDIP2	60	1132	56.5	36 / 411	RPRD1B	27	689	37.7	30 / 411
E9KL35	45	1119	78.5		SCAMP3	26	663	41.2	23 / 411
OSGEP	46	1010	58.8	5 / 411	RPUSD3	28	662	44.6	1 / 411
PDHB	68	1000	60.7	52 / 411	ATAD1	33	659	44	6 / 411
PYCR2	43	950	47.2	32 / 411	SUCLG1	28	654	36.9	20 / 411
ALG5	38	936	47.2	2 / 411	WDR82	30	636	34.5	42 / 411
RPP38	39	931	59.7	7 / 411	ACOT8	25	634	49.5	3 / 411
C9orf64	40	910	50.7	3 / 411	TAMM41	27	631	44.8	6 / 411
GPD1L	36	869	40.7	1 / 411	CNN2	21	625	48	38 / 411
CSNK1A1	40	869	54.1	20 / 411	FLJ21103	27	613	31.1	
HAX1	44	838	73.8	29 / 411	PRKRA	23	605	38.7	23 / 411
DIMT1	37	824	55	20 / 411	DYNC2LI1	26	603	35.2	0 / 411
OTUB1	34	824	56.6	16 / 411	VDAC2	19	598	42.6	58 / 411
AIMP1	33	819	69.9	71 / 411	TRUB1	18	595	33	1 / 411
CSNK1A1	36	808	56.3	20 / 411	EMD	22	591	49.6	82 / 411
AASDHPP T	49	806	53.4	7 / 411	ATP6AP2	23	579	44.7	2 / 411
TYMS	36	785	46.3	5 / 411	NDUFA9	27	574	43.2	17 / 411
PRKAG1	31	781	53.2	12 / 411	P15RS	23	571	37.5	
NUBP1	27	765	59.5	9 / 411	FYTTD1	22	557	37.7	13 / 411
NT5C3A	31	761	45.3	0 / 411	PPP6C	25	548	40.3	14 / 411
SFXN1	30	759	49.7	27 / 411	SEH1L	20	539	32.1	38 / 411
CDK6	29	748	44.8	24 / 411	PGP	21	524	36.1	6 / 411
TSEN34	27	720	50.3	0 / 411	MLEC	17	513	67.8	4 / 411

Table 4.6. Interactors from DUSP domains band 1

Gene	Count	Score	% cover	CRAPome count
SSBP1	24	519	60.8	78 / 411
UBE2N	8	181	39.5	46 / 411
CSTA	7	173	74.6	45 / 411
Similar to TRAPPC1	7	139	24.8	0 / 411
LGALS7B	5	138	28.7	39 / 411
NDUFA13	6	135	13.5	6 / 411
ATP5D	4	135	13.7	21 / 411
MGST1	5	119	16.8	0 / 411
PIP	4	111	14.4	46 / 411
CASP14	5	104	13.6	53 / 411
UBE2D1	8	101	19.7	3 / 411
FABP5	5	95	25.7	38 / 411
ISCU	3	86	16.7	0 / 411
CDKN2A	4	74	31.8	25 / 411
LYZ	4	70	10.8	58 / 411
ISCA1	3	70	15.3	0 / 411

5.2.1 Gene ontology

In order to assess what gene ontologies are enriched in the data obtained from the mass spectroscopy, the DAVID server [366, 367] was used to annotate the proteins and form functional clusters. For this, the data from all bands were collated and entered into the DAVID server to investigate whether any functional clusters were enriched in the mass spec data above the *Homo sapiens* background. Figure 5.21 and 5.22 show the enriched clusters and the number of proteins identified that are associated with the particular ontology. Only clusters that achieved P-values of less than 0.05 are shown. Multiple groups of proteins are identified, which include processes already known for USP20 and also novel roles for the enzyme. Proteins involved with the cell cycle, DNA repair and G-protein-coupled receptors were observed in the mass spec data. These data support USP20's roles with these processes. Also, many proteasomal sub-units were identified, which is interesting as no interaction has as yet been identified between the proteasome and USP20. Novel

processes/locations include proteins involved in cell adhesion, vesicle docking, nuclear transport and nucleotide metabolism.

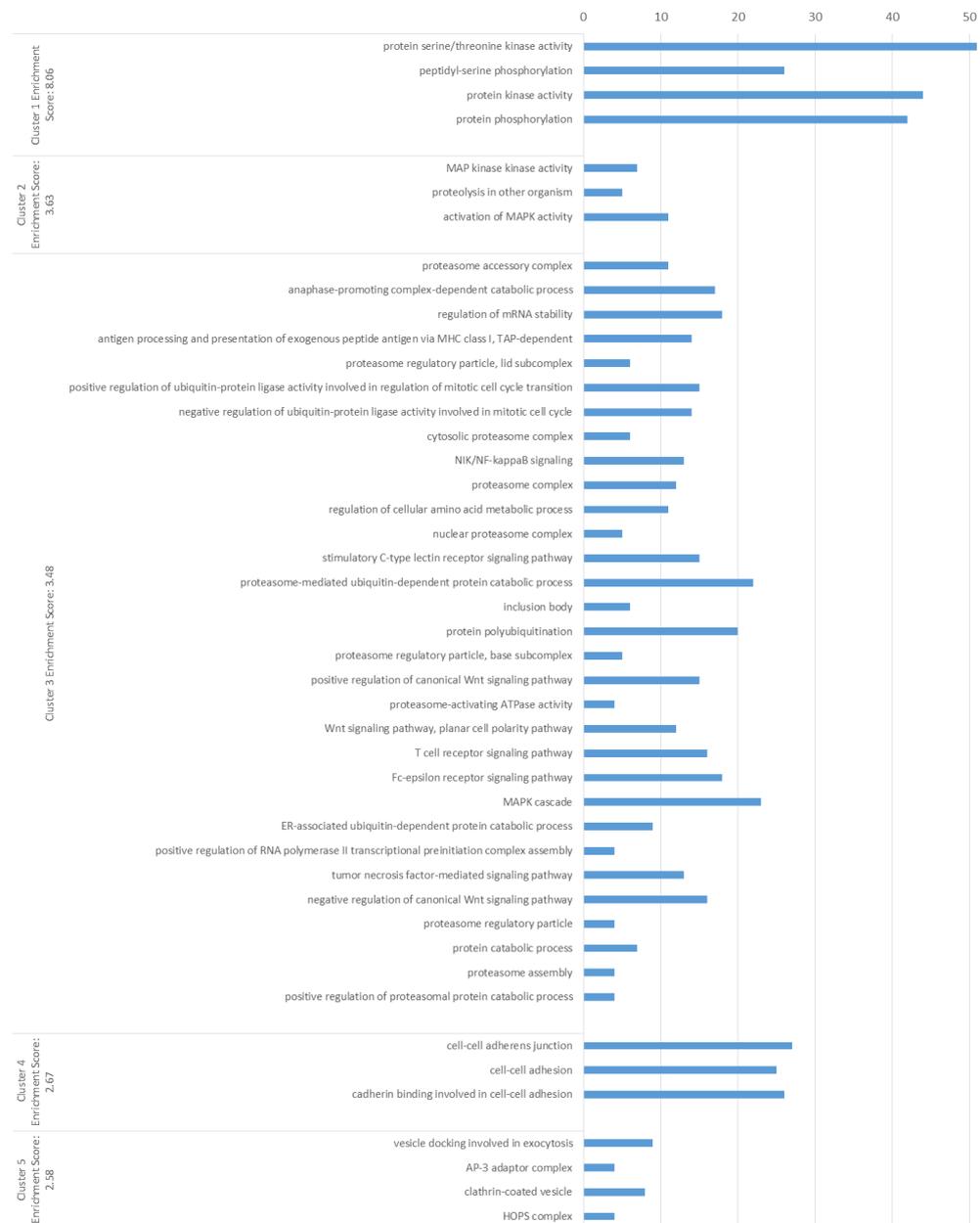


Figure 5.21. Gene ontology clusters 1. Clusters 1-5 are shown with their enrichment score.

The number of proteins in each gene ontology annotation is given (blue bars).

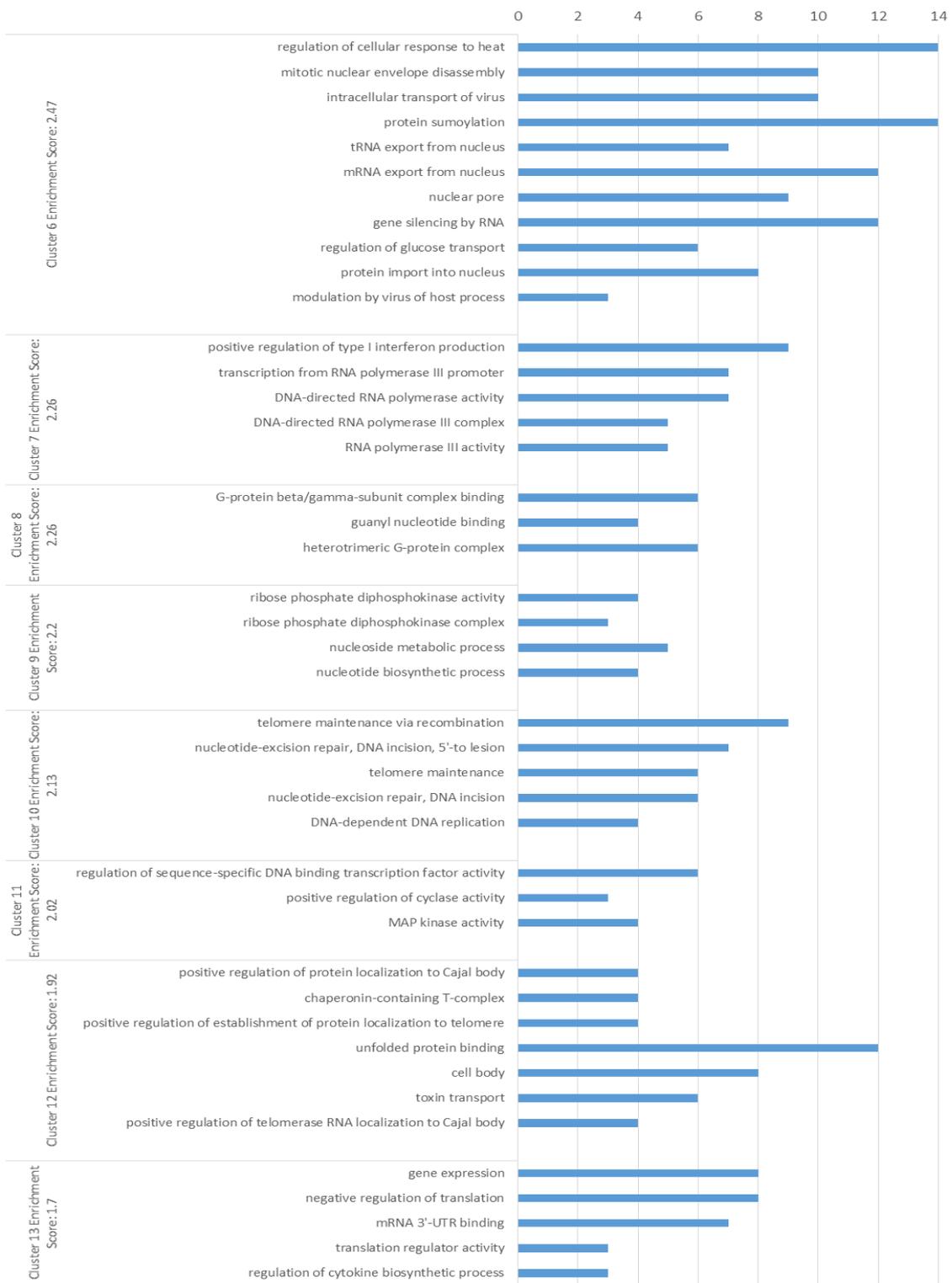


Figure 5.22. Gene ontology clusters 2. Clusters are shown with their enrichment score. The number of proteins in each gene ontology annotation is given (blue bars).

5.3 USP20 Znf-UBP and ubiquitin

The Znf-UBP domain has been implicated in binding RAD17 [368], which is its only mapped interaction to date. As the Znf-UBP domains of HDAC6 and USP5 bind ubiquitin, but the Znf-UBP domains of USP33 do not, it was assessed whether USP20's Znf-UBP domain (residues 1-101) could bind ubiquitin. ELISA, ITC and Thermofluor were used to assess binding. Ubiquitin is routinely purified in the author's lab and so the purification is well optimised and will not be discussed in great detail. Large quantities of ubiquitin is readily purified by reducing the pH of *E. coli* lysate, precipitating most of the proteins. Ubiquitin is very stable and can refold when denatured so this harsh method can be employed to increase the purity of this native purification. The remaining lysate is loaded onto a cation exchange column and gel filtration is used as a polishing step. Very pure protein is obtained, as shown in Figure 5.23.

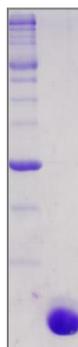


Figure 5.23. Purified ubiquitin. The SDS-PAGE of the final sample of purified ubiquitin is shown. Following pH precipitation of *E.coli* proteins, cation exchange and gel filtration, the ubiquitin sample is very pure.

Thermofluor can be used readily for small-molecule interaction studies, where upon binding a change in protein melting temperature can occur. Often it is not as useful in the analysis of protein-protein interactions because there will be two binding curves in the thermogram, which prevent the individual assessment of each protein. However, ubiquitin is very stable, with a melting temperature over 90 °C (the Thermofluor curve doesn't start to increase until after 80 °C), so it can be used to assess binding of proteins with melting temperatures in a more typical range. The mean melting temperature of the control Znf-UBP was 71.89 °C, which is very close to the control used in the buffer assays (72.24 °C). When ubiquitin was added to the Znf-UBP domain at 10 x molar ratio, there was no significant change in mean melting temperature, 71.51 °C, $p = 0.3706$ (Figure 5.24A). This suggests that no binding is occurring between ubiquitin and the Znf-UBP domain of USP20. However, in some cases, binding can occur without affecting the melting temperature, so ELISA (Figure 5.24B) and ITC (Figure 5.25) were additionally used to investigate binding.

