Investigations into the biophysical properties of the proline rich region of Ubiquilin-2

Bethany Jayne Waddington, MBiomed

Submitted in fulfilment of the requirements for the Degree of Doctor of Philosophy

Institute of Molecular, Cell, and Systems Biology

College of Medical, Veterinary, and Life Sciences

University of Glasgow

May 2022

Abstract

Ubiquitin is a 76 amino acid protein used by eukaryotic cells to selectively label and target proteins to a range of cellular fates, including degradation. Ubiquitin can be attached in many ways and the length of ubiquitin modification creates a code for the outcome of the tagged protein. Furthermore, the type of ubiquitin chain conjugated also directs tagged proteins to a particular fate. This ubiquitin modification code is interpreted by ubiquitin receptors. These receptors contain ubiquitin binding domains which interact with specific chains to direct tagged proteins to the correct pathway. The family of Ubquilin proteins contain such binding domains (termed Ubiquitin Associated domains and shortened to UBA domains from here on). Current literature has shown Ubiquilins target K48 tagged proteins to the proteasome for degradation. Intriguingly however, when expressing the UBA domain in isolation, ubiquitin is bound non-discriminately and with poor affinity. This raises the question of how the Ubiquilins can function as Ubiquitin receptors with such poor affinity and specificity?

Unique to UBQLN2, the proline rich repeat region (PRR region) is hypothesised in this study to be one of the mechanisms conferring the necessary specificity for UBQLN2 to function within the cell. The PRR region houses the majority of mutations causing a rare form of juvenile, X-linked ALS. However, the structure and function of this region remains elusive. Computational modelling conducted in this study has revealed a similarity to bacterial collagen-like proteins, some of which trimerize via their collagen-like domains. The PRR region of UBQLN2 was investigated using a variety of recombinantly expressed proteins and biophysical techniques.

Initial investigation revealed a trimeric structure forming via the PRR region. A previously undescribed secondary structure for the PRR region was also revealed: the polyproline-II helix. This helix is a common precursor for collagen triple helix formation. In order to investigate the potential of this region to multimerize as a collagen-like triple helix, an obligate trimer was engineered. When modelled in *AlphaFold*, this protein revealed a propensity to form a triple helical structure. Upon closer inspection using NMR, no clear conclusions could be drawn regarding its structure, due to the repeated, proline-rich nature of the region. However,

dihedral angle predictions place over 50% of the residues of the PRR region in the Ramachandran space often occupied by residues of the collagen triple helix. Engineering these domains in such a way did not alter the folding of the globular UBA domain, which retained its ability to bind ubiquitin. Therefore, a novel obligate trimer has been created to model the PRR region and UBA domain of UBQLN2. The technologies developed in this study offer new opportunities to elucidate the structure of the PRR region and infer the role it plays in disease progression.

Table of contents

List of tables

List of figures

Acknowledgements

I would like to start by thanking my supervisor Dr Thimo Kurz. Your support and guidance through this changing journey were vital to my completion of this PhD and I will be forever grateful for the opportunity to work in your lab. I would also like to express my deepest appreciation to Dr Brian Smith, who generously provided resources, knowledge, expertise, and (most precious), his time. This endeavour would not have been possible without your support, and I will always credit you with helping me rediscover my love for science. Additionally, the support received from Dr George Gorrie was extremely appreciated, especially during the more turbulent times of my study.

I could not have undertaken this exciting journey without the support of the Euan MacDonald Centre for Motor Neuron Disease Research, a truly special charity responsible for funding this project and providing a welcoming community in which to begin my career.

This PhD project could not have been accomplished without the support and generosity of numerous colleagues, and to mention all by name would be too long. However, I would like to extend a special thanks to Dr Jamie Williams and the Tokatlidis group, in particular Ruairidh and Erik, for all your guidance with the numerous new techniques I encountered.

To my friends and family who have constantly supported me, especially my parents who have always worked hard to provide me with the opportunities to get this far and held unwavering confidence in me throughout the process. And to my friends Fred, Harriet, and Hope. The laughter and love you provided outside of the lab were always what I needed, especially during the tougher times.

And finally, to Ross. I honestly don’t think I would have stayed sane without you. Thank you for showing me I am so much more than this PhD project.

Author's Declaration

I declare that, except where explicit reference is made to the contribution of others, this thesis is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.

Bethany Waddington

Abbreviations

µ	Micro
^{13}C	Carbon-13 Isotope
^{15}N	Nitrogen-15 Isotope
^{1}H	Proton
3-Hyp	3S-Hydroxyproline
4-Hyp	4R-Hydroxyproline
Å	Angstrom
ALS	Amyotrophic Lateral Sclerosis
AUC	Analytical Ultracentrifugation
BLAST	Basic Local Alignment Search Tool
BMRB	Biological Magnetic Resonance Bank
bp	Base Pair
CD	Circular Dichroism
cm	Centimetre
CSP	Chemical Shift Perturbation
DANGLE	Dihedral Angles from Global Likelihood Measurements
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
E. coli	*Escherichia coli*
ER	Endoplasmic Reticulum
FPLC	Fast Protein Liquid Chromatography
FPRR	Foldon-PRR-UBA
g	Gram
g	Relative Centrifugal Force (G-force)
Gly/G	Glycine
gRNA	Guide RNA
HCPRR	His_6-CC-pII-PRR-UBA
HFPRR	His_6-Foldon-PRR-UBA
His/H	Histidine
HSQC	Heteronuclear Single Quantum Coherence

Hz	Hertz
iPSC	Induced Pluripotent Stem Cell
IPTG	Isopropylthio-ß-galactoside
K	Kelvin
K48-	Lys-48 linked
K63-	Lys-63 linked
K_d	Dissociation Constant
kDa	Kilodalton
kHz	Kilohertz
L	Litre
LB	Luria Broth
Leu/L	Leucine
LLPS	Liquid-liquid Phase separation
LMF	Laminin, Matrigel, Fibronectin Coating
m	Milli
M	Molar
mHz	Megahertz
ml	Millilitre
mm	Millimetre
MN	Motor Neuron
MND	Motor Neuron Disease
mRNA	Messenger RNA
mV	Millivolts
MW	Molecular Weight
MWCO	Molecular Weight Cut Off
n	Nano
nm	Nanometre
NMR	Nuclear Magnetic Resonance
PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
polyproline-II	Poly-L-proline type II helix

PPII	Poly-L-proline type II helix
ppm	Parts per Million
PRALINE	Profile Alignment Toolbox
Pro/P	Proline
PRR Region	Proline Rich Repeat Region
RNA	Ribonucleic Acid
rpm	Revolutions per Minute
s	Seconds
S	Sedimentation Coefficient
$S_{20,w}$	Sedimentation coefficient normalised to 20°C in water
SDS	Sodium Dodecyl Sulphate
SE	Sedimentation Equilibrium
SEC	Size Exclusion Chromatography
Ser/S	Serine
SPR	Surface Plasmon Resonance
STI1	Stress-induced protein 1-like
SV	Sedimentation Velocity
TEV Protease	Tobacco Etch Virus Protease
Thr/T	Threonine
TROSY	Transverse Relaxation Optimized Spectroscopy
Ub	Monoubiquitin
Ub_2	Diubiquitin
UBA	Ubiquitin Associated Domain
UBD	Ubiquitin Binding Domain
UBL	Ubiquitin Like Domain
UBQLN1	Ubiquilin 1
UBQLN2	Ubiquilin 2
UV	Ultraviolet
V	Volts
WT	Wild Type
Xaa	Any amino acid
Yaa	Any amino acid

α	Alpha
ß	Beta
ϕ	Phi
ψ	Psi

1 Introduction

1.1 ALS

Motor Neuron Disease (MND) is a fatal neurodegenerative disease characterised by the selective loss of motor neurons (MNs). Amyotrophic Lateral Sclerosis (ALS) is one of four variants of MND and is the most common form, at an incidence of 2 in 100,000 (Chiò *et al.*, 2013). Through selective loss of upper and lower MNs, symptoms such as motor impairments, muscle weakness, spasticity, and dysphagia present (Bruijn, Miller and Cleveland, 2004). Approximately 90% of ALS cases are sporadic (sALS), with no apparent family history of the disease (Renaud *et al.*, 2019). The remaining 5-10% of cases are familial (fALS), and several genes have been implicated in fALS including C9ORF72 (Chromosome 9 open reading frame 72), SOD1 (Super- oxide dismutase 1), TDP-43 (TAR DNA-binding protein 43) and FUS (Fused in Sarcoma) (Rosen *et al.*, 1993; Kabashi *et al.*, 2008; Vance *et al.*, 2009; DeJesus-Hernandez *et al.*, 2011). Although some patients may live many years with the disease, most will have a significantly shortened lifespan of around 3-5 years from diagnosis, with onset occurring around 60 years of age (Zinman and Cudkowicz, 2011). With no cure and only one approved treatment in the UK, it is paramount that more is done to further the research into, and to discover therapeutics for treating, ALS.

A major characteristic of ALS and other forms of neurodegenerative diseases is the presence of protein aggregates in the affected cells (Al-Chalabi *et al.*, 2012; Blokhuis *et al.*, 2013). A number of the aggregated proteins common in ALS are often associated with the cellular homeostatic pathways. For instance, Sequestosome-1 (commonly referred to as p62) and Optineurin (OPTN) are both cargo receptors for selective autophagy pathways (Bjørkøy *et al.*, 2005; Shen *et al.*, 2011), and TDP-43 (TAR DNA-binding protein 43) is a versatile DNA/RNA binding protein involved in RNA-related metabolism (Bhardwaj *et al.*, 2013). This inability to efficiently clear mutated, damaged, or aggregated proteins points to a dysfunction in the homeostatic proteolytic systems within the MNs. Ubiquilin 2 (UBQLN2) is a protein with a variety of cellular roles, including within proteolytic pathways (N'Diaye *et al.*, 2009; Rothenberg *et al.*, 2010; McKinnon and Tabrizi,

2014; Hjerpe *et al.*, 2016; Wu *et al.*, 2020). Mutations in UBQLN2 have been established as a cause of X-linked juvenile- and adult-onset ALS (Deng *et al.*, 2011). Of note is the presence of UBQLN2 positive aggregates in ALS patients with no genetic mutation in UBQLN2 (Deng *et al.*, 2011; Williams *et al.*, 2012). This reaffirms the belief that UQBLN2 activity is central to cellular homeostasis and protein degradation within MNs. The complexity of the formation and clearance of aggregates in ALS is yet to be understood. However, much evidence points to an impairment in protein degradation pathways, as human spinal cord autopsy of ALS patients with a mutation in UBQLN2 found skein-like inclusions containing UBQLN2, Ubiquitin, p62, TDP-43, Fus and OPTN (Deng *et al.*, 2011; Williams *et al.*, 2012). Together, this data provides the justification that further research is needed into the mechanistic actions of UBQLN2, and its impact in the development of ALS.

1.2 The UBQLNs

Originally identified in yeast (as Dsk2) as an essential factor for duplication of the microtubule organising centre (Biggins, Ivanovska and Rose, 1996), there have since been five Ubiquilin proteins discovered in humans, named UBQLN 1-4 and UBQLN L (Marín, 2014). Whilst UBQLN1, UBQLN2 and UBQLN4 are expressed throughout the body, UBQLN3 and UBQLN-L are selectively expressed in the testes, with a potential role in post-meiotic spermatids (Conklin *et al.*, 2000; Yuan *et al.*, 2015). Notably, the three ubiquitously expressed UBQLNs each have an association with the development of disease: UBQLN1 variants may increase the risk of Alzheimer's disease (Bertram *et al.*, 2005), UBQLN4 mutations lead to a hereditary cancer syndrome (Tang *et al.*, 2021), and mutations in UBQLN2 causes hereditary ALS (Deng *et al.*, 2011).

The intronless gene encoding UBQLN2 is located on chromosome Xp11.21 (Kaye and Shows, 2000). The resulting 66 kDa protein has been actively studied in the context of the Ubiquitin Proteasome System (UPS)(McKinnon and Tabrizi, 2014; Hjerpe *et al.*, 2016), but there are also reports implicating UBQLN2 in other cellular pathways such as autophagy function (N'Diaye *et al.*, 2009; Rothenberg *et al.*, 2010; Wu *et al.*, 2020), neuroinflammation (Picher-Martel *et al.*, 2015),

formation of stress granules (Dao *et al.*, 2018), and the mislocalisation of TDP-43 into insoluble cytosolic aggregates (Le *et al.*, 2016).