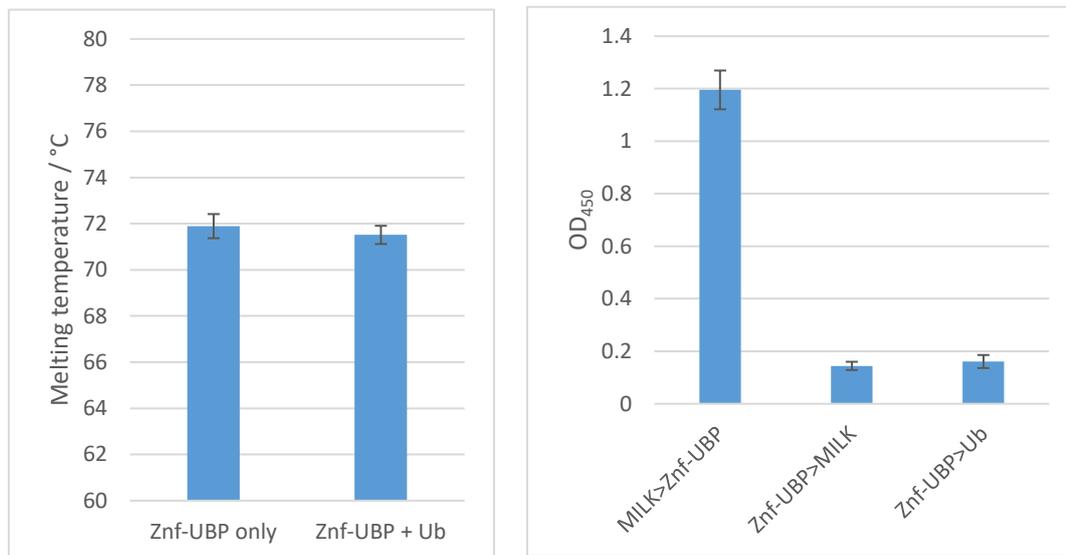


Figure 5.24. Znf-UBP and ubiquitin Thermofluor and ELISA. (A) The Thermofluor assay shows that there is no significant difference between the melting temperature of the Znf-UBP domain with or without ubiquitin present. **(B)** There is no significant interaction between the Znf-UBP domain and ubiquitin on ELISA; the signal when probing untagged ubiquitin with His-tagged Znf-UBP domain is no different than when probing milk.

Untagged ubiquitin was bound to a microtiter plate, and the His-tagged Znf-UBP 1-101 domain was used to probe it. The positive control (plate-bound His-tagged Znf-UBP) showed a strong OD. However, there was no difference in signal when using the Znf-UBP domain to probe ubiquitin or milk ($p = 0.3805$). Again, this assay shows no interaction between ubiquitin and the Znf-UBP domain of USP20.

The Znf-UBP domains of HDAC6 and USP5 were analysed by ITC to measure their equilibrium constant (K_d) with ubiquitin. HDAC6 showed a 60 nM K_d [369] and USP5 showed a 2.8 μ M K_d [86]. If there was binding of ubiquitin and the Znf-UBP domain of USP20, it should be detectable with ITC. 8 μ l and 16 μ l of 400 μ M ubiquitin was

injected into 40 μ M Znf-UBP domain in the cell. A very low signal was observed and there was no detectable binding. If the binding constant of USP20's Znf-UBP domain was even close to that of USP5, some binding should have been observed. The ITC, Thermofluor and ELISA data concludes that the USP20 domain does not bind to ubiquitin.

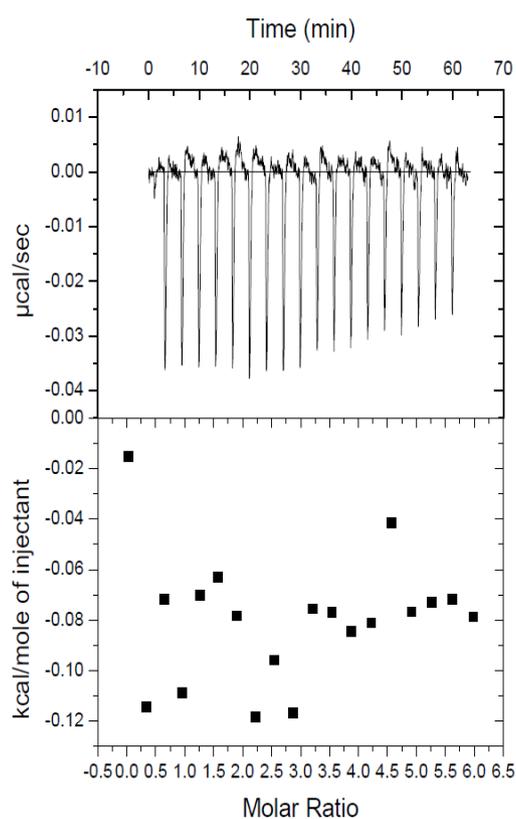


Figure 5.25. Znf-UBP and ubiquitin ITC. The ITC shows a very weak exothermic signal. The lower graph shows that there is no sigmoidicity to the points on this graph. Therefore, even with the large injection volumes of ubiquitin, no binding is observed between the two proteins.

6 Discussion

6.1 Structural studies of USP20

Crystals of USP20 protein have not yet been produced in the crystallisation trials. However, in some cases, crystals may take in excess of 6 months to form, so it is still possible that the trials performed may yield crystals. The chosen method for protein expression in this thesis has been *E. coli*. This is because, due to its ability to grow to high densities in inexpensive media, it boasts a cheap and relatively quick method to achieve pure protein in quantities high enough for crystallisation. Also, many cloning options and expression strains are available, which can increase the proportions of soluble protein produced. However, *E. coli* has many drawbacks as a system for recombinant human protein expression. Firstly, it doesn't have the chaperones of the human proteome. This often presents problems with the production of soluble protein and correct folding of the soluble proteins. Also, codon bias is observed between heterologous organisms and human genes often present with multiple rare codons in the human sequence, which may need mutagenesis or whole-gene-synthesis to rectify. From the evidence obtained from the expression with any catalytic domain-containing construct of USP20, it appears that *E. coli* is not the ideal method for its production. As the catalytic domain of USP20 was also poorly expressed when fused to the LexA protein in the yeast two-hybrid, it is likely that yeast may not be ideal for the protein's expression either. It is likely that the use of mammalian cells or insect cells would provide better yields, as this is often the case when using these expression systems. However, this would greatly have increased the cost of obtaining the protein.

Initial expression of solely histidine tagged proteins is still the ideal method for downstream purification as it provides an adequate balance between ease of purification, solubility and negative consequence on crystallisation. However, proteins are often screened for the tags that provide the most soluble protein from the beginning and only the best fusions are utilised further. This increases the cost initially as all of the fusions have to be cloned. This thesis started with his-tagged proteins and only when these were adequate would another tag be utilised. This reduced the cost, but also increased the time involved dramatically as each construct was cloned sequentially rather than concomitantly. In the case of USP20, using a trigger factor tag did not effectively yield monodisperse protein, and the use of the tags, although increasing the solubility of the DUSP domains, did also not yield crystals.

The design of the novel tags was an interesting process and the two tags that were produced have promise in the field of structural biology. The lack of solubility tags that can significantly increase the chance of crystallisation as a fusion is seriously lacking. Other than thioredoxin, very few small tags for carrier-driven crystallisation have been employed. Thioredoxin is very useful because it acts both as a chaperone and a reducing enzyme to prevent disulphide formation. It has been used successfully to crystallise fusion proteins [370, 371]. However, the efficacy of each protein tag is completely dependent on the protein of interest that is being fused. One tag may work for one protein whereas another may not. For this reason, there is a need for a choice of proteins that can carrier-drive crystallisation. For the design of the two novel tags in this thesis, the desired characteristics were determined and

the PDB was screened. Essentially, the premise was to identify small, rigid, globular proteins that diffracted to a high resolution during X-ray diffraction. Two of the design choices that may not have been ideal were the choices for no enzymatic activity or ligands. Although this ensured rigidity, which was the original reason for this stipulation, it actually has limited the potential of the tags with regards to their purification, and possibly their fusion-crystallisation potential. Having small-molecule ligands in the crystal structure could allow the development of novel purification methods, similar to that achieved with GST and MBP fusions (Glutathione and dextran columns). However, although enzymatic activity and ligand binding was avoided because of conformational changes and flexibility that is possible, the different conformations could provide different crystal contacts and either conformation could allow or prevent crystallisation. This was observed in at least in one case with an MBP fusion, where only the apo form (no maltose bound) of MBP yielded crystals of the MBP-sAgIB fusion [372]. Therefore, if a rigid protein (low B-factors) could be identified that underwent a strict change between multiple conformations, this could increase the possibility of forming crystals as a fusion partner.

As the use of solubility tags did not yield crystals, other techniques to increase the prospect of crystallisation include buffer optimisation, altering the boundaries of the domain, using different expression systems, attempting to co-crystallise the protein of interest. Buffer optimisation was performed, and, although the solubility was increased using ethylene glycol as an additive, did not yield crystals DUSP domains. The DUSP domains were expressed using the predicted boundaries and

additionally the remaining C-terminus of USP20 (an additional 20 amino acids). Both constructs behaved similarly, and there was no other obvious point to extend or retract the boundaries of the domain. The proteins were expressing well in the *E. coli* expression systems and so it was deemed necessary to express them in either mammalian or insect cells. And a protein binding partner was sought so that the DUSP domains could be co crystallised, but due to the low affinity of the interaction, it was not possible to produce a complex of the two proteins for crystallisation trials. Ultimately it is difficult to compare how the fusion of the novel tags compared to any of these other methods for obtaining crystals. Not only because crystals were not obtained in any case, but also because the observed characteristics of solubility and protein yield may not truly represent a protein's ability to crystallise. For example, altering a protein's boundaries may affect these parameters, but actually the most readily crystallisable protein could be the more unstable.

The expression and purification of the Znf-UBP domain produced monodisperse protein, but the protein would not crystallise. Using XtalPred [373] to predict the likelihood of crystallisation, both the USP20 1-101 and 1-108 show low scores for both constructs. This is largely due to the amount of coil in the protein, as well as having high predicted instability indices. The focus of this thesis was to analyse the structure of USP20 domains by X-ray crystallography. However, it is quite possible that using NMR to analyse the structure would have yielded results because it had previously been used for the USP33 Znf-UBP. Attempts to optimise the buffer and to use alternative domain boundaries were employed to produce crystals. Although it is likely that the buffer optimisations stabilised the protein (USP20 1-101), the

increase in thermostability did not allow the formation of crystals. The buffer optimisation was not performed on the USP20 1-108 construct and it is possible that the buffer optimisation could have improved this protein's stability and yielded crystals. Additionally, circular dichroism should be used to assess whether the protein is correctly folded. Confirmation of folding was taken as a monodisperse peak on the gel filtration, which actually may not be the case. Circular dichroism would show what proportions of alpha and beta secondary structure are present, which would could be compared to structure prediction and the USP33 Znf-UBP structure.

The catalytic domain leads to aggregation of most of the protein constructs, which prevents crystallisation. Also, all purifications of catalytic domain-containing constructs are very impure, likely through breakdown. In hindsight, it is quite possible that the C-terminus was too short on the expression constructs that ended at the catalytic domain (Catalytic domain and USP20 Δ DUSPs). This may have led to the aggregation seen in these constructs, but did not account for the lack of monodispersity of the other catalytic domain-containing constructs. In hindsight, it would have been better to extend the terminus/termini and then use limited proteolysis to produce a construct with minimal-length terminal extensions. However, this would also remove the 182-residue and 52-residue inserts, which on one hand may help crystallisation and protein stability, but on the other would remove interesting parts of the protein structure.

The cause for the DUSP domains' inability to crystallise is likely due to aggregation because of an unstable N-terminus. Although circular dichroism was not performed

on this protein either, which would have confirmed that the protein was properly folded. In addition, the 686-894 construct has a poor instability and GRAVY index. Although buffer modification enhanced the overall solubility of the protein, it did not aid in its crystallisation. The novel solubility tags had quite divergent effects on the solubility of the DUSP domains. 2GKG had produced an impressive improvement on solubility, almost to the degree of MBP. The 1BKR tag produced only a modest improvement on the solubility of the domains. The tags alone showed that 1BKR is the most crystallisable of the tags, and thus may be the most beneficial in crystallising proteins in general. These tags are being tested on other proteins to assess their ability to act as solubility tags, and fusion tags. As the primary aim of their use was to aid in carrier-driven crystallisation of the DUSP domains, they have not yet achieved their purpose. However, with the use of further fusions, they may yet prove effective.

6.2 Interaction of USP20 and PLK1

The putative interaction between USP20 and PLK1, seen in the high-throughput proteomic analysis by Sowa *et al.*[172], was validated by both cellular and *in vitro* assays in this thesis. The yeast two-hybrid was initially used to identify this interaction. This technique is notorious for false positives and thus, it was essential to prove *in vitro* that the interaction was true. This provided a difficult problem as the protein interaction was obviously too weak to be identified by pull down assays with any certainty and also would not produce a complex in solution that could be observed on gel filtration. Only ELISA was able to confirm the interaction, and this showed a weak signal, further corroborating the weak interaction. Ultimately, the

interaction was mapped to the poloboxes and DUSP domains of USP20, and was proven to be a binary interaction. The poloboxes have been shown to bind to specific phosphopeptides containing the sequence S-{pS/pT}-[P/X] [213]. Interestingly, the DUSP domains of USP20 do not contain this motif and the interaction occurs even though the DUSP domains are not phosphorylated (as proven by the binding of *E. coli* purified proteins). However, the histidine residue of the phosphosensing pincer (H538 and K540) was shown to be important to the interaction. This suggests that proteins can bind to the PLK1 poloboxes in a non-canonical manner.

Studies show that the serine residue in the P-1 position is essential for maximal binding of the poloboxes to the peptides [213]. However, this motif was identified for peptides produced from a pThr-Pro oriented library. It may be that other sequences can be detected by this cleft if the phosphorylated residue is followed by a residue other than proline, due to restraints on the conformation of the binding peptide. If other sequences could bind, it could be that, phosphorylation of the DUSP domains occurs in mammalian cells, which would allow the DUSP domains to interact with the buried lysine of the pincer. This would not have been identified in the assays performed in this thesis because the assays were either in yeast, or used purified protein from *E. coli*.

If the binding site does only accommodate peptides of the identified motif, the DUSPs may bind to the region surrounding the cleft and only interact with the histidine as it is well exposed on the surface of the poloboxes. Without more

investigation to identify the residues of the DUSP domains involved in the interaction, it is not clear in what mode it binds.

Due to the weak nature of the interaction identified with the DUSPs, the following hypothesis could be made: the DUSP domains of USP20 mediate only a transient interaction with PLK1. Although the catalytic domain showed no binding to PLK1, there are multiple sites of the large disordered insert that fulfil the residues for the motif. In addition, multiple serine residues in these motifs have been identified to be phosphorylated in large scale proteomics screens.