1.2.1 UBLQN2 Structure

UBQLN2 is composed of three key domains (Figure 1.1). The Ubiquitin Like (UBL) domain is capable of binding to the proteasome due to its structural, but not sequence, similarity to the protein *ubiquitin* (Kang, Zhang, *et al.*, 2007). The disordered central domain contains four stress-induced protein 1-like (STI-1) domains, which have been reported to have roles in chaperone binding (Hjerpe *et al.*, 2016) and oligomerization (Dao *et al.*, 2018). Finally, the Ubiquitin Associated (UBA) domain is a highly conserved domain between the UBQLNs, necessary for the binding of ubiquitinated substrates for shuttling to the proteasome (Walters *et al.*, 2002).

Within the central domain lies the unique P-X-X motif (referred to as PRR region in this study), a tandem of 12 Pro-Xaa-Xaa repeats (where Xaa is any amino acid). The PRR region is unique to UBQLN2, differentiating it from other UBQLNs. The high proline and glycine content with mostly hydrophobic residues in-between mimics that of collagen (Ramachandran and Kartha, 1954). Similarly, this enriched Pro and Gly region bears similarities to elastin and elastin-like polypeptides (ELPs), derived from the hydrophobic region of tropoelastin (Roberts, Dzuricky and Chilkoti, 2015). These ELPs are self-interacting biopolymers which undergo temperature-based phase transition (Quiroz *et al.*, 2019). Together, these are numerous examples of how the PRR region may behave structurally and functionally. However, investigations to date have determined the PRR region to be largely disordered (Dao *et al.*, 2018). The PRR region has been implicated in protein-protein interactions (Aitio *et al.*, 2010; Deng *et al.*, 2011) and houses nearly all the ALS disease-causing UBQLN2 mutations. Despite the obvious impact mutations have on the development of disease, the function of the PRR region remains elusive (Deng *et al.*, 2011).

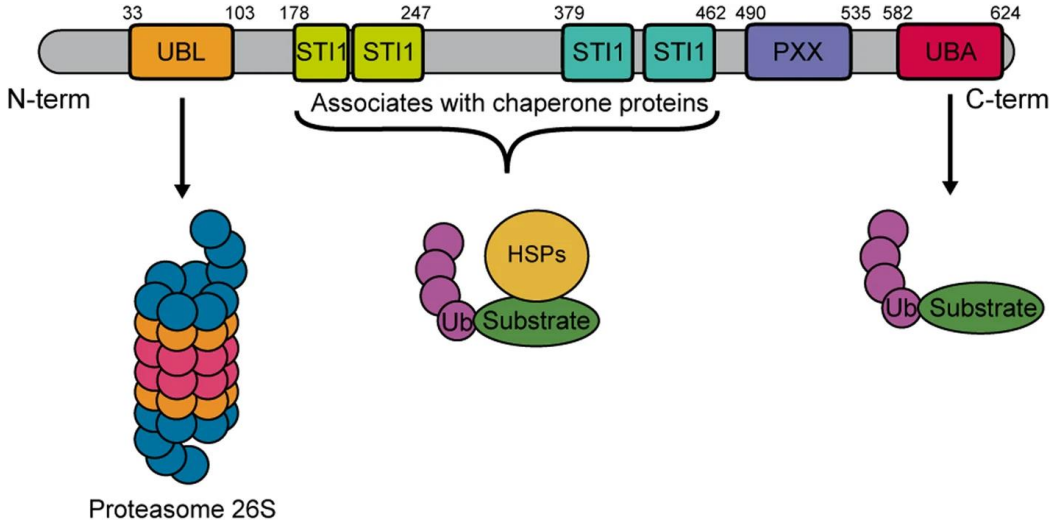


Figure 1.1 Domain architecture of UBQLN2. Schematic outlining the key domains and regions in UBQLN2 along with their associated function. Figure taken and adapted from (Renaud *et al.*, 2019).

1.2.2 UBLQN2 in the development of disease

Mutations in UBQLN2 were first identified in five unrelated families suffering with ALS/FTD (Deng *et al.*, 2011). Since this initial identification, many more mutations have been identified in ALS patients (Table 1.1). However, whilst some papers and reviews would classify these as disease-causing mutations, caution should be exerted when including some of these in the established mutations of UBQLN2. For instance, whilst the mutations were identified in diagnosed ALS cases, there were often reports of the existence of these mutations in unaffected relatives, as in the case of UBQLN2^{S400G} and UBQLN2^{P440L} (Dillen *et al.*, 2013), or of different UBQLN2 mutations identified in the control cohort (Millecamps *et al.*, 2012). In other cases, the patient also carried previously established mutations in OPTN (UBQLN2^{P533L}) or TARDBP (UBQLN2^{M446R})(Gellera *et al.*, 2013). This therefore casts doubt over the conclusive contribution of these particular UBQLN2 mutations to the development of disease. The mutations verified by *in vitro* and *in vivo* analysis or inheritance patterns analysis are more concrete in their contributions to the development of ALS (P494L, P506A, T487I, P497H, P506T, P509S, P525S)(Deng *et al.*, 2011; Williams *et al.*, 2012; Fahed *et al.*, 2014; Teyssou *et al.*, 2017; Gkazi *et al.*, 2019). Analysis of published UBQLN2 mutations revealed mutations in the proline-rich regions cause significant earlier onset in male patients (Gkazi *et al.*, 2019). The mutations for which no biological investigation was carried out, establishment of pathological contribution was predicted through *in silico*

analysis. Various algorithms were used, often with conflicting predictions for the same mutation within the same study (Gellera *et al.*, 2013). The contribution of these mutations to the development of disease cannot be definitively proven only with bioinformatic analysis. This is especially true for UBQLN2, where no structure has been determined and various computational modelling predicts disorder for the PRR region (Hjerpe *et al.*, 2016; Dao *et al.*, 2018).

Table 1.1 ALS-causing mutations in UBQLN2. Identified ALS-causing mutations in UBQLN2. Bioinformatic analyses used are indicated within the table. Mutations with the strongest evidence of pathogenesis are highlighted in blue within the table.

Mutation	Phenotype	Bioinformatic predictions	*In vitro* or *in vivo* studies	Reference
S400G	sALS	Neutral		(Dillen *et al.*, 2013)
P440L	sALS	Pathological by PMUT, tolerated by SIFT		(Dillen *et al.*, 2013)
M466R	fALS	Pathological by PMUT, non-neutral by SNAP		(Gellera *et al.*, 2013)
T487I	fALS		Colocalization with ubiquitin, TDP-43 and FUS in post-mortem spinal cord tissue.	(Williams *et al.*, 2012)
A488T	sALS	Deleterious by SIFT		(Teyssou *et al.*, 2017)
P497L	fALS/FTD/ALS	Deleterious by SIFT	Expression impairs autophagy and Hsp70 binding in the lymphoblast. Gliosis in corticospinal tracts.	(Fahed *et al.*, 2014; Teyssou *et al.*, 2017)
P497H	ALS	Pathological by PMUT, non-neutral by SNAP	Alters ubiquitin binding in cells, impairs the UPS, exacerbates TDP-43 pathology in rats, hyper ubiquitinated in cells, reduces FUS interaction, impairs stress granule formation, disrupts ERAD.	(Deng *et al.*, 2011; Gellera *et al.*, 2013; Xia *et al.*, 2014; Picher-Martel *et al.*, 2015; Alexander *et al.*, 2018; Kim and Goldberg, 2018)
P497S	fALS	Deleterious by SIFT	Impairs autophagy and Hsp70 binding in lymphoblast, and reduces autophagosome acidification in mice.	(Deng *et al.*, 2011; Teyssou *et al.*, 2017; Wu *et al.*, 2020)
P500S	fALS	Deleterious by SIFT		(Teyssou *et al.*, 2017)

Mutation	Phenotype	Bioinformatic predictions	*In vitro* or *in vivo* studies	Reference
P506A	fALS	Deleterious by SIFT	Impairs autophagy and Hsp70 binding in lymphoblast.	(Teyssou *et al.*, 2017)
P506S	ALS	Pathological by PMUT, neutral by SNAP, not tolerated by SIFT, benign by PolyPhen	Localization to neuronal cytoplasmic inclusions in post-mortem tissue.	(Gellera *et al.*, 2013; Vengoechea *et al.*, 2013; Gkazi *et al.*, 2019)
P506T	ALS		Increases ubiquitination, reduces Hsp70 interaction, impairs UPS, reduces FUS interaction, impairs stress granule formation, increases UBQLN2 aggregation propensity, alters droplet morphology, cognitive deficits in mice, increased neuronal death in mice.	(Deng *et al.*, 2011; Hjerpe *et al.*, 2016; Alexander *et al.*, 2018; Kim and Goldberg, 2018; Sharkey *et al.*, 2018)
P509S	ALS		Similar ubiquitination level and solubility compared to WT.	(Deng *et al.*, 2011; Kim and Goldberg, 2018)
P525S	ALS		Slight increase in ubiquitination level and solubility, increase neuronal toxicity.	(Deng *et al.*, 2011; Kim and Goldberg, 2018)
P533L	fALS	Pathological by PMUT and SNAP		(Gellera *et al.*, 2013)
V538L	sALS	Neutral by PMUT and SNAP	Located after PXX domain, does not affect phase separation.	(Dao *et al.*, 2018)

As stated earlier, UBQLN2 has been implicated in numerous cellular pathways. It translocates polyubiquitinated substrates to the proteasome for degradation (McKinnon and Tabrizi, 2014; Hjerpe *et al.*, 2016), plays a role in the ER-associated degradation (ERAD) pathways (Xia *et al.*, 2014), and can maintain mitochondrial function in stress conditions by indirectly inducing Ca^{2+} homeostasis (Kim and

Goldberg, 2018). Finally, it has a role in autophagy, where it indirectly promotes autophagosome formation and lysosomal fusion (N'Diaye *et al.*, 2009; Rothenberg *et al.*, 2010). The actions of UBQLN2 are clearly widespread within the cell, leading to a possible notion that a loss of UBQLN2 would be detrimental. Interestingly, the opposite occurs in both *in vitro* and *in vivo* studies (Picher-Martel *et al.*, 2015; Huang, Shen and Fan, 2017). Overexpression of either wild-type or mutated UBQLN2 caused neuronal death, whilst deletion of UBQLN2 did not. This may suggest a toxic gain of function in mutated UBQLN2, or potentially a loss of function blocks cellular systems in a way which prevents compensatory or complementary mechanisms acting. Another interpretation of this result may be due to the overexpression of UBQLN2 impacting cell viability in a mechanism distinct to that of the disease-causing mutant. A common phenotype when overexpressing UBQLNs is cell death, likely due to sequestration of ubiquitin, as overexpression of UBL-UBA proteins can cause proteasomal degradation inhibition (Kleijnen et al., 2000; Chen and Madura, 2002; Raasi and Pickart, 2003), even at small increases (Verma *et al.*, 2004). Therefore, studies investigating the contribution of UBQLN2 mutants to disease states should be cautious when employing overexpression techniques. Together, this highlights the fine balance required for cellular homeostasis and the implications one protein can have in neuronal survival.

1.3 Protein Quality Control

Protein turnover is the net result of protein synthesis and protein degradation. It allows for the removal and replacement of damaged or superfluous proteins with new ones. Protein turnover is a constant phenomenon in cells, occurring continuously even under steady-state conditions (Price *et al.*, 2010; Cambridge *et al.*, 2011; Boisvert *et al.*, 2012; Cohen *et al.*, 2013). Protein turnover rates vary between protein function, location within the cell, and cell type (Price *et al.*, 2010; Dörrbaum *et al.*, 2018; Mathieson *et al.*, 2018) and are also influenced by both the extracellular and intracellular environment (Dörrbaum *et al.*, 2018). These differences in turnover rates likely correlate to an evolutionary balance between energy-saving stability and dynamic plasticity within the cell. For example, proteome remodelling is required in the brain for learning and memory function (Sutton and Schuman, 2006), and several forms of synaptic plasticity

studied *in vitro* required both protein synthesis and degradation (Kang and Schuman, 1996; Ehlers, 2003; Rosenberg *et al.*, 2014; Schanzenbächer *et al.*, 2016). Dysregulation of protein homeostasis has been associated with a variety of diseases, such as Crohn's disease (Fujimoto *et al.*, 2016), some forms of cancer (van Drie, 2011), and a number of neurodegenerative diseases, including ALS (Douglas and Dillin, 2010; Deng et al., 2011; Taylor, Brown and Cleveland, 2016; Renaud et al., 2019).

Protein degradation is crucial in post-mitotic cells such as neurons; without proliferation the misfolded proteins or aggregations will not be diluted out, resulting in toxicity within the cell (Hara *et al.*, 2006; Komatsu *et al.*, 2006). Therefore, to ensure efficient progression through the system, numerous proteins and complexes for protein quality control (PQC) are formed and utilised. Two major systems have evolved in eukaryotes to achieve degradation: the Autophagy-Lysosome Pathway (ALP) and the Ubiquitin Proteasome System (UPS). The UPS handles approximately 80 % of degradation in eukaryotes (Lilienbaum, 2013), with the remaining 20 % achieved through lysosomal proteolysis (Mizushima and Komatsu, 2011). Both pathways depend on ubiquitination as the mechanism by which proteins are tagged and recognised as substrates destined for degradation. Both pathways also utilise shuttle factors, such as the UBQLNs, to close the gap between substrates and degradation machineries to facilitate degradation (Jantrapirom, Piccolo and Yamaguchi, 2019). However, the mechanisms by which this proteolysis is achieved varies greatly.

1.3.1 Autophagy-Lysosome Pathway

Macroautophagy (often referred to as autophagy) involves the formation of a transient organelle, called an autophagosome, to envelop cytosolic material and degrade it (Kiel, 2010). The autophagosome is a double-membrane vesicle formed on the target which undergoes elongation to envelop cytosolic material. This fuses with the endosomal-lysosomal system, forming an autolysosome (Kiel, 2010). The encased material is degraded by lysosomal proteases into amino acids, which are subsequently transported back across the membrane into the cytosol for re-use (Kuma and Mizushima, 2010).

In general, autophagy is considered to be a non-selective process. However, selective and specialized types of autophagy exist which have an important role in health and disease. For example, xenophagy sequesters bacteria (Baxt, Garza-Mayers and Goldberg, 2013) and viruses (Kim, Lee and Jung, 2010) whilst aggrephagy handles aggregated proteins (Lamark and Johansen, 2012). Degradation of some organelles is also achieved through selective autophagy; pexophagy sequesters peroxisomes (Iwata *et al.*, 2006) whilst mitophagy sequesters and degrades mitochondria (Narendra *et al.*, 2008). Macroautophagy and selective autophagy both use a hierarchy of autophagy related (ATG) protein machinery to form the autophagosome and utilise identical machinery to form the autolysosome (Lamb, Yoshimori and Tooze, 2013). The major difference between the two lies in the identification of particular substrates for cargo-specific degradation. Selective autophagy uses a variety of cargo receptors and adaptors such as p62, Optineurin and NBR1 (Bjørkøy *et al.*, 2005; Kirkin *et al.*, 2009; Korac *et al.*, 2013). These bind the selected cargo to the Microtubule-associated protein 1A/1B-light chain 3 (LC3) family members which are associated with the autophagosome membrane (Birgisdottir, Lamark and Johansen, 2013). Cargo receptors typically contain one or more LC3-interacting region (LIR)(Pankiv *et al.*, 2007) and a ubiquitin-binding UBA domain to act as a cargo-binding site for ubiquitinated substrates. Furthermore, cargo receptors have an intrinsic ability to oligomerize, which enables the formation of dense cargo-containing clusters. The activity of cargo receptors is increased by phosphorylation (such as in phosphorylation of Ser43 in the p62) and with the use of cargo adaptors (such as autophagy-linked FYVE (ALFY) (Filimonenko *et al.*, 2010) which binds ubiquitinated aggregates through p62 and links to the autophagic machinery via binding of ATG5.

Though not very well understood, the UBQLN proteins may play a role in the regulation of autophagy, as UBQLNs associate with autophagy components and substrates. Endogenous UBQLN1, UBQLN2 and UBQLN4 bind to the autolysosome component LC3 and facilitate the maturation of LC3-I into LC3-II (Rothenberg *et al.*, 2010; Yun Lee, Arnott and Brown, 2013; Wu *et al.*, 2020), which then initiates formation and elongation of the autophagosome (Parzych and Klionsky, 2014). Furthermore, this mechanistic action is perturbed in disease; overexpression of an ALS-linked mutation (P497H) in UBQLN2 causes a decreased LC3-II to LC3-I ratio

and reduces expression levels of several autophagy proteins over time (Chen *et al.*, 2018). This could suggest an additional autophagy mechanism involving UBQLN2 or may be an artefact of the overexpression system. UBQLN presence has been detected in autophagosome isolations from mouse liver, indicating a role for UBQLN in macroautopahgy *in vivo*. However, due to the localisation of UBQLN2 to aggregates (Deng *et al.*, 2011; Williams *et al.*, 2012) and the subsequent clearance of aggregates by autophagy (Lamark and Johansen, 2012), it is possible this association is indirect. UBQLN2 has also been reported to interact with mutated Huntingtin protein, a substrate of autophagy, facilitating its clearance through the ALP (Chuang *et al.*, 2016). A final mechanism by which UBQLN2 contributes to autophagy is through the interaction with the proton pump V-ATPase to protect it from degradation and promote stable V-ATPase formation (Şentürk *et al.*, 2019). V-ATPase function is essential for lysosome acidification and autophagic degradation (Şentürk, Mao and Bellen, 2019; Wu *et al.*, 2020), and loss of UBQLN2 or ALS-linked mutations (UBQLN2^{P497S}) resulted in reduced autophagosome acidification (Wu *et al.*, 2020).

Whilst it is clear the UBQLNs have some level of involvement in autophagy, the molecular mechanisms and the downstream consequences are not yet fully apparent. The role of UBQLNs, particularly UBQLN2, is better studied and understood in the other major degradation pathway: the Ubiquitin Proteasome System.

1.3.2 Ubiquitin Proteasome System

The UPS is the primary degradation system in eukaryotes (Husnjak *et al.*, 2008), playing a role in cell cycle progression, signal transduction, transcription, and protein homeostasis (Glickman and Ciechanover, 2002). Multiple stages and proteins are required for full function of the UPS, including enzymes (E1, E2, E3) for the ubiquitination of target substrates, shuttle proteins to direct these substrates to the proteasome, and the proteolytic activity of the proteasome.

Ubiquitination is a post-translational modification which covalently attaches ubiquitin onto protein substrates at specific positions which act as signalling for a variety of cellular pathways (Pickart, 2001; Weissman, 2001). One of these destinations is the 26S proteasome, where misfolded, damaged, or aggregated

proteins will be degraded. The 26S proteasome is a huge enzyme complex ~2.5 MDa in size, composed of the 20S core particle (CP) (Coux, Tanaka and Goldberg, 1996) and one or two 19S regulatory particles (RP)(Glickman *et al.*, 1998). The CP is barrel shaped, with two sets of three proteolytic active sites within the cavity of the barrel (Löwe *et al.*, 1995; Pickart, 2001). The RP recognizes ubiquitin chains via intrinsic ubiquitin receptors to unfold the substrate and translocate into the catalytic compartment of the CP (Voges, Zwickl and Baumeister, 1999). Proteasome-interacting proteins, such as shuttling factors, ubiquitin ligases, and deubiquitinating enzymes, transiently associate with the 26S proteasome to modulate its function.

In summary, the UPS consists of 3 key stages: Conjugation, translocation, and degradation. Conjugation is the first step in the degradative pathway, covalently attaching mono or multimeric ubiquitin chains onto a protein through a series of increasingly specific, sequential steps (Hershko *et al.*, 1983). This tagged substrate is then bound by chaperone proteins to translocate it to the proteasome (Jantrapirom, Piccolo and Yamaguchi, 2019). In some cases, aggregates may first be solubilised by HSP70-HSP110 disaggregase activity, followed by shuttling factor binding (Nillegoda *et al.*, 2015; Hjerpe *et al.*, 2016). UBQLN2 is one of these shuttling proteins, capable of binding ubiquitinated substrates and the proteasome via its UBA and UBL domains, respectively (Hjerpe *et al.*, 2016; Zheng, Yang and Castañeda, 2020). Finally, the ubiquitinated substrate is degraded by the large protease into short peptide strands of ~2-10 amino acids in length (Kisselev *et al.*, 1999). These peptides are rapidly digested into amino acids by cytosolic peptidases (Collins and Goldberg, 2017) for repurposing, with the ubiquitin tag also cleaved off and re-used (Wing, 2003; Lee *et al.*, 2011; Eletr and Wilkinson, 2014). Whilst the ALP is capable of removing large substrates, and sometimes pathogens *in toto*, the UPS is much more efficient at degrading substrates, working on a scale of seconds as opposed to minutes (Bard *et al.*, 2019). Whilst many factors affect protein turnover, such as cellular stress, an average speed of degradation for the 26S proteasome has been suggested at 20 s per 300 amino acid protein (Bard *et al.*, 2019). Considering the high concentration of proteasomes within a single cell (nm - μM range) (Albert *et al.*, 2017), and the dynamic assembly of these (Marshall and Vierstra, 2019), it is clear the UPS is a vital mechanism in proteostasis.

1.3.3 Protein quality control crosstalk

Of course, numerous proteins identified above must also undergo protein turnover to maintain a healthy proteome. Notably, the UBQLNs are a substrate of both macroautophagy and chaperone-mediated autophagy (CMA), with the CMA targeting motif (KFERQ) present in UBQLN1, UBQLN2, and UBQLN4 (Rothenberg *et al.*, 2010). As such, the correct degradation of the UBQLNs via the autophagic pathway may provide an element of control over the homeostatic regulatory pathways; if compromised, any problems caused by perturbances in autophagic flux could be compounded further.

1.3.3.1 Crosstalk

On a larger scale, turnover of the proteasome itself must occur in order to control the quality of the machinery and abundance of the proteasome pool. In non-apoptotic cells, this can occur in two ways. First, non-functional proteasome subunits can be removed by the UPS itself, prior to their integration into the complete 26S proteasome. HSP42 appears to play a crucial role in yeast, combining these subunits into cytoplasmic condensates from which they are cleared by active 26S proteasomes (Peters *et al.*, 2015; Nahar *et al.*, 2019). The second pathway of degradation is through a selective autophagy pathway, termed proteaphagy (Marshall and Vierstra, 2015; Marshall *et al.*, 2015; Marshall, McLoughlin and Vierstra, 2016). As the preferred catabolic route for large, heterogenous material, it is not surprising that autophagy has been discovered to play a hand in the degradation of the proteasome. Degradation of the proteasome upon nutrient starvation is a likely universal event, albeit with varied mechanisms and receptors involved between eukaryotes. Conversely, plant and yeast cells can utilise a second proteophagic pathway to allow clearance of non-functional 26S proteasomes (Marshall *et al.*, 2015; Marshall, McLoughlin and Vierstra, 2016; Nemec *et al.*, 2017), with the human counterpart of the receptors implicated in aggrephagy (Lu, Psakhye and Jentsch, 2014; Marshall, McLoughlin and Vierstra, 2016). This suggests an overlap between the two autophagic machineries. Whilst study of these degradative pathways often occurs in isolation, it should always be kept in mind the interplay between the systems, and the downstream implications that changes in one may have on the other.

The UPS and ALP act simultaneously and share some components of their molecular machinery. Primarily, the post-translational modification ubiquitination is employed by both the ALP and the UPS to label substrates for degradation, with both processes sharing the same ubiquitin pool (Dikic, 2017). Furthermore, multiple substrates require an additional complex (p97/VCP/Cdc48 ATP complex) in order to extract ubiquitinated proteins from membranes for proteasomal degradation (Dantuma and Hoppe, 2012) or selective autophagy (Stolz *et al.*, 2011; Meyer, Bug and Bremer, 2012). Interestingly, the E3 ligase Parkin has a dual role in the co-ordination of the UPS and ALP. Classically known for its central role in mitophagy, Parkin mediates two subsets of mitochondrial substrates, directing one towards proteasomal degradation (Chan *et al.*, 2011; Yoshii *et al.*, 2011) and the other towards autophagic degradation (Heo *et al.*, 2015).