These data fit a model where USP20 is kept in close locality to PLK1 with the weak interaction between the DUSPs and the poloboxes. Upon phosphorylation, these motifs could massively increase the affinity of PLK1 with USP20. In addition, Shanmugam *et al.* [368] showed that decreasing cellular USP20 levels also decreases levels of PLK1, although they did not show that this was a direct cause of USP20's deubiquitination of PLK1. If the interaction between USP20 and PLK1 is promoted by phosphorylation, then it is likely that this leads to PLK1's stabilisation through USP20-mediated deubiquitination. This would promote the transition through the G2/M checkpoint. One of these S-[pS/pT]-[P/X] motifs, S304-S305-Q306, is phosphorylated by ATR in the response to DNA damage [374], but it's unclear as to why stabilisation of PLK1 would occur following DNA damage. However, the peptide spot assays performed by Elia *et al.* [213] show that a glutamine residue in the P+1 position reduces the binding of the poloboxes to the peptide, so this may not be a motif that interacts with PLK1. More likely, the motif S407-S-408-P409 is a possible putative PLK1-interacting sequence. Residues S407

and S408 have been identified as phosphorylated. This sequence (SSPPR) forms both a PLK1 polobox motif and cyclin dependent kinase (CDK) substrate motif: [pS/pT]-P-X-[K/R]. It seems possible that this region is under control of CDKs. Therefore, the full hypothesis is that the DUSPs ensure localisation of PLK1 and USP20, and, when the cell cycle checkpoint is satisfied, a CDK phosphorylates the motif, promoting the interaction between USP20 and PLK1. This stabilisation then leads to the transition of G2/M (shown in Figure 6.1).

Interestingly, many of the phosphorylated motifs were identified in cancer proteomes. Aberrant stabilisation of PLK1 by constitutive phosphorylation of USP20 would promote transition through the G2/M checkpoint, promoting the proliferation of cancer cells.

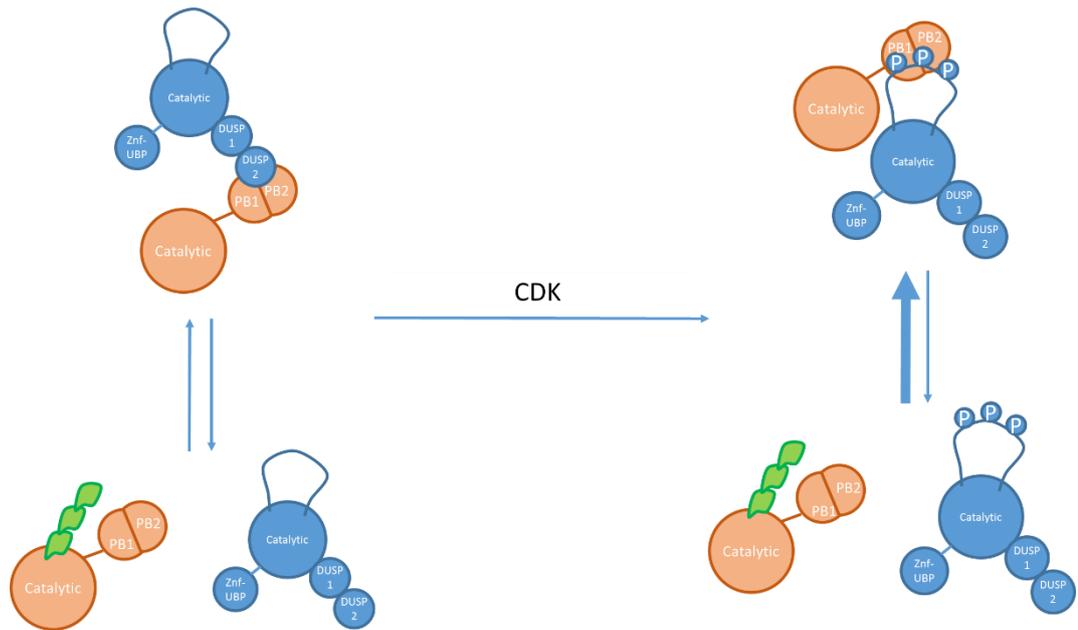


Figure 6.1. Proposed model for USP20-PLK1 interaction. USP20 (blue) interacts with PLK1 through the DUSP domains and poloboxes (left). The low affinity means that the stabilization of PLK1 is kept low. The transient interaction could ensure that USP20 is kept in close locality to PLK1 in the cell. Upon phosphorylation (possibly by CDKs) the affinity for USP20 and PLK1 increases, through interaction between the poloboxes and the catalytic domain insert. This stabilizes PLK1, promoting the transition through the cell cycle.

PLK2 and PLK3 all contain homologous residues to H358 of PLK1 (H629 in PLK2 H590 in PLK3). USP20 could also bind to these poloboxes, depending on which other residues are involved in the interaction. Also, mutation of USP20 D894 to an arginine reduced the interaction slightly. USP20 and USP33 are highly conserved throughout this region (and most of the poloboxes) and thus may also bind to the poloboxes. Further investigation would be required to know whether these homologous proteins could interact with each other.

6.3 Binding partners and processes

The pull down assays revealed a large number of putative interacting proteins. These proteins have roles in multiple cellular processes; some of which are already known to USP20, some of which may highlight novel roles for the enzyme. The protocol used for the pull down assays has been validated previously in our lab. However, the mass spectroscopy performed on each excised band of the SDS-PAGE identified far more proteins than were expected. A possible reason for this is that the presence of any beads on the gel produced smeared lanes (data not shown) and the procedure for eluting the proteins and loading them onto the gel was optimised in such a way that smears were not present. It is visible that there is a constant stain all the way through each lane of the gel and thus it may be that all beads were not prevented from loading onto the SDS-PAGE. Therefore, it is possible that the highest scoring peptides were those from the visible band that was excised, but the proteins present due to the smear were also detected. These proteins were likely present in the eluted sample and are possibly proteins that were pulled down specifically by the USP20 proteins. To make the validation easier, it would have been ideal to excise adjacent bands from the control lane. By performing the same analysis on these sections, it would have identified the background signal from the pull downs. This was not done so the use of immunoprecipitation black-lists and the CRAPome server had to be utilised instead.

Other techniques such as using a library in the yeast two-hybrid screen or using SILAC would have also produced a large amount of data for interacting proteins. The yeast two-hybrid is useful as it primarily shows binary interactions, and SILAC

may have been a more useful protocol as it allows the identification of enriched peptides over the background. However, as the protocol was validated and we had no access to a yeast two-hybrid library, this was the protocol that was chosen at the time.

A discussion of some of the proteins identified are given below. As the list is so diverse, gene ontology was used to investigate functional clusters within the proteins identified by mass spectroscopy. Of the most enriched clusters, protein kinases were the highest. 51 protein kinases were observed in the pull downs. These include those involved in cell cycle regulation such as CDKs, CHEK2 and PLK1. Interestingly, PLK1 was pulled down with the Znf-UBP domain rather than the DUSP domains, indicating that there may be a complex formation with some indirect interaction of the Znf-UBP domain and PLK1. The interaction between USP20 and these kinases may further elucidate how USP20 is involved in the DNA damage response and cell cycle. In addition, cluster 10 shows proteins involved in DNA repair (nucleotide excision), telomere maintenance and DNA replication; including BRCA2, POLD2, POLD3, RAD50, RAD51C, XRCC3, FEN1, PCNA, RPA1, MCMBP and PARP1. As yet, these proteins have not been implicated with USP20 in its role in DNA repair, but further investigation could explain how USP20 regulates or is regulated by these proteins.

In addition to known processes, the mass spec highlighted proteins involved in nuclear transport/the nuclear pore, genes involved in expression and proteins involved with location to the Cajal bodies. This is interesting as neither the nuclear

localisation of USP20 or USP20's role in nuclear transport have been observed. It also makes an interesting link with the observation that the DUSP domains act as a transcription activator if USP20 is observed within the nucleus. However, this is purely hypothetical and a great deal of investigation would be required to assess this.

With regards to USP20 and its role in innate immunity, no obvious functional groups were observed that could enhance the current understanding of how it would function to regulate any part of the process.

6.4 Ubiquitin binding

This thesis has shown that, like USP33, USP20 does not bind ubiquitin. Allen and Bycroft [84] concluded that the interaction is not observed with USP33 because of replacement of the critical ubiquitin-interacting Arg residue. However, the interaction between HDAC6/USP5 and ubiquitin includes many more residues. HDAC6 and USP5 Znf-UBPs have a pocket where the C-terminus of ubiquitin inserts. In both proteins, arginine and tyrosine residues (R221 and Y223 in USP5; R1155 and Y1156 in HDAC6) from one side of the Znf-UBP pocket form critical hydrogen bonds with arginine and glycine of the C-terminal L-R-G-G motif of ubiquitin. The NMR structure of USP33's Znf-UBP domain [84] shows serine (S86) and glutamic acid (E87) residues in place of the tyrosine and arginine residues, respectively. These smaller

side chains do not form the same tight pocket as seen in HDAC6 and USP5 (Figures 6.2 and 6.3), and are unlikely to be able to hydrogen bond with ubiquitin.

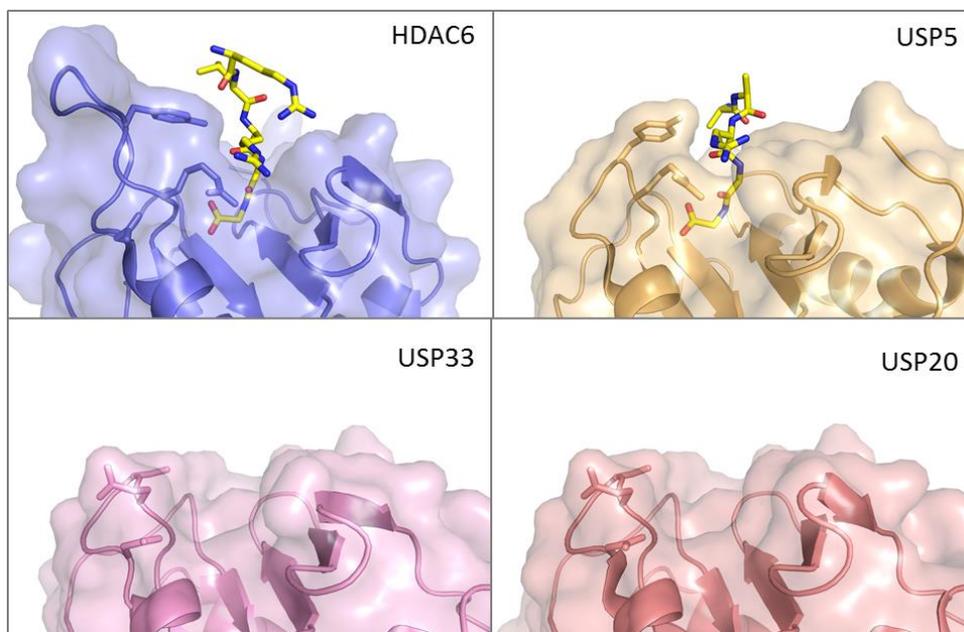


Figure 6.2. Ubiquitin-Znf-UBPs interactions. HDAC5 and USP5 crystal structures are shown (top). The NMR structure of USP33 is shown bottom left. A homology model of USP20 is shown bottom right. The surface of each Znf-UBP is shown. Defined binding pockets (white box) can be seen in HDAC6 and USP5, but not in USP20 or USP33. Side chains of selected residues are shown in stick form: R221, Y223 and N231 of USP5; R1155 and Y1156 of HDAC; S86, E87 and D90 of USP33; S55, E56 and D59 of USP20.

In addition, USP33 has an aspartic acid (D90) where a glycine and asparagine are found in HDAC6 and USP5, respectively (N231 in USP5 and G1159 in HDAC6). The side chain of this residue is found close to where the C-terminus of ubiquitin would be if it bound. The two negative charges would likely make it energetically unfavourable for ubiquitin to bind to the Znf-UBP domain of USP33. USP20 has the identical residues according to an alignment of the USP20 and USP33 sequences

(USP20 S55, E56 and D59). It is likely that all of these amino acid replacements are the cause of USP20's (and USP33's) inability to bind ubiquitin.

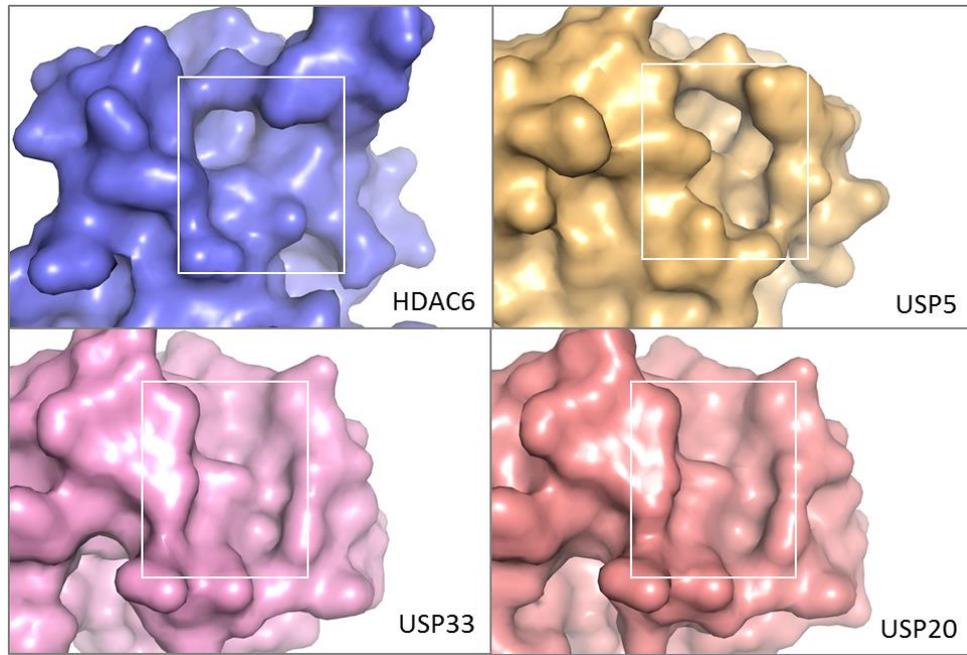


Figure 6.3. Binding pockets of Znf-UBPs. HDAC5 and USP5 crystal structures are shown (top). The NMR structure of USP33 is shown bottom left. A homology model of USP20 is shown bottom right. The surface of each Znf-UBP is shown. Defined binding pockets (white box) can be seen in HDAC6 and USP5, but not in USP20 or USP33.

7 Further studies

In order to crystallise the DUSP domains, it may be necessary to use a protein sequence from other species. Other species may be less hydrophobic, or have a more stable N-terminus, and thus expressing the DUSP domains from these orthologous genes may produce a stable, crystallisable protein. Also, thermophilic species, such as *Tetrahymena sp.*, may have similar proteins that are more stable due to the evolution of the organism into a thermophilic environment. This increase in stability can translate into increasing success of crystallisation in some cases [375]. Alternatively, other structural analysis could be performed, such as NMR. The structure of the Znf-UBP domain of USP33 was solved in this manner, so it should be achievable to obtain the structure of USP20's Znf-UBP with NMR.

In order to further characterise the interaction with USP20 and PLK1, crystallography, cellular assays and further *in vitro* assays would be valuable. Assessing the binding of phosphopeptides from the catalytic domain polobox motifs would show whether these sequences could bind to PLK1 in normal physiology. Also, the purified poloboxes protein used for the binding assays is a crystallisable fragment {Elia, 2003 #513}. This could be crystallised to show how these interactions occur, which would likely be similar to the canonical binding mode of the known phosphopeptides. Cellular assays could show that interactions occur inside cells. Mammalian vectors containing USP20 and PLK1 sequences could be used for immunoprecipitation assays. By truncation of the DUSP domains and poloboxes, and mutation of the catalytic insert motifs, the extent of their role

to the interaction could be seen. Split beta systems could also be used to visualise location and see temporal interactions between USP20 and PLK1.

Validation of the proteins identified by the pull down assays and mass spectroscopy is required. The proteins could be expressed and purified, and the binding could be investigated using the purified USP20 constructs and the putative interacting proteins.

8 References

1. Pray, L.A., *Eukaryotic Genome Complexity*. Nature Education, 2008. **1**(1): p. 96.
2. Milo, R., *What is the total number of protein molecules per cell volume? A call to rethink some published values*. Bioessays, 2013. **35**(12): p. 1050-5.
3. Lothrop, A.P., M.P. Torres, and S.M. Fuchs, *Deciphering post-translational modification codes*. FEBS Lett, 2013. **587**(8): p. 1247-57.
4. Khoury, G.A., R.C. Baliban, and C.A. Floudas, *Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database*. Sci Rep, 2011. **1**.
5. Patterson, H., et al., *Protein kinase inhibitors in the treatment of inflammatory and autoimmune diseases*. Clin Exp Immunol, 2014. **176**(1): p. 1-10.
6. Grant, S.K., *Therapeutic protein kinase inhibitors*. Cell Mol Life Sci, 2009. **66**(7): p. 1163-77.
7. Eldridge, A.G. and T. O'Brien, *Therapeutic strategies within the ubiquitin proteasome system*. Cell Death Differ, 2010. **17**(1): p. 4-13.
8. Field-Smith, A., G.J. Morgan, and F.E. Davies, *Bortezomib (Velcade/trade mark) in the Treatment of Multiple Myeloma*. Ther Clin Risk Manag, 2006. **2**(3): p. 271-9.
9. D'Arcy, P., X. Wang, and S. Linder, *Deubiquitinase inhibition as a cancer therapeutic strategy*. Pharmacol Ther, 2015. **147**: p. 32-54.
10. Pfoh, R., I.K. Lacroix, and V. Saridakis, *Deubiquitinases and the new therapeutic opportunities offered to cancer*. Endocr Relat Cancer, 2015. **22**(1): p. T35-54.
11. Goldstein, G., et al., *Isolation of a polypeptide that has lymphocyte-differentiating properties and is probably represented universally in living cells*. Proc Natl Acad Sci U S A, 1975. **72**(1): p. 11-5.
12. Schlesinger, D.H., G. Goldstein, and H.D. Niall, *The complete amino acid sequence of ubiquitin, an adenylate cyclase stimulating polypeptide probably universal in living cells*. Biochemistry, 1975. **14**(10): p. 2214-8.
13. Low, T.L. and A.L. Goldstein, *The chemistry and biology of thymosin. II. Amino acid sequence analysis of thymosin alpha1 and polypeptide beta1*. J Biol Chem, 1979. **254**(3): p. 987-95.
14. Ciechanover, A., *Tracing the history of the ubiquitin proteolytic system: the pioneering article*. Biochem Biophys Res Commun, 2009. **387**(1): p. 1-10.
15. Goldknopf, I.L. and H. Busch, *Isopeptide linkage between nonhistone and histone 2A polypeptides of chromosomal conjugate-protein A24*. Proc Natl Acad Sci U S A, 1977. **74**(3): p. 864-8.
16. Hunt, L.T. and M.O. Dayhoff, *Amino-terminal sequence identity of ubiquitin and the nonhistone component of nuclear protein A24*. Biochem Biophys Res Commun, 1977. **74**(2): p. 650-5.
17. Ciechanover, A., Y. Hod, and A. Hershko, *A heat-stable polypeptide component of an ATP-dependent proteolytic system from reticulocytes*. 1978. Biochem Biophys Res Commun, 2012. **425**(3): p. 565-70.
18. Wilkinson, K.D., M.K. Urban, and A.L. Haas, *Ubiquitin is the ATP-dependent proteolysis factor I of rabbit reticulocytes*. J Biol Chem, 1980. **255**(16): p. 7529-32.
19. Hershko, A., et al., *Proposed role of ATP in protein breakdown: conjugation of protein with multiple chains of the polypeptide of ATP-dependent proteolysis*. Proc Natl Acad Sci U S A, 1980. **77**(4): p. 1783-6.
20. Neefjes, J., T.A. Groothuis, and N.P. Dantuma, *[The 2004 Nobel Prize in Chemistry for the discovery of ubiquitin-mediated protein degradation]*. Ned Tijdschr Geneesk, 2004. **148**(52): p. 2579-82.
21. Baker, R.T. and P.G. Board, *The human ubiquitin-52 amino acid fusion protein gene shares several structural features with mammalian ribosomal protein genes*. Nucleic Acids Res, 1991. **19**(5): p. 1035-40.
22. Lund, P.K., et al., *Nucleotide sequence analysis of a cDNA encoding human ubiquitin reveals that ubiquitin is synthesized as a precursor*. J Biol Chem, 1985. **260**(12): p. 7609-13.
23. Ozkaynak, E., et al., *The yeast ubiquitin genes: a family of natural gene fusions*. EMBO J, 1987. **6**(5): p. 1429-39.

24. Ozkaynak, E., D. Finley, and A. Varshavsky, *The yeast ubiquitin gene: head-to-tail repeats encoding a polyubiquitin precursor protein*. *Nature*, 1984. **312**(5995): p. 663-6.
25. Baker, R.T. and P.G. Board, *The human ubiquitin gene family: structure of a gene and pseudogenes from the Ub B subfamily*. *Nucleic Acids Res*, 1987. **15**(2): p. 443-63.
26. Wiborg, O., et al., *The human ubiquitin multigene family: some genes contain multiple directly repeated ubiquitin coding sequences*. *EMBO J*, 1985. **4**(3): p. 755-9.
27. Sloper-Mould, K.E., et al., *Distinct functional surface regions on ubiquitin*. *J Biol Chem*, 2001. **276**(32): p. 30483-9.
28. Vijay-Kumar, S., C.E. Bugg, and W.J. Cook, *Structure of ubiquitin refined at 1.8 Å resolution*. *J Mol Biol*, 1987. **194**(3): p. 531-44.
29. Kiel, C. and L. Serrano, *The ubiquitin domain superfold: structure-based sequence alignments and characterization of binding epitopes*. *J Mol Biol*, 2006. **355**(4): p. 821-44.
30. Vijay-Kumar, S., et al., *Three-dimensional structure of ubiquitin at 2.8 Å resolution*. *Proc Natl Acad Sci U S A*, 1985. **82**(11): p. 3582-5.
31. Bayer, P., et al., *Structure determination of the small ubiquitin-related modifier SUMO-1*. *J Mol Biol*, 1998. **280**(2): p. 275-86.
32. Rao-Naik, C., et al., *The rub family of ubiquitin-like proteins. Crystal structure of Arabidopsis rub1 and expression of multiple rubs in Arabidopsis*. *J Biol Chem*, 1998. **273**(52): p. 34976-82.
33. Aramini, J.M., et al., *Solution NMR Structure of the C-terminal Domain of the Interferon Alpha-inducible Isg15 Protein From Homo Sapiens. Northeast Structural Genomics Target Hr2873b*. 2006.
34. Paramanathan, R., et al., *Crystal Structure of Human Gamma-aminobutyric Acid Receptor-Associated Protein-like 1 (Gabarap1), Isoform Cra_a*. 2007.
35. Komander, D., M.J. Clague, and S. Urbe, *Breaking the chains: structure and function of the deubiquitinases*. *Nat Rev Mol Cell Biol*, 2009. **10**(8): p. 550-63.
36. Pickart, C.M., *Mechanisms underlying ubiquitination*. *Annu Rev Biochem*, 2001. **70**: p. 503-33.
37. Komander, D. and M. Rape, *The ubiquitin code*. *Annu Rev Biochem*, 2012. **81**: p. 203-29.
38. Haas, A.L. and I.A. Rose, *The mechanism of ubiquitin activating enzyme. A kinetic and equilibrium analysis*. *J Biol Chem*, 1982. **257**(17): p. 10329-37.
39. Haas, A.L., et al., *Ubiquitin-activating enzyme. Mechanism and role in protein-ubiquitin conjugation*. *J Biol Chem*, 1982. **257**(5): p. 2543-8.
40. Groettrup, M., et al., *Activating the ubiquitin family: UBA6 challenges the field*. *Trends Biochem Sci*, 2008. **33**(5): p. 230-7.
41. Schafer, A., M. Kuhn, and H. Schindelin, *Structure of the ubiquitin-activating enzyme loaded with two ubiquitin molecules*. *Acta Crystallogr D Biol Crystallogr*, 2014. **70**(Pt 5): p. 1311-20.
42. Lee, I. and H. Schindelin, *Structural insights into E1-catalyzed ubiquitin activation and transfer to conjugating enzymes*. *Cell*, 2008. **134**(2): p. 268-78.
43. Ye, Y. and M. Rape, *Building ubiquitin chains: E2 enzymes at work*. *Nat Rev Mol Cell Biol*, 2009. **10**(11): p. 755-64.
44. Michelle, C., et al., *What was the set of ubiquitin and ubiquitin-like conjugating enzymes in the eukaryote common ancestor?* *J Mol Evol*, 2009. **68**(6): p. 616-28.
45. Gray, K.A., et al., *Genenames.org: the HGNC resources in 2015*. *Nucleic Acids Res*, 2014.
46. van Wijk, S.J. and H.T. Timmers, *The family of ubiquitin-conjugating enzymes (E2s): deciding between life and death of proteins*. *FASEB J*, 2010. **24**(4): p. 981-93.
47. Olsen, S.K. and C.D. Lima, *Structure of a ubiquitin E1-E2 complex: insights to E1-E2 thioester transfer*. *Molecular cell*, 2013. **49**(5): p. 884-896.
48. Metzger, M.B., V.A. Hristova, and A.M. Weissman, *HECT and RING finger families of E3 ubiquitin ligases at a glance*. *J Cell Sci*, 2012. **125**(Pt 3): p. 531-7.
49. Huang, L., et al., *Structure of an E6AP-UbcH7 complex: insights into ubiquitination by the E2-E3 enzyme cascade*. *Science*, 1999. **286**(5443): p. 1321-6.

50. Zheng, N., et al., *Structure of a c-Cbl-UbcH7 complex: RING domain function in ubiquitin-protein ligases*. Cell, 2000. **102**(4): p. 533-9.
51. Chau, V., et al., *A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein*. Science, 1989. **243**(4898): p. 1576-83.
52. Sorokin, A.V., E.R. Kim, and L.P. Ovchinnikov, *Proteasome system of protein degradation and processing*. Biochemistry (Mosc), 2009. **74**(13): p. 1411-42.
53. Deveraux, Q., et al., *A 26 S protease subunit that binds ubiquitin conjugates*. J Biol Chem, 1994. **269**(10): p. 7059-61.
54. Husnjak, K., et al., *Proteasome subunit Rpn13 is a novel ubiquitin receptor*. Nature, 2008. **453**(7194): p. 481-8.
55. Nathan, J.A., et al., *Why do cellular proteins linked to K63-polyubiquitin chains not associate with proteasomes?* Embo j, 2013. **32**(4): p. 552-65.
56. Erpapazoglou, Z., O. Walker, and R. Haguenuer-Tsapis, *Versatile roles of k63-linked ubiquitin chains in trafficking*. Cells, 2014. **3**(4): p. 1027-88.
57. Lauwers, E., C. Jacob, and B. Andre, *K63-linked ubiquitin chains as a specific signal for protein sorting into the multivesicular body pathway*. J Cell Biol, 2009. **185**(3): p. 493-502.
58. Messick, T.E. and R.A. Greenberg, *The ubiquitin landscape at DNA double-strand breaks*. J Cell Biol, 2009. **187**(3): p. 319-26.
59. Brown, J.S. and S.P. Jackson, *Ubiquitylation, neddylation and the DNA damage response*. Open Biol, 2015. **5**(4): p. 150018.
60. Harikumar, K.B., et al., *K63-linked polyubiquitination of transcription factor IRF1 is essential for IL-1-induced production of chemokines CXCL10 and CCL5*. Nat Immunol, 2014. **15**(3): p. 231-8.
61. Locke, M., J.I. Toth, and M.D. Petroski, *K11- and K48-Linked Ubiquitin Chains Interact with p97 during Endoplasmic Reticulum-Associated Degradation*. The Biochemical journal, 2014. **459**(1): p. 205-216.
62. Wickliffe, K.E., et al., *K11-linked ubiquitin chains as novel regulators of cell division*. Trends Cell Biol, 2011. **21**(11): p. 656-63.
63. Morris, J.R. and E. Solomon, *BRCA1 : BARD1 induces the formation of conjugated ubiquitin structures, dependent on K6 of ubiquitin, in cells during DNA replication and repair*. Hum Mol Genet, 2004. **13**(8): p. 807-17.
64. Komander, D., *Mechanism, specificity and structure of the deubiquitinases*. Subcell Biochem, 2010. **54**: p. 69-87.
65. Reyes-Turcu, F.E., K.H. Ventii, and K.D. Wilkinson, *Regulation and cellular roles of ubiquitin-specific deubiquitinating enzymes*. Annu Rev Biochem, 2009. **78**: p. 363-97.
66. Amerik, A.Y. and M. Hochstrasser, *Mechanism and function of deubiquitinating enzymes*. Biochim Biophys Acta, 2004. **1695**(1-3): p. 189-207.
67. Hu, M., et al., *Crystal Structure of a UBP-Family Deubiquitinating Enzyme in Isolation and in Complex with Ubiquitin Aldehyde*. Cell. **111**(7): p. 1041-1054.
68. Clague, M.J., et al., *Deubiquitylases from genes to organism*. Physiol Rev, 2013. **93**(3): p. 1289-315.
69. Zhang, W., et al., *Contribution of active site residues to substrate hydrolysis by USP2: insights into catalysis by ubiquitin specific proteases*. Biochemistry, 2011. **50**(21): p. 4775-85.
70. García-Santisteban, I., et al., *USP1 deubiquitinase: cellular functions, regulatory mechanisms and emerging potential as target in cancer therapy*. Molecular Cancer, 2013. **12**: p. 91-91.
71. Thorne, C., et al., *Isoform-specific localization of the deubiquitinase USP33 to the Golgi apparatus*. Traffic, 2011. **12**(11): p. 1563-74.
72. Valero, R., et al., *Characterization of alternatively spliced products and tissue-specific isoforms of USP28 and USP25*. Genome Biol, 2001. **2**(10): p. Research0043.
73. Sahtoe, D.D. and T.K. Sixma, *Layers of DUB regulation*. Trends in Biochemical Sciences, 2015. **40**(8): p. 456-467.