These distinct degradative pathways can also influence each other's activity. An example of this is the termination of autophagy, mediated by the proteasomal degradation of ULK1 (Antonioli *et al.*, 2014; Liu *et al.*, 2016). As part of a complex, ULK1 (unc-51-like kinase 1) drives formation of the phagophore through activation of the VPS34 (vacuolar protein sorting 34) and via mediation of ATG trafficking (Zachari and Ganley, 2017). Once autophagy has been stimulated, the autophosphorylation of ULK1 facilitates its ubiquitination and subsequent degradation. Thus, proteasomal degradation acts to mediate the amplitude and duration of autophagy (Liu *et al.*, 2016). The autophagy receptor p62 is another protein frequently involved in crosstalk between the systems. p62 has a well described role in delivering cargo for autophagy. It also plays a role in the UPS by both delivering cargo to the 19S particle of the 26S proteasome (Seibenhener et al., 2004), or by targeting the proteasome for autophagy. An interesting aspect of this is the apparent dual role that the UBA domain plays: it determines the function of p62 as either substrate targeting to the proteasome or targeting the proteasome to autophagy (Cohen-Kaplan *et al.*, 2016). Finally, not only can proteins impact on the mechanisms within which they are acting, but the systems can impact on the proteins. In the case of p62, phosphorylation can be induced by activation of xenophagy, mitophagy, and aggrephagy (Ichimura *et al.*, 2013), as well as proteasomal dysfunction (Kageyama *et al.*, 2014) to increase its binding affinity to ubiquitin (Lim *et al.*, 2015). The crosstalk of the degradative pathways

and the proteins within them co-ordinates the entire degradative flux of the cell (McEwan and Dikic, 2011). Whilst p62 appears to have a diverse role amongst the mechanisms (Matsumoto *et al.*, 2011; Reidick *et al.*, 2014; Dikic, 2017), ubiquitination is the clear commonality across these major degradation pathways. As such, a mechanism for encoding, decoding, and propagating the information contained within the ubiquitin code must be employed.

1.4 The Ubiquitin Code

The specific identification and tagging of proteins intended for proteolysis is the first step in protein degradation. The initial events governing which proteins will be destined for degradation remains unclear in some cases, though several factors have been determined as playing a governing role in degradation determination. These include the presence of specific sequences (such as the destruction box in the cyclins)(Murray, 2004), or the N-End rule (where a protein containing small amino acid at the N-terminus is more stable than its equivalent with a basic amino acid)(Rao *et al.*, 2001). Other degradation triggers include oxygen levels and proline hydroxylation (Hirsilä *et al.*, 2003), or phosphorylation (Vlach, Hennecke and Amati, 1997). Aged and denatured, or misfolded, proteins may also present a hydrophobic patch which identifies the protein as a substrate for degradation (Varshavsky, 2005). Whilst the selection process for labelling some proteins for degradation remains elusive, the mechanism by which this label is achieved is not: ubiquitination.

1.4.1 Ubiquitin and Ubiquitination

Ubiquitin, named due to its ubiquitous expression throughout eukaryotes, is a highly conserved 76 amino acid protein. Ubiquitination is the shared mechanism by which all major degradation pathways determine the protein's fate (Dikic, 2017). Different lengths and linkage types of ubiquitin may favour interaction with various receptors. For example, mono-ubiquitination is required for receptor endocytosis (Hoeller *et al.*, 2006) whilst the attachment of different poly-ubiquitin regulates numerous cellular processes, such as protein degradation (Thrower *et al.*, 2000) and immune signalling (Gerlach *et al.*, 2011) (Figure 1.2).

This ubiquitin code and the proteins decoding it directs substrates toward a particular fate (Dikic, 2017).

Ubiquitin is covalently attached to substrates via an increasingly specific enzymatic cascade. Ubiquitin is activated by attachment to E1 activating enzyme which is then transferred to an E2 conjugating enzyme. Each E2 enzyme is associated with several E3 ligases which recognise different target proteins (Hershko and Ciechanover, 1998). This cascade uses increasing numbers of ligases in each step; only two E1 encoding genes (UBA1 and UBA6) were revealed in comparative genome analysis, but ~40 E2 and ~800 E3 ligase genes have been discovered (Semple *et al.*, 2003). Activated ubiquitin is attached via it's C-terminus to the side chain of a lysine on substrate proteins, via interaction with an E2-E3-ubiquitin complex (Nandi *et al.*, 2006). Modifications can result in mono-ubiquitination, where a single ubiquitin molecule is covalently attached to one lysine residue, or multiple lysine residues of the substrate, in the case of multimono-ubiquitination. Additional ubiquitin can be coupled to its own lysine residues, forming a poly-ubiquitin chain. These differential chain lengths create one layer of the ubiquitin code (Figure 1.2).

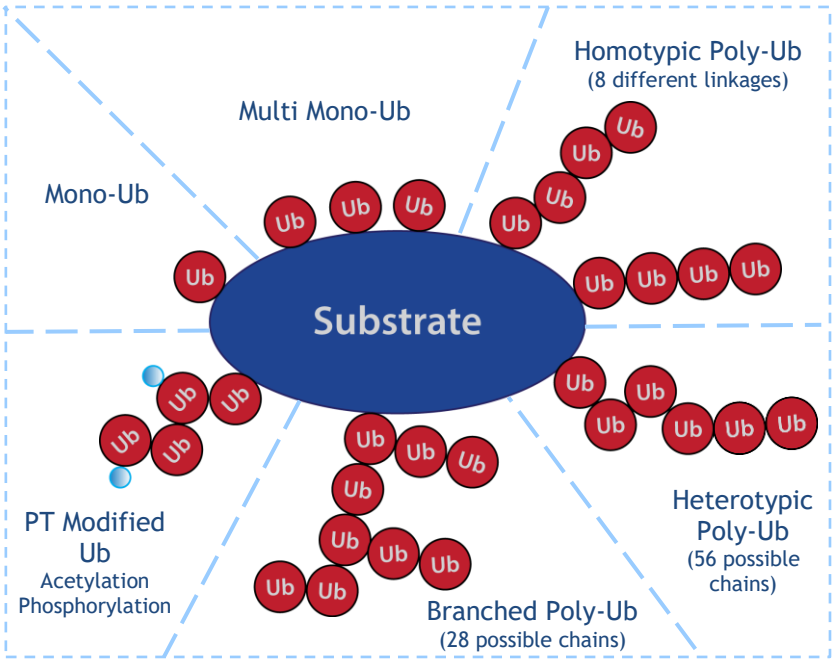


Figure 1.2 The first layer of the Ubiquitin code. Schematic demonstrating the different variations of ubiquitination possible, including further post-translational modifications of ubiquitin. Image taken and adapted from (Pérez Berrocal *et al.*, 2020).

The vast scope of the ubiquitin code is apparent when the varieties of ubiquitination are considered. Poly-ubiquitination can occur via seven lysine residues which allow the formation of distinct isopeptide chain linkages. These linkages are K6, K11, K27, K29, K33, K48 and K63. Linkage can also occur via the N-terminal methionine, termed M1 linkage. Ubiquitination was first described as a post-translational modification targeting proteins for proteasomal degradation in 1980 (Ciechanover *et al.*, 1980). Since then, both distinct and over-lapping roles have been discovered for the different chain linkages. For instance, K48-tetraubiquitin was initially believed to be the minimum requirement for proteasomal degradation (Thrower *et al.*, 2000). However, following experimentation in yeast (Xu *et al.*, 2009) and HeLa cell lines (Jacobson *et al.*, 2009), it is now suggested the proteasome will accept all other chain types except K63. Differential pathways for the various chain linkages becomes apparent outside of proteolysis (Figure 1.3). For example, K11-linked chains have a role in mitotic exit alongside proteasomal degradation (Matsumoto *et al.*, 2010), whilst K27-linked chains are involved in DNA damage response and innate immunity (Wang *et al.*, 2014; Gatti *et al.*, 2015). K29-linked chains acts as an inhibitor of Wnt signalling, with downstream implications in embryogenesis, tumorigenesis, and other human diseases (Clevers and Nusse, 2012). K33-linked chains are implicated in post-Golgi trafficking (Yuan *et al.*, 2014), acting as a negative regulator of both T-cell antigen receptor (Huang *et al.*, 2010) and AMPK-related protein kinases (Al-Hakim *et al.*, 2008). The role of K6-linkages is less clear, but these have been reported to increase with UV radiation (Elia *et al.*, 2015) and DNA repair (Morris and Solomon, 2004; Nishikawa *et al.*, 2004), alongside tagging mitochondrial outer membrane proteins upon depolarization of the organelle (Ordureau et al., 2014). Better understood are the K48 and K63 linked chains, implicated in proteasomal degradation and lysosomal degradation respectively (Thrower *et al.*, 2000; Mukhopadhyay and Riezman, 2007). K63 chain linkages have also been implicated in DNA damage (Al-Hakim *et al.*, 2010) and have been reported to promote translation through stabilisation of the polysomes (Spence *et al.*, 2000). Finally, K48 linked chains are the most abundant linkage in all organisms and Lys48 is the only essential lysine in yeast (Chau *et al.*, 1989). K48 linkage levels rapidly increase following proteasome inhibition (Peng *et al.*, 2003; Xu *et al.*, 2009; Kaiser *et al.*, 2011; Kim *et al.*, 2011), thus demonstrating their

key role in proteasomal degradation. These variations in the chain linkages adds a second layer of complexity to the ubiquitin code.

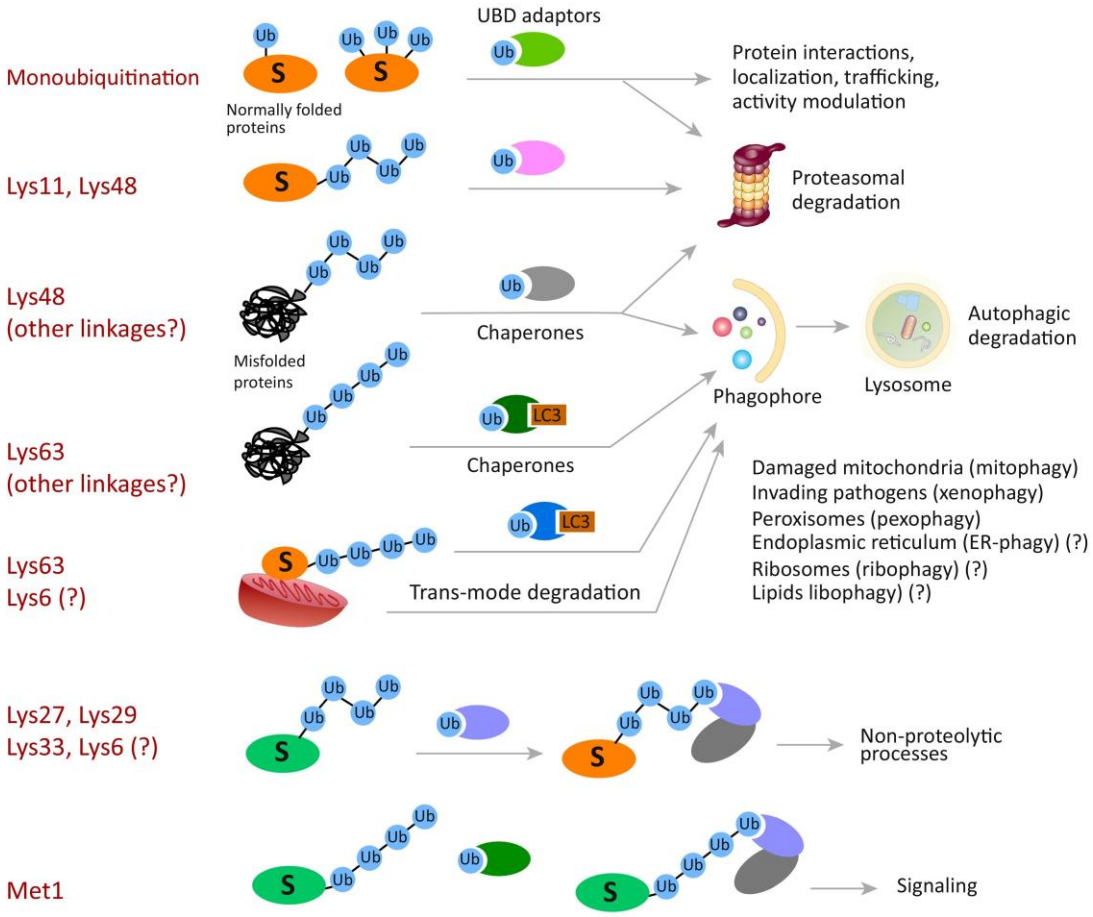


Figure 1.3 Converting the ubiquitin code into biological processes. Diagram depicting some of the cell fates of ubiquitinated proteins and the role of UBDs in facilitating this. Image taken and adapted from (Kwon and Ciechanover, 2017).