74. Lee, M.J., et al., *Trimming of Ubiquitin Chains by Proteasome-associated Deubiquitinating Enzymes*. Molecular & Cellular Proteomics : MCP, 2011. **10**(5): p. R110.003871.
75. Coyne, E.S. and S.S. Wing, *The business of deubiquitination – location, location, location*. F1000Research, 2016. **5**: p. F1000 Faculty Rev-163.
76. Kim, R.Q., W.J. van Dijk, and T.K. Sixma, *Structure of USP7 catalytic domain and three Ubl-domains reveals a connector α -helix with regulatory role*. Journal of Structural Biology, 2016. **195**(1): p. 11-18.
77. Hutti, J.E., et al., *IkappaB kinase beta phosphorylates the K63 deubiquitinase A20 to cause feedback inhibition of the NF-kappaB pathway*. Mol Cell Biol, 2007. **27**(21): p. 7451-61.
78. Huang, O.W., et al., *Phosphorylation-dependent activity of the deubiquitinase DUBA*. Nat Struct Mol Biol, 2012. **19**(2): p. 171-5.
79. Huang, X., et al., *Deubiquitinase USP37 is activated by CDK2 to antagonize APC(CDH1) and promote S phase entry*. Mol Cell, 2011. **42**(4): p. 511-23.
80. Khoronenkova, S.V., et al., *ATM-dependent downregulation of USP7/HAUSP by PPM1G activates p53 response to DNA damage*. Mol Cell, 2012. **45**(6): p. 801-13.
81. Hutti, J.E., et al., *Phosphorylation of the tumor suppressor CYLD by the breast cancer oncogene IKKepsilon promotes cell transformation*. Mol Cell, 2009. **34**(4): p. 461-72.
82. Villamil, M.A., et al., *Serine phosphorylation is critical for the activation of ubiquitin-specific protease 1 and its interaction with WD40-repeat protein UAF1*. Biochemistry, 2012. **51**(45): p. 9112-23.
83. Li, Z., et al., *Identification of a deubiquitinating enzyme subfamily as substrates of the von Hippel-Lindau tumor suppressor*. Biochem Biophys Res Commun, 2002. **294**(3): p. 700-9.
84. Allen, M.D. and M. Bycroft, *The solution structure of the ZnF UBP domain of USP33/VDU1*. Protein Sci, 2007. **16**(9): p. 2072-5.
85. Ouyang, H., et al., *Protein aggregates are recruited to aggresome by histone deacetylase 6 via unanchored ubiquitin C termini*. J Biol Chem, 2012. **287**(4): p. 2317-27.
86. Reyes-Turcu, F.E., et al., *The ubiquitin binding domain ZnF UBP recognizes the C-terminal diglycine motif of unanchored ubiquitin*. Cell, 2006. **124**(6): p. 1197-208.
87. Ye, Y., et al., *Dissection of USP catalytic domains reveals five common insertion points*. Mol Biosyst, 2009. **5**(12): p. 1797-808.
88. Berthouze, M., et al., *The deubiquitinases USP33 and USP20 coordinate beta2 adrenergic receptor recycling and resensitization*. Embo j, 2009. **28**(12): p. 1684-96.
89. Curcio-Morelli, C., et al., *Deubiquitination of type 2 iodothyronine deiodinase by von Hippel-Lindau protein-interacting deubiquitinating enzymes regulates thyroid hormone activation*. J Clin Invest, 2003. **112**(2): p. 189-96.
90. Yuasa-Kawada, J., et al., *Midline crossing and Slit responsiveness of commissural axons require USP33*. Nat Neurosci, 2009. **12**(9): p. 1087-1089.
91. Li, Z., et al., *VHL protein-interacting deubiquitinating enzyme 2 deubiquitinates and stabilizes HIF-1alpha*. EMBO Rep, 2005. **6**(4): p. 373-8.
92. Koscielny, G., et al., *The International Mouse Phenotyping Consortium Web Portal, a unified point of access for knockout mice and related phenotyping data*. Nucleic Acids Research, 2014. **42**(Database issue): p. D802-D809.
93. Chatzinikolaou, G., I. Karakasilioti, and G.A. Garinis, *DNA damage and innate immunity: links and trade-offs*. Trends Immunol, 2014. **35**(9): p. 429-35.
94. Tarr, A.J., et al., *β -ADRENERGIC RECEPTOR MEDIATED INCREASES IN ACTIVATION AND FUNCTION OF NATURAL KILLER CELLS FOLLOWING REPEATED SOCIAL DISRUPTION*. Brain, behavior, and immunity, 2012. **26**(8): p. 1226-1238.
95. Nizet, V. and R.S. Johnson, *Interdependence of hypoxic and innate immune responses*. Nat Rev Immunol, 2009. **9**(9): p. 609-617.
96. Yasunaga, J., et al., *Ubiquitin-specific peptidase 20 targets TRAF6 and human T cell leukemia virus type 1 tax to negatively regulate NF-kappaB signaling*. J Virol, 2011. **85**(13): p. 6212-9.

97. Jean-Charles, P.Y., et al., *Ubiquitin-specific Protease 20 Regulates the Reciprocal Functions of beta-Arrestin2 in Toll-like Receptor 4-promoted Nuclear Factor kappaB (NFkappaB) Activation*. J Biol Chem, 2016. **291**(14): p. 7450-64.
98. Dev, A., et al., *NF-kappaB and innate immunity*. Curr Top Microbiol Immunol, 2011. **349**: p. 115-43.
99. Arrojo, E.D.R. and A.C. Bianco, *Type 2 deiodinase at the crossroads of thyroid hormone action*. Int J Biochem Cell Biol, 2011. **43**(10): p. 1432-41.
100. Bianco, A.C. and P.R. Larsen, *Cellular and structural biology of the deiodinases*. Thyroid, 2005. **15**(8): p. 777-86.
101. Darras, V.M. and S.L. Van Herck, *Iodothyronine deiodinase structure and function: from ascidians to humans*. J Endocrinol, 2012. **215**(2): p. 189-206.
102. Flamant, F., K. Gauthier, and J. Samarut, *Thyroid hormones signaling is getting more complex: STORMs are coming*. Mol Endocrinol, 2007. **21**(2): p. 321-33.
103. Bassett, J.H., C.B. Harvey, and G.R. Williams, *Mechanisms of thyroid hormone receptor-specific nuclear and extra nuclear actions*. Mol Cell Endocrinol, 2003. **213**(1): p. 1-11.
104. Sagar, G.D., et al., *Ubiquitination-induced conformational change within the deiodinase dimer is a switch regulating enzyme activity*. Mol Cell Biol, 2007. **27**(13): p. 4774-83.
105. St Germain, D.L., V.A. Galton, and A. Hernandez, *Minireview: Defining the roles of the iodothyronine deiodinases: current concepts and challenges*. Endocrinology, 2009. **150**(3): p. 1097-107.
106. Casula, S. and A.C. Bianco, *Thyroid hormone deiodinases and cancer*. Front Endocrinol (Lausanne), 2012. **3**: p. 74.
107. Murakami, M., et al., *Expression of type II iodothyronine deiodinase in brain tumors*. J Clin Endocrinol Metab, 2000. **85**(11): p. 4403-6.
108. Pugh, T.J., et al., *Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations*. Nature, 2012. **488**(7409): p. 106-10.
109. Bamford, S., et al., *The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website*. British Journal of Cancer, 2004. **91**(2): p. 355-358.
110. Pause, A., et al., *The von Hippel-Lindau tumor-suppressor gene product forms a stable complex with human CUL-2, a member of the Cdc53 family of proteins*. Proceedings of the National Academy of Sciences of the United States of America, 1997. **94**(6): p. 2156-2161.
111. Cai, W. and H. Yang, *The structure and regulation of Cullin 2 based E3 ubiquitin ligases and their biological functions*. Cell Division, 2016. **11**: p. 7.
112. Nguyen, H.C., et al., *Insights into Cullin-RING E3 ubiquitin ligase recruitment: Structure of the VHL-EloBC-Cul2 complex*. Structure (London, England : 1993), 2015. **23**(3): p. 441-449.
113. Kamura, T., et al., *Rbx1, a component of the VHL tumor suppressor complex and SCF ubiquitin ligase*. Science, 1999. **284**(5414): p. 657-61.
114. Kamura, T., et al., *VHL-box and SOCS-box domains determine binding specificity for Cul2-Rbx1 and Cul5-Rbx2 modules of ubiquitin ligases*. Genes & Development, 2004. **18**(24): p. 3055-3065.
115. Wang, G.L., et al., *Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellular O2 tension*. Proceedings of the National Academy of Sciences of the United States of America, 1995. **92**(12): p. 5510-5514.
116. Groulx, I. and S. Lee, *Oxygen-dependent ubiquitination and degradation of hypoxia-inducible factor requires nuclear-cytoplasmic trafficking of the von Hippel-Lindau tumor suppressor protein*. Mol Cell Biol, 2002. **22**(15): p. 5319-36.
117. Harris, A.L., *Hypoxia--a key regulatory factor in tumour growth*. Nat Rev Cancer, 2002. **2**(1): p. 38-47.
118. Birk, D.M., et al., *Current insights on the biology and clinical aspects of VEGF regulation*. Vasc Endovascular Surg, 2008. **42**(6): p. 517-30.

119. Semenza, G.L., *HIF-1: upstream and downstream of cancer metabolism*. *Curr Opin Genet Dev*, 2010. **20**(1): p. 51-6.
120. Srinivas, V., et al., *Characterization of an oxygen/redox-dependent degradation domain of hypoxia-inducible factor alpha (HIF-alpha) proteins*. *Biochem Biophys Res Commun*, 1999. **260**(2): p. 557-61.
121. Jaakkola, P., et al., *Targeting of HIF-alpha to the von Hippel-Lindau ubiquitylation complex by O2-regulated prolyl hydroxylation*. *Science*, 2001. **292**(5516): p. 468-72.
122. Masson, N., et al., *Independent function of two destruction domains in hypoxia-inducible factor-alpha chains activated by prolyl hydroxylation*. *Embo j*, 2001. **20**(18): p. 5197-206.
123. Yu, F., et al., *HIF-1alpha binding to VHL is regulated by stimulus-sensitive proline hydroxylation*. *Proc Natl Acad Sci U S A*, 2001. **98**(17): p. 9630-5.
124. Masson, N. and P.J. Ratcliffe, *HIF prolyl and asparaginyl hydroxylases in the biological response to intracellular O(2) levels*. *J Cell Sci*, 2003. **116**(Pt 15): p. 3041-9.
125. Ke, Q. and M. Costa, *Hypoxia-inducible factor-1 (HIF-1)*. *Mol Pharmacol*, 2006. **70**(5): p. 1469-80.
126. Pugh, C.W. and P.J. Ratcliffe, *The von Hippel-Lindau tumor suppressor, hypoxia-inducible factor-1 (HIF-1) degradation, and cancer pathogenesis*. *Semin Cancer Biol*, 2003. **13**(1): p. 83-9.
127. Ivan, M., et al., *HIFalpha targeted for VHL-mediated destruction by proline hydroxylation: implications for O2 sensing*. *Science*, 2001. **292**(5516): p. 464-8.
128. Hon, W.C., et al., *Structural basis for the recognition of hydroxyproline in HIF-1 alpha by pVHL*. *Nature*, 2002. **417**(6892): p. 975-8.
129. Min, J.H., et al., *Structure of an HIF-1alpha -pVHL complex: hydroxyproline recognition in signaling*. *Science*, 2002. **296**(5574): p. 1886-9.
130. Kamura, T., et al., *Activation of HIF1alpha ubiquitination by a reconstituted von Hippel-Lindau (VHL) tumor suppressor complex*. *Proc Natl Acad Sci U S A*, 2000. **97**(19): p. 10430-5.
131. Tanimoto, K., et al., *Mechanism of regulation of the hypoxia-inducible factor-1 alpha by the von Hippel-Lindau tumor suppressor protein*. *Embo j*, 2000. **19**(16): p. 4298-309.
132. Paltoglou, S. and B.J. Roberts, *HIF-1alpha and EPAS ubiquitination mediated by the VHL tumour suppressor involves flexibility in the ubiquitination mechanism, similar to other RING E3 ligases*. *Oncogene*, 2007. **26**(4): p. 604-9.
133. *Ongoing and future developments at the Universal Protein Resource*. *Nucleic Acids Res*, 2011. **39**(Database issue): p. D214-9.
134. Li, Z., et al., *Ubiquitination of a novel deubiquitinating enzyme requires direct binding to von Hippel-Lindau tumor suppressor protein*. *J Biol Chem*, 2002. **277**(7): p. 4656-62.
135. Barontini, M. and P.L. Dahia, *VHL disease*. *Best Pract Res Clin Endocrinol Metab*, 2010. **24**(3): p. 401-13.
136. Gnarr, J.R., et al., *Mutations of the VHL tumour suppressor gene in renal carcinoma*. *Nat Genet*, 1994. **7**(1): p. 85-90.
137. Rosenbaum, D.M., S.G.F. Rasmussen, and B.K. Kobilka, *The structure and function of G-protein-coupled receptors*. *Nature*, 2009. **459**(7245): p. 356-363.
138. Johnson, M., *Molecular mechanisms of beta(2)-adrenergic receptor function, response, and regulation*. *J Allergy Clin Immunol*, 2006. **117**(1): p. 18-24; quiz 25.
139. Pierce, K.L., R.T. Premont, and R.J. Lefkowitz, *Seven-transmembrane receptors*. *Nat Rev Mol Cell Biol*, 2002. **3**(9): p. 639-650.
140. Benovic, J.L., *Novel beta2-adrenergic receptor signaling pathways*. *J Allergy Clin Immunol*, 2002. **110**(6 Suppl): p. S229-35.
141. Penn, R.B., A.N. Pronin, and J.L. Benovic, *Regulation of G protein-coupled receptor kinases*. *Trends Cardiovasc Med*, 2000. **10**(2): p. 81-9.
142. Mohan, M.L., et al., *G-Protein Coupled Receptor Resensitization – Appreciating the Balancing Act of Receptor Function*. *Current molecular pharmacology*, 2012.