Each chain linkage adopts a distinct 3D conformation and is present at varying frequencies within eukaryotic cells (Xu *et al.*, 2009). Recruitment of particular combinations of E2-E3 enzymes dictates the chain linkage type, and chains can be either homo- or heterotypic: Ubiquitin chains composed of a single linkage type form homotypic chains, where a polymer of varying linkages is a heterotypic chain. Branching can also occur in heterotypic chains, where one ubiquitin molecule is ubiquitylated at two or more sites (Meyer and Rape, 2014). Further ubiquitin modifications can also occur in the form of acetylation (Ohtake *et al.*, 2015) or phosphorylation (Kane et al., 2014; Kazlauskaite et al., 2014; Koyano et al., 2014; Ordureau et al., 2014), adding increased complexity to the ubiquitin code (Figure 1.2). Considering the huge number of distinct chain linkage types, lengths, and modifications, a mechanism for interpreting the ubiquitin code and propagating

the information it carries must be employed to ensure correct substrate fate. This decoding of the ubiquitin code is achieved by ubiquitin binding proteins.

1.4.2 Ubiquitin binding proteins

Recognition of the various ubiquitin chains is a crucial step in protein degradation. This recognition is required by shuttling factors (to translocate the substrate to the correct location) (Jantrapirom, Piccolo and Yamaguchi, 2019), deubiquitinating enzymes (to cleave ubiquitin from substrates) (Komander, Clague and Urbé, 2009), and intrinsic ubiquitin receptors within the proteasome (to allow degradation through the proteasome to occur). Each of these proteins contains a ubiquitin binding domain (UBD) to interpret the ubiquitin code.

1.4.2.1 Intrinsic Ubiquitin Receptors

A number of UBDs have evolved to recognise and translate the ubiquitin code. The 26S proteasome contains two sets of three intrinsic ubiquitin receptors named Rpn1, Rpn10 and Rpn13 (Figure 1.4A). These receptors contain a ubiquitin interacting motif (UIM) domain at their C-terminus which recognizes ubiquitinated substrates on the proteasome. The UIM domains are also capable of binding ubiquitin-like (UBL) domains of other proteasome-interacting proteins, such as shuttling factors (Tsuchiya, Endo and Saeki, 2020).

1.4.2.2 Extrinsic Ubiquitin Receptors- Shuttling Factors

Originally identified in yeast, shuttling factors such as the UV excision repair protein Rad23 (Rad23) (Perozzi and Prakash, 1986), DNA damage inducible protein 1 (Ddi1) (Kaplun *et al.*, 2005) and Ubiquitin domain-containing protein DSK2 (Dsk2) (Biggins, Ivanovska and Rose, 1996) are known as UBL-UBA proteins. The human homologues of these proteins are Rad23A/B, Ddi1/2, and UBQLN1-4 respectively. These proteins contain a UBL domain N-terminally and one or two UBA domains C-terminally. They simultaneously bind to intrinsic receptors of the proteasome via UBL domains and to ubiquitin of tagged proteins via their UBA domains (Bertolaet et al., 2001b; Wilkinson et al., 2001).

The number of UBA domains can vary between UBL-UBA proteins (Figure 1.4B). Rad23 has two UBA domains where Dsk2 and its human homologues contain only one. In yeast, Rad23 and Dsk2 synergistically contribute to shuttling ubiquitinated substrates to the proteasome (Tsuchiya *et al.*, 2017). Ddi1/2 is less understood, however both its UBL and UBA domains are capable of binding ubiquitin as well as the proteasome (Nowicka *et al.*, 2015).

The association of UBL-UBA proteins with intrinsic ubiquitin receptors can vary between proteins (Figure 1.4C). The intrinsic ubiquitin receptor Rpn10 has two UIM domain available for UBL binding. In humans, Rad23A has a preferential binding affinity for UIM-2 in Rpn10 (Kang, Chen, *et al.*, 2007). In contrast, the UBL domain of UBQLN2 has a 25-fold stronger affinity for UIM-1 over UIM-2 of the same receptor (Chen *et al.*, 2019). Despite this differential UIM binding, both Rad23A and UBQLN2 UBL domains are capable of activating the mammalian proteasome (Kim and Goldberg, 2018; Collins and Goldberg, 2020). However, it is currently unclear how the co-ordination of the UIM/UBL interaction is achieved. Human Rpn13 binds ubiquitin via a conserved anti-parallel ß-sheet region called the Pleckstrin-like receptor for ubiquitin (Pru) domain. This domain binds K48-diubiquitin whilst also binding UBL domains in Rad23A and UBQLN2 (Husnjak *et al.*, 2008). Human Rpn1 is a large proteasomal subunit composed of 11 proteasome/cyclosome (PC) repeats of 30-40 residues. Each unit forms a helix-turn helix hairpin (Kajava, 2002; He *et al.*, 2012). These PC repeats form closed toroid domains of T1 and T2 sites. Interestingly, T1 is capable of binding both ubiquitin and Rad23A/UBQLN2 UBL domains (Shi *et al.*, 2016). This suggests the T1 site of human Rpn1 acts as a receptor for both ubiquitinated substrates and shuttling factors. Whilst there is some functional redundancy in the intrinsic ubiquitin receptors of yeast, other studies have discovered preferential binding of the receptors to different chain linkages. Rpn10 is the primary receptor for K48-linked ubiquitin substrates whilst Rpn1 and Rpn13 are co-receptors for multi-ubiquitinated substrates (Martinez-Fonts *et al.*, 2020). Rpn1 also binds K11 and K48 branched ubiquitin chains more efficiently than K48-linked homotypic chains (Boughton, Krueger and Fushman, 2020). Together, these intrinsic receptors form a versatile platform for ubiquitin decoding.

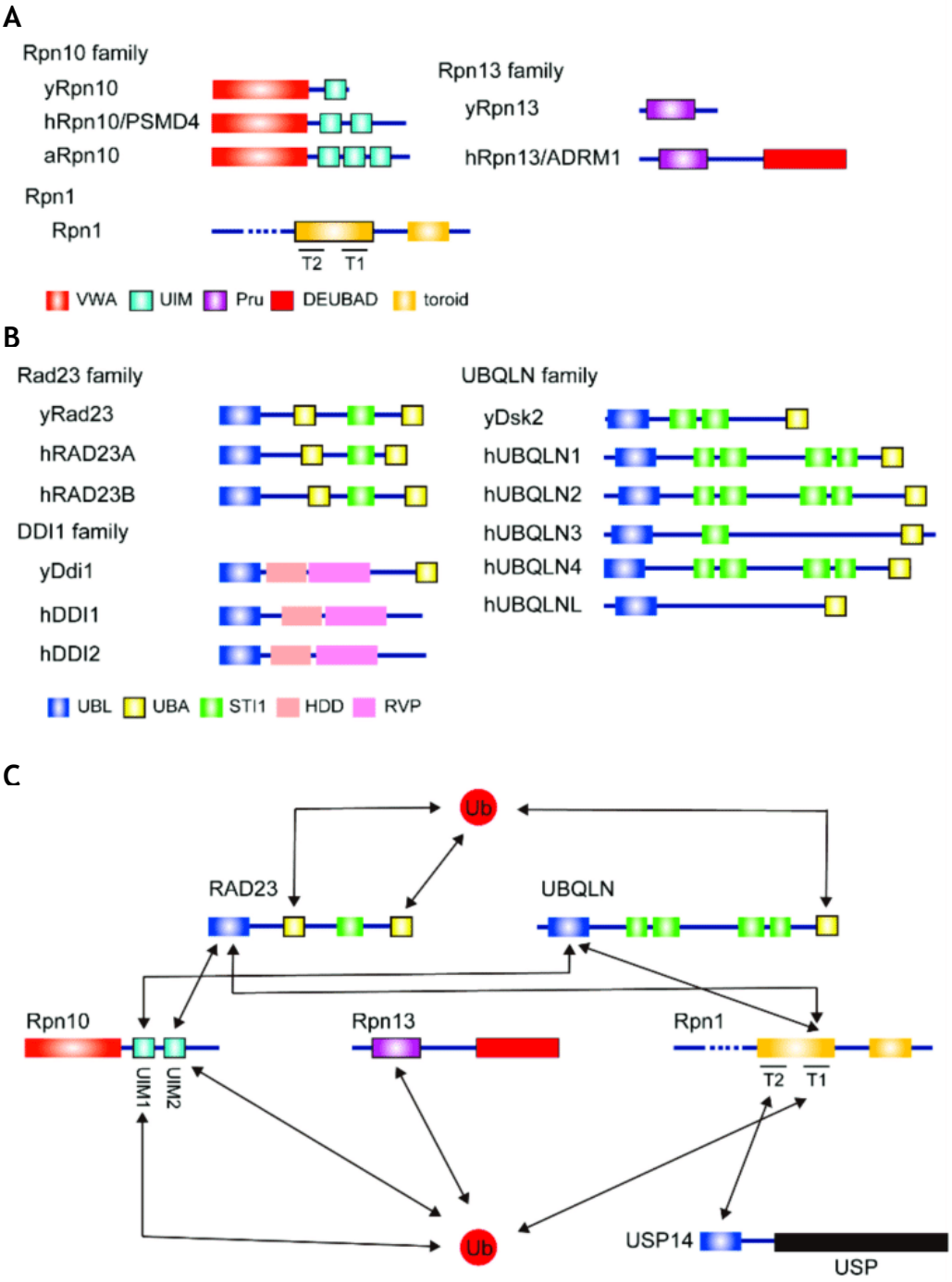


Figure 1.4 Ubiquitin receptors of the UPS. Schematic outlining the domain architecture of intrinsic ubiquitin receptors (A), extrinsic ubiquitin receptors (B), and the interplay between them (C). The prefix "h" indicates the human protein; "y" indicates the yeast protein; "a" indicates the arabidopsis protein. *UIM*, ubiquitin-interacting motif. *Pru*, Pleckstrin-like receptor for ubiquitin. *DEUBAD*, deubiquitinase adaptor. *UBL*, ubiquitin-like. *UBA*, ubiquitin associated. Not discussed in this thesis are *HDD*, helical domain of Ddi1; *RVP*, predicted retroviral protease-like; *VWA*, von Willebrand factor A; USP, ubiquitin-specific protease. Image taken and adapted from (Tsuchiya, Endo and Saeki, 2020).

1.4.2.3 Extrinsic Ubiquitin Receptors- Deubiquitinating Enzymes

Ubiquitination is a reversible modification, where the ubiquitin chain or its attachment to a substrate can be hydrolysed by deubiquitinating enzymes (DUBs) (Clague *et al.*, 2013). The mammalian RP of the proteasome has three associated DUBs: POH1/PSMD14, USP14, and UCH37 (Komander, Clague and Urbé, 2009). DUBs disassemble the ubiquitin chains and remove the proximal ubiquitin from the substrate to maintain the ubiquitin pool prior to degradation. They are also involved in editing ubiquitin chains and processing ubiquitin precursors (Komander, Clague and Urbé, 2009). Similar to other ubiquitin biding proteins, DUBs have distinct binding affinities conferred by ubiquitin chain length and linkage type.

1.4.3 Specificity within the code

A common theme amongst UBDs is the ability to differentially bind substrates to direct them to their ultimate fate. UBDs recognize and bind distinct surface regions of ubiquitin (Husnjak and Dikic, 2012). X-ray crystallography and NMR studies have revealed hydrophobic patches on residues I44 and I36 are positioned differently between ubiquitin polymers. For example, K63 and M1-linked di-ubiquitin polymers adopt an open conformation where the only contact sites are those existing at the linkage site (Komander, Clague and Urbé, 2009). In contrast, dimers of K6, K11, K29, K33 and K48 form intramolecular interfaces between the two ubiquitin molecules (Cook *et al.*, 1992; Bremm, Freund and Komander, 2010; Matsumoto *et al.*, 2010; Hospenthal, Freund and Komander, 2013; Kristariyanto, Abdul Rehman, *et al.*, 2015; Kristariyanto, Choi, *et al.*, 2015; Michel *et al.*, 2015). These chain variations confer the specificity of UBD-binding, aiding ubiquitin binding proteins in discerning the ubiquitin code needed for cellular survival.

The human genome encodes more than 20 UBDs, and linkage specificity of ubiquitin binding is well established (Callis, 2014). This specificity may be conferred through spatial arrangement of UBDs, either encoded within a single protein such as in Rad23, or by combining domains within a multimeric complex. Combining domains may restrict simultaneous occupancy of the two binding sites to particular chain configurations, thus increasing chain specificity.

1.4.4 Role of UBQLN2 as a ubiquitin receptor

The molecular role of the UBQLN family and their interactions with their substrates are only slowly being uncovered, with many roles linked to proteasomal degradation: UBQLN4 is important for the DNA damage response by facilitating the proteasomal degradation of the repair factor MRE 11 (Jachimowicz and Reinhardt, 2019), whilst UBQLN2 operates with chaperones to degrade misfolded and aggregated proteins via the proteasome (Hjerpe *et al.*, 2016). There is, however, a growing body of evidence suggesting roles for the UBQLNs outside of the classical proteasomal degradation pathway, including autophagy, ERAD pathway, DNA/RNA metabolism, protein trafficking and cytoskeletal regulation (Wu et al., 1999; Lim et al., 2009; Xia et al., 2014; Gilpin, Chang and Monteiro, 2015; Halloran et al., 2019; Jachimowicz and Reinhardt, 2019; Şentürk et al., 2019). Despite the wide array of cellular processes requiring UBQLN function, it is unclear how UBQLNs differentiate between the different roles.

Rad23 contains two UBA domains which are both required for ubiquitin binding (Chen and Madura, 2002) and potentially provide a regulatory function through dimerization (Bertolaet *et al.*, 2001a) or conferring the functional conformation of the protein (Chen and Madura, 2002). Whilst Rad23 contains multiple UBA domains to bind ubiquitin, the UBQLNs contain only one. This UBA domain is extremely well conserved between the UBQLNs. Given the high degree of similarity (Zhang, Raasi and Fushman, 2008), it is assumed the UBA domain of UBQLN2 will behave in a similar manner to UBQLN1. When studied in isolation, the UBA domain of UBQLN1 binds mono-ubiquitin with a K_d of ~20 μM and ubiquitin chains are bound indiscriminately with a K_d of ~4 μM (Zhang, Raasi and Fushman, 2008). This lack of specificity between polyubiquitin chains is surprising, as the UBA domains of Rad23A/B and p62 both confer their own specificity with no external influence (Raasi et al., 2004; Seibenhener et al., 2004). However, when the UBA domain of UBQLN1 is expressed in tandem as a head-to-tail tetramer (referred to as TUBEs), the affinity for K63-tetraubiquitin chains increases by ~1000 fold whilst the affinity for K48-linked tetraubiquitin increased by ~100 fold (Hjerpe *et al.*, 2009). The increased affinity for K63-linked chains is not surprising, given that in the context of the full-length protein, UBQLN1 UBA domains show a greater specificity for K63 linked ubiquitin chains over K48 linkages, and a further

preference to binding longer K63 poly-ubiquitin chains (Harman and Monteiro, 2019). This does however create an interesting question, would a multimer of UBQLN2 UBA domains have a larger increase in affinity for K48-linked over K63-linked polyubiquitin chains? Furthermore, the increased affinity of the TUBEs for tetraubiquitin was not mimicked when increasing the concentration of single UBA domain by a 6-fold molar excess to the TUBEs under investigation (Hjerpe *et al.*, 2009). This signifies the importance of tandem arrangement of UBA domains for optimal binding of polyubiquitin.

It would therefore be logical to hypothesise a similar arrangement mechanism may be in play with UBQLN2 *in vivo*. It has been reported that ubiquitin binds to one face of the UBA domain of UBQLN2, established through NMR titration studies and mapping chemical shift perturbations (CSPs) (Dao *et al.*, 2018). However, multimerization of UBQLN2 may organise the UBA domains and their ubiquitin binding sites to confer chain specificity. Finally, tandem UBA domains protect ubiquitinated proteins from deconjugation by DUBs and proteasomal degradation (Hjerpe *et al.*, 2009). The mechanism by which UBQLN2 triages proteins towards different fates is still unclear, though this presents a potential protective mechanism whilst protein fate is determined. Together, this is compelling evidence that multimerization of UBA domains can impact on the affinity, and potentially specificity, of ubiquitin binding.

The UBA domains of UBQLN1 and UBQLN2 are 98 % identical at the sequence level (Zhang, Raasi and Fushman, 2008), allowing for assumptions to be extrapolated when investigating the UBA domain alone. However, UBQLN2 contains a unique proline rich repeat (PRR) region N-terminally to its UBA domain. This is the region to which the majority of ALS-causing mutations locate (Deng *et al.*, 2011), with numerous reports of degradation dysfunction as a result (Table 1.1). One study found ALS-linked UBQLN2 mutants interacted less effectively with the intrinsic receptor of the proteasome, Rpn10 (Chang and Monteiro, 2015). However, the mutations did not appear to impact the ability of the UBA domain to bind to substrates, nor did it affect overall proteasome activity. Thus, it was hypothesised that the ALS-linked mutants were defective in transporting the ubiquitinated substrates to the proteasome, possibly due to a decreased affinity (Chang and Monteiro, 2015).

Whilst the PRR region is conserved in UBQLN2 from other species, the sequence is not readily identifiable in any other human protein. The unique existence of this polyproline region, and the severe impact mutations within it have, raises an important question regarding the function and its role in disease development.

1.5 Polyproline Regions

1.5.1 Poly-L-proline type II helix

The poly-L-proline type II helix (referred to as polyproline-II from here on in) is remarkably different in its structure from the other two major secondary structures (α-helix and ß-strand). A number of protein regions historically believed to be intrinsically disordered in nature have since been characterised as polyproline-II helices (Rath, Davidson and Deber, 2005), with increasing evidence that polyproline-II helices can be found within most proteins (Adzhubei and Sternberg, 1993; Stapley and Creamer, 1999; Jha *et al.*, 2005; Berisio *et al.*, 2006). This relatively newly characterised protein structure has been identified as playing a role in structure, function, and interactions of proteins and peptides (Siligardi and Drake, 1995a; Kay, Williamson and Sudol, 2000; Hicks and Hsu, 2004).

The "random coil" model is often still used to describe the conformation of unfolded peptides and proteins. The model implies amino acid residues are distributed randomly within the sterically permitted regions of the Ramachandran plot (Schweitzer-Stenner, 2011). However, early work by Tiffany and Krimm observed a change of poly-L-glutamic acid and poly-L-lysine structure from an α-helical to an extended helical conformation at high pH and low salt (Tiffany and Krimm, 1968, 1969; Tiffany and Krimm, 1972). This suggested the peptides were not fully unstructured and could adopt a conformation similar to the polyproline-II helix. This led to the conclusion that the peptides investigated did not fit the "random coil" model. Instead, the high dielectric constant and the balance of electrostatic and steric interactions of charged side-chains favoured polyproline-II helix in water (Tiffany and Krimm, 1969). Whilst controversial at the time, the notion of polyproline-II presence in regions previously characterised as unordered

has since been supported extensively (Drake, Siligardi and Gibbons, 1988; Dukor and Keiderling, 1991; Woody, 1992).

At 3.1 Å per residue in comparison to the α-helical 1.5 Å per residue, the polyproline-II helix is an extended left-handed helix defined by the φ, ψ torsional angle cluster with the distribution maximum existing at -75° and 145°. (Siligardi and Drake, 1995b; Adzhubei, Sternberg and Makarov, 2013). In its ideal conformation, the polyproline-II helix has three residues per turn with a 3-fold rotational symmetry (Adzhubei, Sternberg and Makarov, 2013). The polyproline-II helix does not support regular patterns of interchain hydrogen bonds, leading to a more flexible conformation than that of an α-helix or ß-sheet (Stapley and Creamer, 1999; Kelly *et al.*, 2001; Cubellis *et al.*, 2005). Counter-intuitively to its name, the polyproline-II helix conformation may be adopted by peptides which are not dominated by, or even necessarily contain, proline (Makarov *et al.*, 1975). Nonetheless, when containing proline, the structural flexibility is limited slightly with the proline pyrrolidone ring. Due to this restriction, polyproline-II helices are the dominant conformation for proline rich regions (Williamson, 1994).

Polyproline-II helices often occur at the boundaries of α-helices (Adzhubei and Sternberg, 1993), in linker regions (Adzhubei and Sternberg, 1994; Dyson and Wright, 2005), interaction interfaces and protein terminal regions (Williamson, 1994). The polyproline-II helix occurs frequently in natural polypeptides (Makarov *et al.*, 1984, 1992; Adzhubei *et al.*, 1987) and is one of the common constituents of the collagen triple helix (Shoulders and Raines, 2009).

1.5.2 Collagen-Like Domains

Collagen is the most abundant protein in the human body, constituting ~2% of all proteins. The repetitive sequence of amino acids drives spontaneous self-assembly of three similar chains into a unique triple helical structure (Ricard-Blum, 2011). Further self-assembly into large fibrils occurs with longer triple helical structures, whilst shorter helices may provide rigidity to the protein or complement the biological activity of other functional domains (Brodsky and Persikov, 2005). Since the identification of collagen, and its iconic triple helix, numerous proteins have been discovered to contain collagen-like domains.

1.5.2.1 The Collagen Triple Helix

The distinctive collagen triple helix is composed of three left-handed polyproline-II helices supercoiled into a tight right-handed triple helix (Brodsky and Persikov, 2005). This tightly wound structure has a tenfold helical symmetry with a 28.6 Å axial repeat (Shoulders and Raines, 2009) with varying helical pitch between subsets of collagen: Proline-rich collagen possess a 7/2 helical pitch, whilst proline-poor regions possess a 10/3 pitch (Kramer *et al.*, 1999; Boudko *et al.*, 2008). This variability may play a role in the interaction of collagenous domains with other biomolecules. The internal residue position of the helix can only be accommodated by glycine, as the smallest amino acid. This produces the characteristic amino acid sequence of $(Gly\text{-}Xaa\text{-}Yaa)_n$, where Xaa and Yaa can be any amino acid, but are often favoured by Proline in human collagen (Shoulders and Raines, 2009). The formation of the collagen triple helix is entropy driven by the loss of surface water as the larger structure is formed (Prockop, 2004). It involves an initial, slow formation of a small nucleus followed by rapid propagation where the molecular architecture of the nucleus is extensively repeated to form the collagen triple helix (Prockop, 2004).

The amino acid composition of the collagen triple helix impacts the stability of the helix in numerous ways. The hydrogens on the Cα of the glycine allow a single interchain hydrogen bond formation per triplet to occur. The closed ring structures of imino acids (proline and hydroxyproline) limit rotation of the polypeptide backbone, thus providing rigidity to the structure (Prockop, 2004). The polyproline-II conformation decreases the entropic cost of collagen folding (Cram, 1988) by pre-organising the ring pucker of the imino acids to stabilise the collagen triple helix (Vitagliano *et al.*, 2001). Therefore, polyproline-II helices are common prerequisites for collagen triple helix formation.

The occurrence of proline in the Yaa position leads to stabilisation of the collagen triple helix via post-translational modification by the enzyme Prolyl-4-hydroxylase, forming a 4R-hydroxy-L-proline (4-Hyp). Thus, a characteristic motif of human collagen triple helices is $(Gly\text{-}Pro\text{-}4Hyp)_n$. 4R-Hydroxyproline is rarely found in other proteins, but its existence in the collagen triple helix locks the slightly flexible ring of the imino acid into a conformation that further stabilises the backbone, increasing the T_m by ~30ºC (Rosenbloom, Harsch and Jimenez,

1973; Sakakibara *et al.*, 1973). This extra thermal stability is due to stereoelectronic effects (Bretscher *et al.*, 2001; Improta, Benzi and Barone, 2001). Furthermore, hydroxyproline appears to be an aid in the nucleation and efficiency of collagen triple helix folding and receptor interactions (Xu *et al.*, 2003; Khoshnoodi *et al.*, 2006). Regions of the collagen triple helix with low 4-Hyp content are able to rapidly fold and unfold (Prockop, 2004), retaining some of the necessary flexibility required for function.

There are 28 genetically distinct types of human collagen, which can be classified into two broad categories. Fibrillar collagen is the most abundant, with perfect, non-interrupted $(Gly\text{-}Xaa\text{-}Yaa)_n$ repeats (Shoulders and Raines, 2009). Non-fibrillar collagen often contain interruptions in the $(Gly\text{-}Xaa\text{-}Yaa)_n$ repeats, such as type IV collagen which contains 20 interruptions (Traub and Piez, 1971). However, the collagen triple helix is always synthesised with non-collagenous domains adjacent to the helix. These domains may play a role in trimerization, chain registration, assembly, and biological interactions.

The number of proteins identified containing collagenous domains is increasing. This suggests the collagen triple helix is a basic motif, adaptable to a range of proteins and functions. Variations in the helix twist may alter recognition features or change the orientation of the collagen triple helix for binding.