143. Shenoy, S.K., *Seven-Transmembrane Receptors and Ubiquitination*. *Circulation research*, 2007. **100**(8): p. 1142-1154.
144. Shenoy, S.K., et al., *Regulation of receptor fate by ubiquitination of activated beta 2-adrenergic receptor and beta-arrestin*. *Science*, 2001. **294**(5545): p. 1307-13.
145. Nobles, K.N., et al., *Distinct phosphorylation sites on the beta(2)-adrenergic receptor establish a barcode that encodes differential functions of beta-arrestin*. *Sci Signal*, 2011. **4**(185): p. ra51.
146. Nabhan, J.F., H. Pan, and Q. Lu, *Arrestin domain-containing protein 3 recruits the NEDD4 E3 ligase to mediate ubiquitination of the beta2-adrenergic receptor*. *EMBO Reports*, 2010. **11**(8): p. 605-611.
147. Shenoy, S.K., et al., *Nedd4 mediates agonist-dependent ubiquitination, lysosomal targeting, and degradation of the beta2-adrenergic receptor*. *J Biol Chem*, 2008. **283**(32): p. 22166-76.
148. Xiao, K. and S.K. Shenoy, *Beta2-adrenergic receptor lysosomal trafficking is regulated by ubiquitination of lysyl residues in two distinct receptor domains*. *J Biol Chem*, 2011. **286**(14): p. 12785-95.
149. Shenoy, S.K., et al., *Beta-Arrestin-dependent signaling and trafficking of 7-transmembrane receptors is reciprocally regulated by the deubiquitinase USP33 and the E3 ligase Mdm2*. *Proceedings of the National Academy of Sciences of the United States of America*, 2009. **106**(16): p. 6650-6655.
150. Xie, L., et al., *Oxygen-regulated beta(2)-adrenergic receptor hydroxylation by EGLN3 and ubiquitylation by pVHL*. *Sci Signal*, 2009. **2**(78): p. ra33.
151. Yasunaga, J., et al., *Ubiquitin-Specific Peptidase 20 Targets TRAF6 and Human T Cell Leukemia Virus Type 1 Tax To Negatively Regulate NF-kappaB Signaling*. *Journal of Virology*, 2011. **85**(13): p. 6212-6219.
152. Boxus, M., et al., *The HTLV-1 Tax interactome*. *Retrovirology*, 2008. **5**: p. 76-76.
153. Grassmann, R., M. Aboud, and K.T. Jeang, *Molecular mechanisms of cellular transformation by HTLV-1 Tax*. *Oncogene*, 2005. **24**(39): p. 5976-85.
154. Jeang, K.T., et al., *Life, death, and tax: role of HTLV-I oncoprotein in genetic instability and cellular transformation*. *J Biol Chem*, 2004. **279**(31): p. 31991-4.
155. McGuire, K.L., et al., *Influence of human T-cell leukemia virus type I tax and rex on interleukin-2 gene expression*. *J Virol*, 1993. **67**(3): p. 1590-9.
156. Chiari, E., et al., *Stable ubiquitination of human T-cell leukemia virus type 1 tax is required for proteasome binding*. *J Virol*, 2004. **78**(21): p. 11823-32.
157. Lamsoul, I., et al., *Exclusive ubiquitination and sumoylation on overlapping lysine residues mediate NF-kappaB activation by the human T-cell leukemia virus tax oncoprotein*. *Mol Cell Biol*, 2005. **25**(23): p. 10391-406.
158. Wilkinson, K.A. and J.M. Henley, *Mechanisms, regulation and consequences of protein SUMOylation*. *The Biochemical journal*, 2010. **428**(2): p. 133-145.
159. Nasr, R., et al., *Tax ubiquitylation and sumoylation control critical cytoplasmic and nuclear steps of NF-kappaB activation*. *Blood*, 2006. **107**(10): p. 4021-9.
160. Peloponese, J.-M., et al., *Ubiquitination of Human T-Cell Leukemia Virus Type 1 Tax Modulates Its Activity*. *Journal of Virology*, 2004. **78**(21): p. 11686-11695.
161. Chu, Z.L., et al., *IKKgamma mediates the interaction of cellular I kappaB kinases with the tax transforming protein of human T cell leukemia virus type 1*. *J Biol Chem*, 1999. **274**(22): p. 15297-300.
162. O'Mahony, A.M., et al., *Human T-cell lymphotropic virus type 1 tax induction of biologically Active NF-kappaB requires I kappaB kinase-1-mediated phosphorylation of RelA/p65*. *J Biol Chem*, 2004. **279**(18): p. 18137-45.
163. Brockman, J.A., et al., *Coupling of a signal response domain in I kappa B alpha to multiple pathways for NF-kappa B activation*. *Mol Cell Biol*, 1995. **15**(5): p. 2809-18.

164. Gatza, M.L., T. Dayaram, and S.J. Marriott, *Ubiquitination of HTLV-I Tax in response to DNA damage regulates nuclear complex formation and nuclear export*. *Retrovirology*, 2007. **4**: p. 95-95.
165. Alefantis, T., et al., *Characterization of a nuclear export signal within the human T cell leukemia virus type I transactivator protein Tax*. *J Biol Chem*, 2003. **278**(24): p. 21814-22.
166. Gatza, M.L. and S.J. Marriott, *Genotoxic stress and cellular stress alter the subcellular distribution of human T-cell leukemia virus type 1 tax through a CRM1-dependent mechanism*. *J Virol*, 2006. **80**(13): p. 6657-68.
167. Deng, L., et al., *Activation of the I κ B kinase complex by TRAF6 requires a dimeric ubiquitin-conjugating enzyme complex and a unique polyubiquitin chain*. *Cell*, 2000. **103**(2): p. 351-61.
168. Wang, C., et al., *TAK1 is a ubiquitin-dependent kinase of MKK and IKK*. *Nature*, 2001. **412**(6844): p. 346-51.
169. Chen, Z.J., *Ubiquitin signalling in the NF-kappaB pathway*. *Nat Cell Biol*, 2005. **7**(8): p. 758-65.
170. Gorello, P., et al., *Combined interphase fluorescence in situ hybridization elucidates the genetic heterogeneity of T-cell acute lymphoblastic leukemia in adults*. *Haematologica*, 2010. **95**(1): p. 79-86.
171. Li, J., et al., *USP33 regulates centrosome biogenesis via de-ubiquitylation of a centriolar protein, CP110*. *Nature*, 2013. **495**(7440): p. 10.1038/nature11941.
172. Sowa, M.E., et al., *Defining the human deubiquitinating enzyme interaction landscape*. *Cell*, 2009. **138**(2): p. 389-403.
173. Schmidt, T.I., et al., *Control of centriole length by CPAP and CP110*. *Curr Biol*, 2009. **19**(12): p. 1005-11.
174. Spektor, A., et al., *Cep97 and CP110 Suppress a Cilia Assembly Program*. *Cell*. **130**(4): p. 678-690.
175. Franz, A., et al., *CP110 exhibits novel regulatory activities during centriole assembly in Drosophila*. *J Cell Biol*, 2013. **203**(5): p. 785-99.
176. Chen, Z., et al., *CP110, a cell cycle-dependent CDK substrate, regulates centrosome duplication in human cells*. *Dev Cell*, 2002. **3**(3): p. 339-50.
177. D'Angiolella, V., et al., *SCF(Cyclin F) controls centrosome homeostasis and mitotic fidelity through CP110 degradation*. *Nature*, 2010. **466**(7302): p. 138-42.
178. Skaar, J.R., J.K. Pagan, and M. Pagano, *Mechanisms and function of substrate recruitment by F-box proteins*. *Nat Rev Mol Cell Biol*, 2013. **14**(6): p. 369-381.
179. Li, J., et al., *Neurl4, a novel daughter centriole protein, prevents formation of ectopic microtubule organizing centres*. *EMBO Reports*, 2012. **13**(6): p. 547-553.
180. Buchholz, M., et al., *Transcriptome analysis of microdissected pancreatic intraepithelial neoplastic lesions*. *Oncogene*, 2005. **24**(44): p. 6626-36.
181. Badea, L., et al., *Combined gene expression analysis of whole-tissue and microdissected pancreatic ductal adenocarcinoma identifies genes specifically overexpressed in tumor epithelia*. *Hepatology*, 2008. **55**(88): p. 2016-27.
182. Shanmugam, I., et al., *Ubiquitin-specific peptidase 20 regulates Rad17 stability, checkpoint kinase 1 phosphorylation and DNA repair by homologous recombination*. *J Biol Chem*, 2014. **289**(33): p. 22739-48.
183. Yuan, J., et al., *HERC2-USP20 axis regulates DNA damage checkpoint through Claspin*. *Nucleic Acids Res*, 2014. **42**(21): p. 13110-21.
184. Zhu, M., et al., *HERC2/USP20 coordinates CHK1 activation by modulating CLASPIN stability*. *Nucleic Acids Res*, 2014. **42**(21): p. 13074-81.
185. Plimmer, R.A., *CXLIV.-The Proteins of egg-yolk*. *Journal of the Chemical Society, Transactions*, 1908. **93**: p. 1500-1506.
186. Krebs, E.G., *Historical perspectives on Protein Phosphorylation and a Classification System for Protein Kinases*. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 1983. **302**(1108): p. 3-11.

187. Cohen, P., *The origins of protein phosphorylation*. Nat Cell Biol, 2002. **4**(5): p. E127-30.
188. Heilmeyer, L.M.G., *Signal Transduction and Protein Phosphorylation*. 1987, New York and London: Plenum Press.
189. Kiessling, W., Biochem. Z., 1939. **302**: p. 50.
190. Cori, C.G. and A.A. Green, *Crystalline Muscle Phosphorylase II Prosthetic Group*. J. Biol. Chem., 1943. **151**: p. 31-38.
191. Fischer, E.H., et al., *Structure of the site phosphorylated in the phosphorylase b to a reaction*. J Biol Chem, 1959. **234**(7): p. 1698-704.
192. Krebs, E.G. and E.H. Fischer, *The phosphorylase b to a converting enzyme of rabbit skeletal muscle*. Biochim Biophys Acta, 1956. **20**(1): p. 150-7.
193. Fischer, E.H. and E.G. Krebs, *Conversion of phosphorylase b to phosphorylase a in muscle extracts*. J Biol Chem, 1955. **216**(1): p. 121-32.
194. Sutherland, E.W., Jr. and W.D. Wosilait, *Inactivation and activation of liver phosphorylase*. Nature, 1955. **175**(4447): p. 169-70.
195. Burnett, G. and E.P. Kennedy, *The enzymatic phosphorylation of proteins*. J Biol Chem, 1954. **211**(2): p. 969-80.
196. Lindsten, J. and N. Ringhertz, *The Nobel prize in physiology or medicine, 1901-2000*. The Nobel Prize: the first, 2001. **100**: p. 111-137.
197. Endicott, J.A., M.E. Noble, and L.N. Johnson, *The structural basis for control of eukaryotic protein kinases*. Annu Rev Biochem, 2012. **81**: p. 587-613.
198. Fabbro, D., S.W. Cowan-Jacob, and H. Moebitz, *Ten things you should know about protein kinases: IUPHAR Review 14*. Br J Pharmacol, 2015. **172**(11): p. 2675-700.
199. Berg, J., J. Tymoczko, and L. Stryer, *Phosphorylase is Regulated by Allosteric Interactions and Reversible Phosphorylation.*, in *Biochemistry*. 2002, WH Freeman: New York.
200. Johnson, L.N. and D. Barford, *The effects of phosphorylation on the structure and function of proteins*. Annu Rev Biophys Biomol Struct, 1993. **22**: p. 199-232.
201. Laskowski, R.A., F. Gerick, and J.M. Thornton, *The structural basis of allosteric regulation in proteins*. FEBS Lett, 2009. **583**(11): p. 1692-8.
202. Manning, G., et al., *The protein kinase complement of the human genome*. Science, 2002. **298**(5600): p. 1912-34.
203. Hanks, S.K. and T. Hunter, *Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification*. Faseb j, 1995. **9**(8): p. 576-96.
204. Hubbard, S.R. and J.H. Till, *Protein tyrosine kinase structure and function*. Annu Rev Biochem, 2000. **69**: p. 373-98.
205. Taylor, S.S. and A.P. Kornev, *Protein kinases: evolution of dynamic regulatory proteins*. Trends Biochem Sci, 2011. **36**(2): p. 65-77.
206. Ubersax, J.A. and J.E. Ferrell Jr, *Mechanisms of specificity in protein phosphorylation*. Nat Rev Mol Cell Biol, 2007. **8**(7): p. 530-541.
207. Nakajima, H., et al., *Identification of a consensus motif for Plk (Polo-like kinase) phosphorylation reveals Myt1 as a Plk1 substrate*. J Biol Chem, 2003. **278**(28): p. 25277-80.
208. Toyoshima-Morimoto, F., E. Taniguchi, and E. Nishida, *Plk1 promotes nuclear translocation of human Cdc25C during prophase*. EMBO Reports, 2002. **3**(4): p. 341-348.
209. Kim, S.T., et al., *Substrate specificities and identification of putative substrates of ATM kinase family members*. J Biol Chem, 1999. **274**(53): p. 37538-43.
210. O'Neill, T., et al., *Utilization of oriented peptide libraries to identify substrate motifs selected by ATM*. J Biol Chem, 2000. **275**(30): p. 22719-27.
211. Traven, A. and J. Heierhorst, *SQ/TQ cluster domains: concentrated ATM/ATR kinase phosphorylation site regions in DNA-damage-response proteins*. Bioessays, 2005. **27**(4): p. 397-407.
212. Elia, A.E., L.C. Cantley, and M.B. Yaffe, *Proteomic screen finds pSer/pThr-binding domain localizing Plk1 to mitotic substrates*. Science, 2003. **299**(5610): p. 1228-31.