1.5.2.2 Bacterial Collagen-like Proteins

Bacterial collagen-like proteins are a relatively new discovery, identified through database searches of the bacterial genome (Rasmussen, Jacobsson and Björck, 2003) following identification of the first bacterial collagen-like proteins Scl1 and Scl2 (Lukomski *et al.*, 2000; Rasmussen, Eden and Bjorck, 2000). These proteins form stable collagen triple helices, despite bacteria lacking the prolyl-4-hydroxylase enzyme required for the post-translational formation of 4R-hydroxyproline.

136 eubacterial genomes were analysed for sequence homology to $(Gly\text{-}Pro\text{-}Pro)_n$, with 56 hits identified in 25 different bacterial genomes (Rasmussen, Jacobsson and Björck, 2003). The number of $(Gly\text{-}Xaa\text{-}Yaa)_n$ repeats varied in length, from $_n$=7-745, with an average length of 76 repeats. All the bacterial collagen-like

sequences are flanked by non-collagenous domains, which is comparable to mammalian collagen. When looking at non-Gly residues in the collagen-like sequence, bacterial collagens have a proline content above 20%, with Pro preferentially located in the Xaa position and Thr and Gln frequently found in the Yaa position (Rasmussen, Jacobsson and Björck, 2003). This contrasts with animal collagen, where the proline is often located in the Yaa position and selectively hydroxylated into the stabilising 4R-hydroxyproline (Rosenbloom, Harsch and Jimenez, 1973; Sakakibara *et al.*, 1973). Multiple collagen-like sequences were identified in several bacterial genomes, suggesting the potential for heterotrimers to occur in a similar fashion to animal collagen.

Structural studies using recombinantly expressed bacterial collagen-like proteins were performed on eight different proteins (Xu *et al.*, 2002, 2010; Boydston *et al.*, 2005; Vandersmissen *et al.*, 2010). Triple helix formation was established by protease digestion and Circular Dichroism studies, with the thermal stability of the helix monitored by following the mean residue ellipticity at 220 nm over various temperatures. All bacterial collagen-like proteins expressed in solution formed stable collagen triple helices with a T_m in the range of 35-39ºC, and the presence of non-collagenous domains had little effect in the stability of the triple helix (Xu *et al.*, 2010; Yu *et al.*, 2010). This melting temperature is comparable to human collagen at 37ºC, which is surprising as bacterial collagen-like proteins lack the stabilizing effects of 4R-Hydroxyproline. Therefore, stability must be established via other mechanisms. It has been suggested electrostatic interactions stabilise regions rich in charged residues (Mohs *et al.*, 2007), whilst polar residues have also been found to play a role in stabilisation (Xu *et al.*, 2010). Furthermore, threonine in the Yaa position can be post-translationally glycosylated which stabilises the BclA bacterial collagen-like protein (Boydston *et al.*, 2005). The high calorimetric enthalpy observed for Scl2 (Yoshizumi *et al.*, 2009), despite the lack of 4-Hyp, indicates a high degree of hydrogen bonding mediated by hydration. This suggests bacterial collagen-like proteins may have similar hydration properties of mammalian collagen (Bella *et al.*, 1994). However, no direct studies were carried out on bacterial collagen-like proteins extracted from their natural bacteria. The achievement of this may shed more light on the stabilisation mechanisms involved in bacterial collagen-like proteins.

The stability of the triple helix can also be investigated through studies on collagen-like peptides. Stability of the triple helix appears to be conferred by length and amino acid composition (Persikov, Ramshaw and Brodsky, 2005). A minimum length of the $(Gly\text{-}Xaa\text{-}Yaa)_n$ repeats are required to form a triple helix, with stability levelling off with increasing length to fit a single exponential curve (Persikov, Ramshaw and Brodsky, 2005). The shortest protein of $(Gly\text{-}Xaa\text{-}Yaa)_{20}$ had a 5ºC lower thermal stability than longer constructs, which were equal in stability to mammalian collagen despite the lack of 4-Hyp (Han, Caswell, *et al.*, 2006; Han, Zwiefka, *et al.*, 2006). Furthermore, the stability of collagen constructs depends on amino acid composition, where high proline content is most stable (Yu, Brodsky and Inouye, 2011). It would be easy to presume the evolution of bacterial collagen-like protein stability has evolved to match host body temperatures. However, three of the eight bacterial collagen-like proteins identified belong to non-pathogenic bacteria (Xu *et al.*, 2010), suggesting alternative driving forces in the evolution of stable bacterial collagen-like proteins.

A characteristic property of animal collagen is the self-assembly of collagen molecules into fibrils (Ricard-Blum, 2011), but no natural higher order structure has been observed in bacterial collagen-like proteins. This is potentially due to the lack of 4-Hyp which is implicated in the self-assembly of triple helical molecules (Kramer *et al.*, 2000; Perret *et al.*, 2001). *In vitro* introduction of higher order structures was achieved with the collagenous domain of Scl2 protein (Yoshizumi *et al.*, 2009). Triple helical domains of ¼ and ½ the length of human fibrillar collagens can self-assemble into twisted and staggered fibrillar structures at neutral pH. However, these fibrils are small and do not show the periodic banding typically observed in animal collagen fibrils (Yoshizumi *et al.*, 2009). As bacteria are unicellular organisms, and as such lack an extracellular matrix, the biological role of bacterial collagen-like proteins likely differs from the characteristic structural role of animal collagens.

Natural expression of bacterial collagen-like proteins is evident in only a few cases (Karlström *et al.*, 2004; Karlström, Jacobsson and Guss, 2006), despite numerous species containing collagen-like sequences within their genome (Rasmussen, Jacobsson and Björck, 2003). As such, the biological role of bacterial collagen-like proteins has not been fully elucidated. Some bacterial collagen-like proteins

appear to have roles to increase their pathogenesis, whilst others play a more defensive role in the host system. For example, the sequences of the Scl1 and Scl2 proteins of *S.pyrogenes* indicates these proteins anchor to the cell surface and bind a variety of host proteins to evade phagocytosis (Han, Caswell, *et al.*, 2006; Han, Zwiefka, *et al.*, 2006; Caswell, Barczyk, *et al.*, 2008; Caswell, Han, *et al.*, 2008; Gao *et al.*, 2010; Reuter *et al.*, 2010). The Scl1 collagenous domain is capable of mimicking mammalian collagens to interact with collagen receptor integrins to facilitate adherence to host cells and trigger intracellular signalling (Humtsoe *et al.*, 2005; Caswell, Han, *et al.*, 2008). In contrast, BclA and BclB are structural components of the *Bacillus* exosporium, with a globular C-terminal domain at the distal end of the collagenous filaments (Sylvestre, Couture-Tosi and Mock, 2003), forming a barrier around the spore (Boydston *et al.*, 2005). Collagen-like domains have also been identified in non-pathogenic bacteria, such as *S. usitatus* and *R. palustris,* commonly found in soil (Xu *et al.*, 2010). However, the triple helix structure is not distinctly different between pathogenic and non-pathogenic bacteria (Xu *et al.*, 2010) and the triple helical domains vary in size (105-246 amino acids) with non-collagenous flanking domains also differing greatly (Yu *et al.*, 2014). The variety of domain lengths and amino acid compositions discovered in bacterial collagen-like proteins, and the similarity in stability despite lacking 4-Hyp residues, demonstrates the potential insights to be gained from their further research.

1.5.3 UBQLN2 and Oligomerization

The emergence of the triple helix motif in proteins distinct to mammalian collagen promotes an increasing opportunity of investigations into collagen-like proteins. Furthermore, the existence of stable collagen triple helices in bacteria presents a challenge to the central dogma that 4R-Hydroxyproline is the primary stabilisation mechanism. Expanding the scope of future analysis, to include human proteins fitting the bacterial collagen-like protein model, may reveal interesting roles for the triple helical motif in human health and disease. One avenue of this is explored in this study: investigating the triple helix potential of UBQLN2.

UBQLN2 contains a unique proline-rich repeat (PRR) region which is highly conserved across mammalian species of UBQLN2. Despite this high conservation,

a function for the PRR region has not yet been elucidated. Composed of 12 tandem repeats of proline followed by two other amino acids (Pro-Xaa-Yaa), it has often been named the PXX domain. However, if the region observed is expanded, a collagen-like sequence is revealed within this PRR region (Figure 1.5). Containing >25% proline content, but with no prolines located in the Yaa position of the (Gly-Xaa-Yaa)$_n$ repeat, the collagen-like sequence of UBQLN2 aligns closer to that of bacterial collagen like proteins.

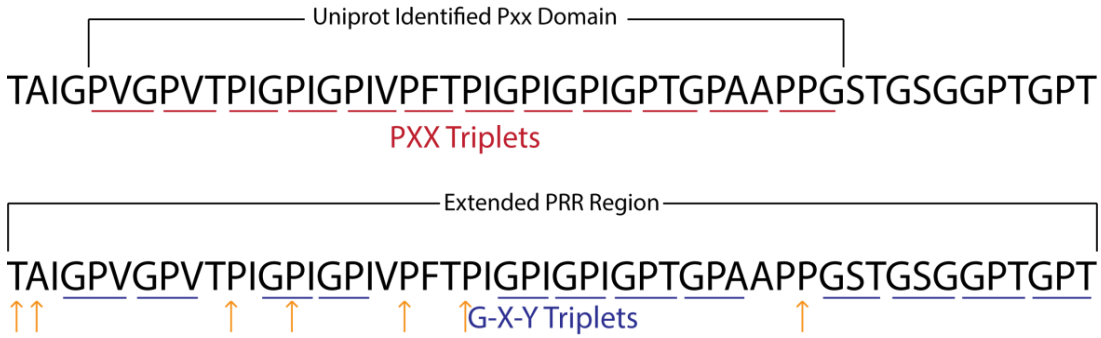


Figure 1.5 The proline rich region (PRR) of UBQLN2. Single letter amino acid code of the PRR region of UBQLN2, with the classical PXX domain identified by UniProt vs the extended bacterial collagen-like PRR region. The P-X-X triplets are highlighted in red whilst the collagen-like (Gly-Xaa-Yaa)$_n$ triplets highlighted in blue. The orange arrows denote locations of ALS-causing mutations, the majority of which fall outside the (Gly-Xaa-Yaa)$_n$ triplets.

Oligomerization via collagen-like domains occurs in numerous proteins. In C1q (Complement C1q tumour necrosis factor-related protein), three protomers trimerize to form a collagen triple helix, which then multimerizes further to form a bouquet of 18 globular domains. When expressing the globular domains of C1q alone, they exhibit differentially independent functions (Kishore *et al.*, 1998). Surfactant proteins in the lungs also increase their combined affinity through trimerization to constitute a high avidity of binding for structures found on the surface of viruses and bacteria, but not on human cells. This enables the protein to distinguish self from pathogen in a relatively simple mechanism (Watson *et al.*, 2019). A last example of oligomerization via collagen-like domains to improve binding of ligands is Mannan Binding Protein (MBP). Although the subunits of MBP exhibit limited affinity (mM range), their assembly as an oligomer provides high avidity. As a result, these proteins can selectively bind ligands with high affinity (nM to pM range) (Ogden *et al.*, 2001).

Similar to the proteins described above, UBQLN2 contains a C-terminal globular domain responsible for ligand binding, the UBA domain (Walters *et al.*, 2002). This domain is highly conserved across the UBQLNs, with one residue difference between UBQLN1 and UBQLN2 in a region unlikely to affect ubiquitin binding (Zhang, Raasi and Fushman, 2008). As such, results gathered from studies in UBQLN1 are likely similar in UBQLN2. When studied in isolation, the UBA domain of UBQLN1 does not demonstrate any chain specificity (Zhang, Raasi and Fushman, 2008). However, when expressed in the context of the full-length protein, the UBA domain of UBQLN1 preferentially binds K63-linked poly-ubiquitin chains (Harman and Monteiro, 2019). This reinforces the notion that regions outside of the UBA domains may have a role in governing ubiquitin binding. Furthermore, head-to-tail expression of four UBA domains greatly improves ubiquitin affinity (100-1000 fold), if not selectivity (Hjerpe *et al.*, 2009). This provides further indication that multiple UBA domains together can have a large impact on affinity.