213. Elia, A.E., et al., *The molecular basis for phosphodependent substrate targeting and regulation of Plks by the Polo-box domain*. Cell, 2003. **115**(1): p. 83-95.
214. Durocher, D., et al., *The FHA domain is a modular phosphopeptide recognition motif*. Mol Cell, 1999. **4**(3): p. 387-94.
215. Durocher, D., et al., *The molecular basis of FHA domain:phosphopeptide binding specificity and implications for phospho-dependent signaling mechanisms*. Mol Cell, 2000. **6**(5): p. 1169-82.
216. Rosano, G.L. and E.A. Ceccarelli, *Recombinant protein expression in Escherichia coli: advances and challenges*. Frontiers in Microbiology, 2014. **5**: p. 172.
217. Woestenenk, E.A., et al., *His tag effect on solubility of human proteins produced in Escherichia coli: a comparison between four expression vectors*. J Struct Funct Genomics, 2004. **5**(3): p. 217-29.
218. Smith, J.C., et al., *Chemical synthesis and cloning of a poly(arginine)-coding gene fragment designed to aid polypeptide purification*. Gene, 1984. **32**(3): p. 321-7.
219. Lichty, J.J., et al., *Comparison of affinity tags for protein purification*. Protein Expression and Purification, 2005. **41**(1): p. 98-105.
220. Wang, Z., et al., *Human SUMO fusion systems enhance protein expression and solubility*. Protein Expr Purif, 2010. **73**(2): p. 203-8.
221. Zhao, X., G. Li, and S. Liang, *Several affinity tags commonly used in chromatographic purification*. J Anal Methods Chem, 2013. **2013**: p. 581093.
222. Cherezov, V., et al., *High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor*. Science, 2007. **318**(5854): p. 1258-65.
223. Rosenbaum, D.M., et al., *GPCR engineering yields high-resolution structural insights into beta2-adrenergic receptor function*. Science, 2007. **318**(5854): p. 1266-73.
224. Zhan, Y., X. Song, and G.W. Zhou, *Structural analysis of regulatory protein domains using GST-fusion proteins*. Gene, 2001. **281**(1-2): p. 1-9.
225. Nauli, S., et al., *Polymer-driven crystallization*. Protein Sci, 2007. **16**(11): p. 2542-51.
226. Nanyes, D.R., et al., *Multiple polymer architectures of human polyhomeotic homolog 3 sterile alpha motif*. Proteins, 2014. **82**(10): p. 2823-30.
227. Waugh, D.S., *Crystal structures of MBP fusion proteins*. Protein Science, 2016. **25**(3): p. 559-571.
228. Smyth, D.R., et al., *Crystal structures of fusion proteins with large-affinity tags*. Protein Science : A Publication of the Protein Society, 2003. **12**(7): p. 1313-1322.
229. Moon, A.F., et al., *A synergistic approach to protein crystallization: combination of a fixed-arm carrier with surface entropy reduction*. Protein Sci, 2010. **19**(5): p. 901-13.
230. Carson, M., et al., *His-tag impact on structure*. Acta Crystallogr D Biol Crystallogr, 2007. **63**(Pt 3): p. 295-301.
231. Mattaini, K.R., et al., *An epitope tag alters phosphoglycerate dehydrogenase structure and impairs ability to support cell proliferation*. Cancer & Metabolism, 2015. **3**: p. 5.
232. Berman, H.M., et al., *The Protein Data Bank*. Nucleic Acids Res, 2000. **28**(1): p. 235-42.
233. Knecht, S., et al., *Oligohis-tags: mechanisms of binding to Ni²⁺-NTA surfaces*. J Mol Recognit, 2009. **22**(4): p. 270-9.
234. Nieba, L., et al., *BIACORE analysis of histidine-tagged proteins using a chelating NTA sensor chip*. Anal Biochem, 1997. **252**(2): p. 217-28.
235. Schmitt, J., H. Hess, and H.G. Stunnenberg, *Affinity purification of histidine-tagged proteins*. Mol Biol Rep, 1993. **18**(3): p. 223-30.
236. Bornhorst, J.A. and J.J. Falke, *[16] Purification of Proteins Using Polyhistidine Affinity Tags*. Methods in enzymology, 2000. **326**: p. 245-254.
237. Kipriyanov, S.M., et al., *Bacterial expression and refolding of single-chain Fv fragments with C-terminal cysteines*. Cell Biophys, 1995. **26**(3): p. 187-204.

238. Janknecht, R., et al., *Rapid and efficient purification of native histidine-tagged protein expressed by recombinant vaccinia virus*. Proc Natl Acad Sci U S A, 1991. **88**(20): p. 8972-6.
239. Sinha, D., M. Bakhshi, and R. Vora, *Ligand binding assays with recombinant proteins refolded on an affinity matrix*. Biotechniques, 1994. **17**(3): p. 509-12, 514.
240. Mori, S. and H.G. Barth, *Size exclusion chromatography*. 1999, Berlin Heidelberg: Springer
241. Erickson, H.P., *Size and Shape of Protein Molecules at the Nanometer Level Determined by Sedimentation, Gel Filtration, and Electron Microscopy*. Biological Procedures Online, 2009. **11**: p. 32-51.
242. Dale, G.E., C. Oefner, and A. D'Arcy, *The protein as a variable in protein crystallization*. J Struct Biol, 2003. **142**(1): p. 88-97.
243. Zulauf, M. and A. D'Arcy, *Light scattering of proteins as a criterion for crystallization*. Journal of Crystal Growth, 1992. **122**(1): p. 102-106.
244. Bragg, W.L., *The Structure of Some Crystals as Indicated by Their Diffraction of X-rays*. Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character, 1913. **89**(610): p. 248-277.
245. Green, D.W., V.M. Ingram, and M.F. Perutz, *The Structure of Haemoglobin. IV. Sign Determination by the Isomorphous Replacement Method*. Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 1954. **225**(1162): p. 287-307.
246. Kendrew, J.C., et al., *A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis*. Nature, 1958. **181**(4610): p. 662-666.
247. Berman, H.M., et al., *Trendspotting in the Protein Data Bank*. FEBS letters, 2013. **587**(8): p. 1036-1045.
248. Chayen, N.E., P.D. Shaw Stewart, and D.M. Blow, *Microbatch crystallization under oil — a new technique allowing many small-volume crystallization trials*. Journal of Crystal Growth, 1992. **122**(1): p. 176-180.
249. Krauss, I.R., et al., *An Overview of Biological Macromolecule Crystallization*. International Journal of Molecular Sciences, 2013. **14**(6): p. 11643-11691.
250. Dessau, M.A. and Y. Modis, *Protein Crystallization for X-ray Crystallography*. Journal of Visualized Experiments : JoVE, 2011(47): p. 2285.
251. Chayen, N.E., *Turning protein crystallisation from an art into a science*. Curr Opin Struct Biol, 2004. **14**(5): p. 577-83.
252. Chayen, N.E., *Methods for separating nucleation and growth in protein crystallisation*. Prog Biophys Mol Biol, 2005. **88**(3): p. 329-37.
253. Blow, D.M., et al., *Control of nucleation of protein crystals*. Protein Science : A Publication of the Protein Society, 1994. **3**(10): p. 1638-1643.
254. Patel, S., B. Cudney, and A. McPherson, *Polymeric Precipitants for the Crystallization of Macromolecules*. Biochemical and Biophysical Research Communications, 1995. **207**(2): p. 819-828.
255. McPherson, A., *A comparison of salts for the crystallization of macromolecules*. Protein Science : A Publication of the Protein Society, 2001. **10**(2): p. 418-422.
256. Juers, D.H. and B.W. Matthews, *Cryo-cooling in macromolecular crystallography: advantages, disadvantages and optimization*. Q Rev Biophys, 2004. **37**(2): p. 105-19.
257. Garman, E.F. and R.L. Owen, *Cryocooling and radiation damage in macromolecular crystallography*. Acta Crystallogr D Biol Crystallogr, 2006. **62**(Pt 1): p. 32-47.
258. Blow, D., *Outline of Crystallography for Biologists*. 2002: OUP Oxford.
259. Rupp, B., *Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology*. 2010: Garland Science.
260. Drenth, J., *Principles of Protein X-ray Crystallography*. 2002: Springer New York.
261. Skarzynski, T., *Collecting data in the home laboratory: evolution of X-ray sources, detectors and working practices*. Acta Crystallographica Section D: Biological Crystallography, 2013. **69**(Pt 7): p. 1283-1288.

262. Bruckner, A., et al., *Yeast two-hybrid, a powerful tool for systems biology*. Int J Mol Sci, 2009. **10**(6): p. 2763-88.
263. Stynen, B., et al., *Diversity in genetic in vivo methods for protein-protein interaction studies: from the yeast two-hybrid system to the mammalian split-luciferase system*. Microbiol Mol Biol Rev, 2012. **76**(2): p. 331-82.
264. Sandrock, B. and J.M. Egly, *A yeast four-hybrid system identifies Cdk-activating kinase as a regulator of the XPD helicase, a subunit of transcription factor IIH*. J Biol Chem, 2001. **276**(38): p. 35328-33.
265. Fields, S. and O. Song, *A novel genetic system to detect protein-protein interactions*. Nature, 1989. **340**(6230): p. 245-6.
266. Serebriiskii, I., V. Khazak, and E.A. Golemis, *A two-hybrid dual bait system to discriminate specificity of protein interactions*. J Biol Chem, 1999. **274**(24): p. 17080-7.
267. Aho, S., et al., *A novel reporter gene MEL1 for the yeast two-hybrid system*. Anal Biochem, 1997. **253**(2): p. 270-2.
268. Kamiya, T., et al., *Quantitative Y2H screening: cloning and signal peptide engineering of a fungal secretory LacA gene and its application to yeast two-hybrid system as a quantitative reporter*. J Biotechnol, 2010. **146**(4): p. 151-9.
269. Gyuris, J., et al., *Cdi1, a human G1 and S phase protein phosphatase that associates with Cdk2*. Cell, 1993. **75**(4): p. 791-803.
270. Le Douarin, B., et al., *A new version of the two-hybrid assay for detection of protein-protein interactions*. Nucleic Acids Res, 1995. **23**(5): p. 876-8.
271. Vojtek, A.B., S.M. Hollenberg, and J.A. Cooper, *Mammalian Ras interacts directly with the serine/threonine kinase Raf*. Cell, 1993. **74**(1): p. 205-14.
272. James, P., J. Halladay, and E.A. Craig, *Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast*. Genetics, 1996. **144**(4): p. 1425-36.
273. Cormack, R.S., K. Hahlbrock, and I.E. Somssich, *Isolation of putative plant transcriptional coactivators using a modified two-hybrid system incorporating a GFP reporter gene*. Plant J, 1998. **14**(6): p. 685-92.
274. Manning, V.A., et al., *Intracellular expression of a host-selective toxin, ToxA, in diverse plants phenocopies silencing of a ToxA-interacting protein, ToxABP1*. New Phytol, 2010. **187**(4): p. 1034-47.
275. Xu, C.W., A.R. Mendelsohn, and R. Brent, *Cells that register logical relationships among proteins*. Proc Natl Acad Sci U S A, 1997. **94**(23): p. 12473-8.
276. Johnsson, N. and A. Varshavsky, *Split ubiquitin as a sensor of protein interactions in vivo*. Proc Natl Acad Sci U S A, 1994. **91**(22): p. 10340-4.
277. Tarassov, K., et al., *An in vivo map of the yeast protein interactome*. Science, 2008. **320**(5882): p. 1465-70.
278. Stefan, E., et al., *Quantification of dynamic protein complexes using Renilla luciferase fragment complementation applied to protein kinase A activities in vivo*. Proc Natl Acad Sci U S A, 2007. **104**(43): p. 16916-21.
279. Atmakuri, K., Z. Ding, and P.J. Christie, *VirE2, a type IV secretion substrate, interacts with the VirD4 transfer protein at cell poles of Agrobacterium tumefaciens*. Mol Microbiol, 2003. **49**(6): p. 1699-713.
280. Lieberman, B.A., *Steroid Receptor Methods: Protocols and Assays*. 2010: Humana Press.
281. Chien, C.T., et al., *The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest*. Proc Natl Acad Sci U S A, 1991. **88**(21): p. 9578-82.
282. Rajagopala, S.V., *Mapping the Protein-Protein Interactome Networks Using Yeast Two-Hybrid Screens*. Adv Exp Med Biol, 2015. **883**: p. 187-214.
283. Zhang, J. and S. Lautar, *A yeast three-hybrid method to clone ternary protein complex components*. Anal Biochem, 1996. **242**(1): p. 68-72.

284. Licitra, E.J. and J.O. Liu, *A three-hybrid system for detecting small ligand-protein receptor interactions*. Proc Natl Acad Sci U S A, 1996. **93**(23): p. 12817-21.
285. SenGupta, D.J., et al., *A three-hybrid system to detect RNA-protein interactions in vivo*. Proc Natl Acad Sci U S A, 1996. **93**(16): p. 8496-501.
286. Li, J.J. and I. Herskowitz, *Isolation of ORC6, a component of the yeast origin recognition complex by a one-hybrid system*. Science, 1993. **262**(5141): p. 1870-4.
287. Stellberger, T., et al., *Improving the yeast two-hybrid system with permuted fusions proteins: the Varicella Zoster Virus interactome*. Proteome Science, 2010. **8**(1): p. 1-9.
288. Osborne, M.A., S. Dalton, and J.P. Kochan, *The Yeast Tribid System[mdash]Genetic Detection of trans-phosphorylated ITAM-SH2-Interactions*. Nat Biotech, 1995. **13**(12): p. 1474-1478.
289. Guo, D., et al., *A tethered catalysis, two-hybrid system to identify protein-protein interactions requiring post-translational modifications*. Nat Biotech, 2004. **22**(7): p. 888-892.
290. Estojak, J., R. Brent, and E.A. Golemis, *Correlation of two-hybrid affinity data with in vitro measurements*. Mol Cell Biol, 1995. **15**(10): p. 5820-9.
291. Yang, M., Z. Wu, and S. Fields, *Protein-peptide interactions analyzed with the yeast two-hybrid system*. Nucleic Acids Research, 1995. **23**(7): p. 1152-1156.
292. Christensen, J.J., et al., *Entropy Titration. A Calorimetric Method for the Determination of ΔG , ΔH , and ΔS from a Single Thermometric Titration*1a,b. The Journal of Physical Chemistry, 1966. **70**(6): p. 2003-2010.
293. Christensen, J.J., R.M. Izatt, and D. Eatough, *Thermodynamics of Metal Cyanide Coordination. V. Log K, ΔH° , and ΔS° Values for the Hg²⁺-CN⁻ System*. Inorganic Chemistry, 1965. **4**(9): p. 1278-1280.
294. Christensen, J.J., et al., *Theoretical evaluation of entropy titration method for calorimetric determination of equilibrium constants in aqueous solution*. Analytical Chemistry, 1968. **40**(11): p. 1713-1717.
295. Beaudette, N.V. and N. Langerman, *An improved method for obtaining thermal titration curves using micromolar quantities of protein*. Anal Biochem, 1978. **90**(2): p. 693-704.
296. Pierce, M.M., C.S. Raman, and B.T. Nall, *Isothermal Titration Calorimetry of Protein-Protein Interactions*. Methods, 1999. **19**(2): p. 213-221.
297. Liang, Y., *Applications of isothermal titration calorimetry in protein science*. Acta Biochim Biophys Sin (Shanghai), 2008. **40**(7): p. 565-76.
298. Freyer, M.W. and E.A. Lewis, *Isothermal Titration Calorimetry: Experimental Design, Data Analysis, and Probing Macromolecule/Ligand Binding and Kinetic Interactions*, in *Methods in Cell Biology*. 2008, Academic Press. p. 79-113.
299. Turnbull, W.B. and A.H. Daranas, *On the value of c: can low affinity systems be studied by isothermal titration calorimetry?* J Am Chem Soc, 2003. **125**(48): p. 14859-66.
300. Ericsson, U.B., et al., *Thermofluor-based high-throughput stability optimization of proteins for structural studies*. Anal Biochem, 2006. **357**(2): p. 289-98.
301. Parks, D.J., et al., *1,4-Benzodiazepine-2,5-diones as small molecule antagonists of the HDM2-p53 interaction: discovery and SAR*. Bioorg Med Chem Lett, 2005. **15**(3): p. 765-70.
302. Cummings, M.D., M.A. Farnum, and M.I. Nelen, *Universal screening methods and applications of ThermoFluor*. J Biomol Screen, 2006. **11**(7): p. 854-63.
303. Sambrook, J. and D.W. Russell, *The Condensed Protocols from Molecular Cloning: A Laboratory Manual*. 2006: Cold Spring Harbor Laboratory Press.
304. Zhang, T., et al., *An improved method for whole protein extraction from yeast Saccharomyces cerevisiae*. Yeast, 2011. **28**(11): p. 795-8.
305. Gasteiger, E., et al., *Protein Identification and Analysis Tools on the ExPASy Server*, in *The Proteomics Protocols Handbook*, J.M. Walker, Editor. 2005, Humana Press: Totowa, NJ. p. 571-607.
306. Lindwall, G., et al., *A sparse matrix approach to the solubilization of overexpressed proteins*. Protein Eng, 2000. **13**(1): p. 67-71.

307. Geer, L.Y., et al., *CDART: Protein Homology by Domain Architecture*. Genome Research, 2002. **12**(10): p. 1619-1623.
308. Linding, R., et al., *Protein disorder prediction: implications for structural proteomics*. Structure, 2003. **11**(11): p. 1453-9.
309. Clerici, M., et al., *The DUSP-Ubl domain of USP4 enhances its catalytic efficiency by promoting ubiquitin exchange*. Nat Commun, 2014. **5**: p. 5399.
310. Walker JR, A.G., Xue S, Butler-Cole C, Weigelt J, Bountra C, Arrowsmith CH, Edwards AM, Bochkarev A, Dhe-Paganon S, *Human ubiquitin specific peptidase 5 in covalent complex with ubiquitin*. SGC, 2009.
311. Avvakumov, G.V., et al., *Amino-terminal dimerization, NRDP1-rhodanese interaction, and inhibited catalytic domain conformation of the ubiquitin-specific protease 8 (USP8)*. J Biol Chem, 2006. **281**(49): p. 38061-70.
312. Luft, J.R., J.R. Wolfley, and E.H. Snell, *What's in a drop? Correlating observations and outcomes to guide macromolecular crystallization experiments*. Crystal growth & design, 2011. **11**(3): p. 651-663.
313. Ritorto, M.S., et al., *Screening of DUB activity and specificity by MALDI-TOF mass spectrometry*. Nat Commun, 2014. **5**: p. 4763.
314. San-Miguel, T., P. Pérez-Bermúdez, and I. Gavidia, *Production of soluble eukaryotic recombinant proteins in E. coli is favoured in early log-phase cultures induced at low temperature*. SpringerPlus, 2013. **2**: p. 89.
315. Arakawa, T., et al., *The critical role of mobile phase composition in size exclusion chromatography of protein pharmaceuticals*. J Pharm Sci, 2010. **99**(4): p. 1674-92.
316. Kramer, Ryan M., et al., *Toward a Molecular Understanding of Protein Solubility: Increased Negative Surface Charge Correlates with Increased Solubility*. Biophysical Journal, 2012. **102**(8): p. 1907-1915.
317. Seddon, A.M., P. Curnow, and P.J. Booth, *Membrane proteins, lipids and detergents: not just a soap opera*. Biochim Biophys Acta, 2004. **1666**(1-2): p. 105-17.
318. Gupta, N., et al., *Analyzing protease specificity and detecting in vivo proteolytic events using tandem mass spectrometry*. Proteomics, 2010. **10**(15): p. 2833-2844.
319. Keil, B., *Specificity of Proteolysis*. 1992: Springer-Verlag.
320. Reinhard, L., et al., *Optimization of protein buffer cocktails using Thermofluor*. Acta Crystallogr Sect F Struct Biol Cryst Commun, 2013. **69**(Pt 2): p. 209-14.
321. Smyth, D.R., et al., *Crystal structures of fusion proteins with large-affinity tags*. Protein Sci, 2003. **12**(7): p. 1313-22.
322. Wright, P.E. and H.J. Dyson, *Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm*. J Mol Biol, 1999. **293**(2): p. 321-31.
323. Chan, P., R.A. Curtis, and J. Warwicker, *Soluble expression of proteins correlates with a lack of positively-charged surface*. Sci Rep, 2013. **3**: p. 3333.
324. Majhi, P.R., et al., *Electrostatically driven protein aggregation: beta-lactoglobulin at low ionic strength*. Langmuir, 2006. **22**(22): p. 9150-9.
325. Fink, A.L., *Protein aggregation: folding aggregates, inclusion bodies and amyloid*. Fold Des, 1998. **3**(1): p. R9-23.
326. Vihinen, M., E. Torkkila, and P. Riikonen, *Accuracy of protein flexibility predictions*. Proteins, 1994. **19**(2): p. 141-9.
327. Banuelos, S., M. Saraste, and K. Djinovic Carugo, *Structural comparisons of calponin homology domains: implications for actin binding*. Structure, 1998. **6**(11): p. 1419-31.
328. Fraser, J.S., et al., *An atypical receiver domain controls the dynamic polar localization of the Myxococcus xanthus social motility protein FrzS*. Mol Microbiol, 2007. **65**(2): p. 319-32.
329. Djinovic Carugo, K., S. Banuelos, and M. Saraste, *Crystal structure of a calponin homology domain*. Nat Struct Biol, 1997. **4**(3): p. 175-9.

330. Sharp, P.M., et al., *Codon usage patterns in Escherichia coli, Bacillus subtilis, Saccharomyces cerevisiae, Schizosaccharomyces pombe, Drosophila melanogaster and Homo sapiens; a review of the considerable within-species diversity*. Nucleic Acids Res, 1988. **16**(17): p. 8207-11.
331. Kane, J.F., *Effects of rare codon clusters on high-level expression of heterologous proteins in Escherichia coli*. Curr Opin Biotechnol, 1995. **6**(5): p. 494-500.
332. Gouy, M. and C. Gautier, *Codon usage in bacteria: correlation with gene expressivity*. Nucleic Acids Res, 1982. **10**(22): p. 7055-74.
333. Baneyx, F., *Recombinant protein expression in Escherichia coli*. Curr Opin Biotechnol, 1999. **10**(5): p. 411-21.
334. Dong, X., et al., *PlasMapper: a web server for drawing and auto-annotating plasmid maps*. Nucleic Acids Research, 2004. **32**(Web Server issue): p. W660-W664.
335. An, X., et al., *Identification and functional characterization of protein 4.1R and actin-binding sites in erythrocyte beta spectrin: regulation of the interactions by phosphatidylinositol-4,5-bisphosphate*. Biochemistry, 2005. **44**(31): p. 10681-8.
336. Berleman, J.E., et al., *FrzS Regulates Social Motility in Myxococcus xanthus by Controlling Exopolysaccharide Production*. PLoS ONE, 2011. **6**(8): p. e23920.
337. Zhan, Y., X. Song, and G.W. Zhou, *Structural analysis of regulatory protein domains using GST-fusion proteins*. Gene, 2001. **281**(1-2): p. 1-9.
338. Waugh, D.S., *Crystal structures of MBP fusion proteins*. Protein Science : A Publication of the Protein Society, 2016. **25**(3): p. 559-571.
339. Stellberger, T., et al., *Improving the yeast two-hybrid system with permuted fusions proteins: the Varicella Zoster Virus interactome*. Proteome Science, 2010. **8**: p. 8-8.
340. Huang, H. and J.S. Bader, *Precision and recall estimates for two-hybrid screens*. Bioinformatics, 2009. **25**(3): p. 372-8.
341. Ma, J. and M. Ptashne, *A new class of yeast transcriptional activators*. Cell, 1987. **51**(1): p. 113-9.
342. Giniger, E. and M. Ptashne, *Transcription in yeast activated by a putative amphipathic alpha helix linked to a DNA binding unit*. Nature, 1987. **330**(6149): p. 670-2.
343. Gill, G., I. Sadowski, and M. Ptashne, *Mutations that increase the activity of a transcriptional activator in yeast and mammalian cells*. Proc Natl Acad Sci U S A, 1990. **87**(6): p. 2127-31.
344. Shenoy, S.K., et al., *Beta-arrestin-dependent signaling and trafficking of 7-transmembrane receptors is reciprocally regulated by the deubiquitinase USP33 and the E3 ligase Mdm2*. Proc Natl Acad Sci U S A, 2009. **106**(16): p. 6650-5.
345. Jean-Charles, P.-Y., et al., *Ubiquitin-specific protease 20 Regulates the Reciprocal Functions of Beta-arrestin2 in Toll-like Receptor 4-promoted NFkappaB Activation*. Journal of Biological Chemistry, 2016.
346. Lowery, D.M., D. Lim, and M.B. Yaffe, *Structure and function of Polo-like kinases*. Oncogene, 2005. **24**(2): p. 248-59.
347. Amanchy, R., et al., *A curated compendium of phosphorylation motifs*. Nat Biotechnol, 2007. **25**(3): p. 285-6.
348. Zhou, H., et al., *Toward a comprehensive characterization of a human cancer cell phosphoproteome*. J Proteome Res, 2013. **12**(1): p. 260-71.
349. Mulhern, D., *CST Curation Set: 12684; Year: 2011; Biosample/Treatment: cell line, Jurkat/calyculin & '||' pervanadate; Disease: T cell leukemia; SILAC: -; Specificities of Antibodies Used to Purify Peptides prior to LCMS: (F/Y/M)Xp[ST](L/I/M)*. 2011.
350. Sharma, K., et al., *Ultradeep human phosphoproteome reveals a distinct regulatory nature of Tyr and Ser/Thr-based signaling*. Cell Rep, 2014. **8**(5): p. 1583-94.
351. Schweppe, D.K., J.R. Rigas, and S.A. Gerber, *Quantitative phosphoproteomic profiling of human non-small cell lung cancer tumors*. J Proteomics, 2013. **91**: p. 286-96.

352. Mertins, P., et al., *Integrated proteomic analysis of post-translational modifications by serial enrichment*. Nat Methods, 2013. **10**(7): p. 634-7.
353. Guo, A., *CST Curation Set: 11984; Year: 2011; Biosample/Treatment: cell line, Jurkat/calyculin &'||' pervanadate; Disease: T cell leukemia; SILAC: -; Specificities of Antibodies Used to Purify Peptides prior to LCMS: p[ST]XP* 2011.
354. Guo, A., *CST Curation Set: 11989; Year: 2011; Biosample/Treatment: cell line, Jurkat/calyculin &'||' pervanadate; Disease: T cell leukemia; SILAC: -; Specificities of Antibodies Used to Purify Peptides prior to LCMS: p[ST]XP Antibodies Used to Purify Peptides prior to LCMS: Phospho-(Ser) 14-3-3 Binding Motif Antibody Cat#: 9601* 2011.
355. Possemato, A., *CST Curation Set: 6922; Year: 2009; Biosample/Treatment: cell line, Jurkat/pervanadate &'||' calyculin; Disease: T cell leukemia; SILAC: -; Specificities of Antibodies Used to Purify Peptides prior to LCMS: p[ST]P* 2009.
356. Dephoure, N., et al., *A quantitative atlas of mitotic phosphorylation*. Proc Natl Acad Sci U S A, 2008. **105**(31): p. 10762-7.
357. Guo, A., *CST Curation Set: 12740; Year: 2011; Biosample/Treatment: cell line, Jurkat/calyculin &'||' pervanadate; Disease: T cell leukemia; SILAC: -; Specificities of Antibodies Used to Purify Peptides prior to LCMS: PXpSP, pSPX(K/R) Antibodies Used to Purify Peptides prior to LCMS: Phospho-MAPK/CDK Substrates (PXSP or SPXR/K) (34B2) Rabbit mAb Cat#: 2325, PTMScan(R) Phospho-MAPK/CDK Substrate Motif (PXS*P, S*PXK/R) Immunoaffinity Beads Cat#: 1982* 2011.
358. Daub, H., et al., *Kinase-selective enrichment enables quantitative phosphoproteomics of the kinome across the cell cycle*. Mol Cell, 2008. **31**(3): p. 438-48.
359. Christensen, G.L., et al., *Quantitative phosphoproteomics dissection of seven-transmembrane receptor signaling using full and biased agonists*. Mol Cell Proteomics, 2010. **9**(7): p. 1540-53.
360. Chen, R.Q., et al., *CDC25B mediates rapamycin-induced oncogenic responses in cancer cells*. Cancer Res, 2009. **69**(6): p. 2663-8.
361. Stokes, M., *CST Curation Set: 4392; Year: 2008; Biosample/Treatment: cell line, K562/untreated; Disease: chronic myelogenous leukemia; SILAC: -; Specificities of Antibodies Used to Purify Peptides prior to LCMS: p[STY])* 2008.
362. Kolker, E., et al., *MOPED: Model Organism Protein Expression Database*. Nucleic Acids Research, 2012. **40**(Database issue): p. D1093-D1099.
363. Wang, M., et al., *PaxDb, a Database of Protein Abundance Averages Across All Three Domains of Life*. Molecular & Cellular Proteomics : MCP, 2012. **11**(8): p. 492-500.
364. Schaab, C., et al., *Analysis of high accuracy, quantitative proteomics data in the MaxQB database*. Mol Cell Proteomics, 2012. **11**(3): p. M111.014068.
365. Trinkle-Mulcahy, L., et al., *Identifying specific protein interaction partners using quantitative mass spectrometry and bead proteomes*. J Cell Biol, 2008. **183**(2): p. 223-39.
366. Huang da, W., B.T. Sherman, and R.A. Lempicki, *Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources*. Nat Protoc, 2009. **4**(1): p. 44-57.
367. Huang da, W., B.T. Sherman, and R.A. Lempicki, *Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists*. Nucleic Acids Res, 2009. **37**(1): p. 1-13.
368. Shanmugam, I., et al., *Ubiquitin-specific Peptidase 20 Regulates Rad17 Stability, Checkpoint Kinase 1 Phosphorylation and DNA Repair by Homologous Recombination*. The Journal of Biological Chemistry, 2014. **289**(33): p. 22739-22748.
369. Boyault, C., et al., *HDAC6-p97/VCP controlled polyubiquitin chain turnover*. The EMBO Journal, 2006. **25**(14): p. 3357-3366.
370. Corsini, L., et al., *Thioredoxin as a fusion tag for carrier-driven crystallization*. Protein Sci, 2008. **17**(12): p. 2070-9.

371. Berndt, C., C.H. Lillig, and A. Holmgren, *Thioredoxins and glutaredoxins as facilitators of protein folding*. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 2008. **1783**(4): p. 641-650.
372. Matsumoto, S., et al., *Crystal structure of the C-terminal globular domain of oligosaccharyltransferase from Archaeoglobus fulgidus at 1.75 Å resolution*. *Biochemistry*, 2012. **51**.
373. Slabinski, L., et al., *XtalPred: a web server for prediction of protein crystallizability*. *Bioinformatics*, 2007. **23**(24): p. 3403-5.
374. Yuan, J., et al., *HERC2-USP20 axis regulates DNA damage checkpoint through Claspin*. *Nucleic Acids Research*, 2014. **42**(21): p. 13110-13121.
375. Deller, M.C., L. Kong, and B. Rupp, *Protein stability: a crystallographer's perspective*. *Acta Crystallographica. Section F, Structural Biology Communications*, 2016. **72**(Pt 2): p. 72-95.