1.6 iPSC models for neurodegeneration

Despite neurological disorders causing the second most fatalities globally (Feigin *et al.*, 2019), very little is understood about the causes, underlying mechanisms of disease, or how to successfully diagnose and treat them. This is largely due to the inaccessible nature of the nervous system. For example, a selection of cancers are more readily studied due to the availability of tumour tissue and the viability of culturing these cells *in vitro*. In contrast, even a healthy individual cannot afford to lose neurons, even if they were accessible. This has resulted in a heavy dependence on animal models for research into neurodegeneration. While animal models are vital tools, especially for clinical trials, they do not always exhibit the same penetrance, progression, or disease pathology as humans (Lutz, 2018).

In 2007, two independent protocols were published which revolutionised the field: the creation of human induced pluripotent stem cells (iPSC) (Takahashi *et al.*, 2007; Yu *et al.*, 2007). Using accessible somatic cells (such as fibroblasts of the skin), the authors discovered the cocktail of factors (known as the Thomson or Yamanaka factors) required to reverse the differentiation process and revert the cell to a pluripotent state. Building on this work, it is now possible to drive the

differentiation of these iPSCs into cell types of all three germ layers, such as cardiac or neural cells (Takahashi *et al.*, 2007).

Pairing this technique with CRISPR/Cas9 gene editing (Jinek *et al.*, 2012), disease-causing mutations contained within the donated patient cells can be rescued. This model will have the exact genetic background of a patient suffering with ALS, with no environmental contributions. Therefore, any conclusions drawn must be attributed to the mutation, rather from external influence (Sandoe and Eggan, 2013). This combination of technologies provides the opportunity for longitudinal study into disease progression, without invasive acquisition of patient tissue.

1.7 Thesis Aims

The original aim of this thesis was to generate an ALS-mutant iPSC line through CRISPR/Cas9 engineering, to further understand the role of ALS mutants in the development and progression of disease. A series of stress assays were planned and developed to investigate the role of UBQLN2 under various stress conditions and identify any phenotypes in a human neuronal model. Complications due to covid-19 prevented completion of these aims, but the progress made deserved acknowledgement and is outlined in Chapter 3.

Full length bacterially expressed UBLQN2 migrates and sediments as a trimer in SEC and AUC experiments (Hjerpe *et al.*, 2016). Immunoprecipitation experiments reveal 50 % more ubiquitin is pulled down in UBQLN2^{P506T} mutants when compared to wild-type, but stress-induced binding of UBQLN2^{P506T} to ubiquitylated substrates was impaired (Hjerpe *et al.*, 2016). However, the linkage type or length was not determined in this study. It is therefore possible that a loss in ubiquitin binding specificity occurs with ALS-causing mutants of UBQLN2, leading to proteasomal dysfunction and increased sensitivity to cellular stress.

Numerous proteins trimerize via a collagen-like domain to confer specificity, whilst other ubiquitin receptors (such as Rad23A/B) contain multiple UBA domains or multimerize to generate the required specificity and/or affinity for a given pathway. It is therefore logical to theorise that similar mechanisms may be employed by UBQLN2 in its function as a ubiquitin receptor. This study hypothesises external governance on ubiquitin binding is provided by the PRR

region N-terminal to the UBA domain. By trimerizing via this bacterial collagen-like PRR region, increased ubiquitin affinity may be conferred as is achieved by multimers of UBA domains (Hjerpe *et al.*, 2009), or naturally by ubiquitin receptors containing multiple UBA domains. ALS-causing mutations located within the PRR region may influence the stability of the trimer formed, resulting in a negative impact on ubiquitin binding specificity.

The aim of this thesis was to answer the following questions:

1. Does UBQLN2 trimerize via its PRR region?

2. Is the structure formed a collagen-like triple helix?

3. Does the formation of the trimer increase ubiquitin affinity or specificity of binding?

In order to answer these questions, the objectives of this thesis were as follows:

1. Create a reductionist model to investigate only the PRR region and UBA domain of UBQLN2.

2. Investigate the structures formed by these regions using biophysical approaches.

3. Elucidate the impact of these structures on ubiquitin binding.

2 Materials and Methods

2.1 Materials

2.1.1 Chemicals

Unless otherwise stated, all chemicals were purchased from Sigma-Aldrich (UK) or Thermo-Fisher Scientific (UK). ^{15}N and ^{13}C sources were purchased from Cambridge Isotope Laboratories (USA). Enzymes used for molecular biology were obtained from Promega (UK), Invitrogen (UK) or NEB (USA). Plasmid DNA purification, PCR purification and gel extraction kits were purchased from Qiagen (UK). Primers were synthesised by Biomers (Germany).

2.1.2 Buffer and media compositions

Buffer	Components
1D Gel buffer	1.5 M Aminocaproic Acid, 150 mM Bis-Tris pH 7.0
6x Agarose gel loading buffer	10 mM Tris-HCl (pH 7.6), 30 % (v/v) glycerol, 0.15 % (w/v) orange G
BN Sample Buffer (2x)	40 mM Tris pH 7.0, 3 M NaCl, 0.5 M EDTA, 20 % (v/v) glycerol
Buffer A (1x)	300 mM NaCl, 50 mM Tris pH 7.4, 5 % (v/v) glycerol
Competent cell storage buffer	100 mM $CaCl_2$, 15 % (v/v) glycerol
Dialysis Buffer	50 mM Tris pH 7.4, 150 mM NaCl, 5 % (v/v) glycerol

Buffer	Components
Elution buffer	150 mM NaCl, 50 mM Tris pH 7.4, 5 % (v/v) glycerol, 250 mM Imidazole
Equilibration buffer	300 mM NaCl, 50 mM Tris pH 7.4, 5 % (v/v) glycerol, 10 mM Imidazole
LB Media	0.1 % (w/v) tryptone, 0.1 % (w/v) NaCl, 0.05 % (w/v) yeast extract. (1.5 % (w/v) agar added for LB agar).
Minimal Media (1x)	50 mM $Na_2HPO_4.7H_2O$, 25 mM KH_2PO_4 (pH 8.0–8.2), 10 mM NaCl, 5 mM $MgSO_4$, 0.2 mM $CaCl_2$, 0.25 x Thiamine, 0.1 % $^{15}NH_4Cl$, 0.3 % Glucose or ^{13}C-Glucose.
MN-Diff Base	Advanced DMEM/F12 (Gibco), Neurobasal Medium (Gibco), 1x Pen/Strep, 2 mM Glutamax (Life Tech), 100 µM ß-mercaptoethanol, 1:50 B27 supplement (Life Tech), 1:100 N2 supplement (Life Tech), 10 µM Ascorbic Acid.
MN-NF Media	Neurobasal media (Gibco), 1x Pen/Strep, 1x MEM-NEAA (Life Tech), 2 mM Glutamax (Life Tech), 1:50 B27 Supplement (Life Tech), 1:100 N2 Supplement (Life Tech), 100 µM ß-Mercaptoethanol (Life Tech); 2.5 µM Ascorbic Acid (Sigma); 1 µM Retinoic Acid (Sigma); 10 ng ml^{-1} BDNF (R&D Systems); 10 ng ml^{-1} GDNF (R&D Systems); 10 ng ml^{-1} CNTF (R&D Systems); 10 ng ml^{-1} IGF-1 (PeproTech).
Native PAGE Dye Buffer	50 mM Aminocaproic acid, 50 mM Bis-Tris

Buffer	Components
NMR Buffer	20 mM KPi buffer pH 7.4, 150 mM NaCl, 0.01 % (v/v) NaN_3
Potassium phosphate buffer (KPi buffer) (1 M, pH 7.4)	802 mM K_2HPO_4, 198 mM KH_2
SDS-PAGE Running Buffer (20x)	50mM MOPS, 50mM Tris Base, 0.1 % SDS, 1 mM EDTA, pH 7.7
SDS-PAGE Sample Buffer (4x)	200 mM Tris-HCl (pH 6.8), 8 % (w/v) SDS, 40 % (w/v) glycerol, 0.2 % (w/v) bromophenol blue, 400 mM ß-mercaptoethanol
SDS-PAGE Transfer Buffer (10x)	25 mM Tris, 192 mM Glycine, pH 8.3
TAE Buffer (50x)	2 M Tris, 5.71% (v/v) glacial acetic acid, 50 mM EDTA.
TBS-T (1x)	200 mM Tris pH 8.1, 1.5 M NaCl, 0.1 % (v/v) Tween® 20
Wash buffer 1	150 mM NaCl, 50 mM Tris pH 7.4, 5 % (v/v) glycerol, 15 mM Imidazole
Wash buffer 2	150 mM NaCl, 50 mM Tris pH 7.4, 5 % (v/v) glycerol, 25 mM Imidazole
Wash buffer 3	150 mM NaCl, 50 mM Tris pH 7.4, 5 % (v/v) glycerol, 40 mM Imidazole

2.1.3 Plasmids

All plasmid vectors used in this study are listed below. The UBQLN2 sequence was obtained from UniProt (Q9UHD9) and foldon/CC-pII sequences obtained from their respective references in text.

Table 2.1 List of plasmids used in this study.

Insert	Plasmid	Source	Use
UBQLN2 PRR-UBA	pET-28a(+)-TEV	GenScript (designed for this study)	Protein Purification (Chapter 4)
UBQLN2 UBA-Linker	pET-28a(+)-TEV	GenScript (designed for this study)	Protein Purification (Chapter 4)
UBQLN2 UBA	pET-28a(+)-TEV	GenScript (designed for this study)	Protein Purification (Chapter 4)
UBQLN2 Foldon PRR-UBA	pET-28a(+)	GenScript (designed for this study)	Protein Purification (Chapter 5)
UBQLN2 CCpII PRR-UBA	pET-28a(+)	GenScript (designed for this study)	Protein Purification (Chapter 5)
UBQLN2 PRR-UBA	prsf-DUET-1	Merck Millipore	SPR (Chapter 7)
UBQLN2 UBA	prsf-DUET-1	Merck Millipore	SPR (Chapter 7)
CRISPR/Cas9 UBQLN2 gRNAs	pSpCas9(BB)-2A-Puro	Addgene	CRISPR/Cas9 (Chapter 3)

2.1.4 Primers

Primers were purchased from Biomers and stored in ddH_2O at 100µM concentration at -20°C. gRNAs for CRISPR/Cas9 are outlined in the chapter.

Table 2.2 List of primers used in this study.

Name	Primer Sequence 5'-3'	Description
Fwd_UQ2_P506T	CCATAGGTCCTATAGTCACTTTTACCCCCATAGGC	Forward primer to introduce P506T mutation into recombinant UBQLN2
Rev_UQ2_P506T	TGGGGGTAAAAGTGACTATAGGACCT	Reverse primer to introduce P506T mutation into recombinant UBQLN2
L619A_Fwd	GCCATTGAAAGGCTGGCGGGCTCCCAGCCATC	Forward primer to introduce L619A mutation into recombinant UBQLN2
L619A_Rev	GATGGCTGGGAGCCCGCCAGCCTTTCAATGGC	Reverse primer to introduce L619A mutation into recombinant UBQLN2
F594V_Fwd	GCTCAACGCAATGGGGGTCTTAAACCGTGAAGC	Forward primer to introduce F594V mutation into recombinant UBQLN2
F594V_Rev	GCTTCACGGTTTAAGACCCCCATTGCGTTGAGC	Reverse primer to introduce F594V mutation into recombinant UBQLN2

Name	Primer Sequence 5'-3'	Description
n996_UBQLN2_487-C	TAAGCAGGATCCGGTGGCTCTACCGCTATAGGCCCTGTAGGCCC	Forward primer to clone PRR-UBA into the modified pRSF-DUET1 for SPR analysis.
n997_UBQLN2_579-C	TAAGCAGGATCCGGTGGCTCTGAAGTCAGATTTCAGCAACAACTGG	Forward primer to clone UBA into the modified pRSF-DUET1 for SPR
n998_UBQLN2	TGCTTAGAATTCTTAAGATGGCTGGGAGCCCAG	Reverse primer to clone PRR-UBA into the modified pRSF-DUET1 for SPR analysis.
UBA_rev	TCTCCAAGCTTTTACGATGGCTGGGAGCCCAG	For sequencing from the C-term of
Frw_UQ2_P506T	CCATAGGTCCTATAGTCACTTTTACCCCCATAGGC	Site-directed mutagenesis of recombinant UBQLN2
Rev_UQ2_P506T	GCCTATGGGGGTAAAAGTGACTATAGGACCTATGG	Site-directed mutagenesis of recombinant UBQLN2
Repair_P506T	GATTCCGAGCTTCACTCCAGGTGTGGGGGTGGGGGTGCTGGGAACCGCTATAGGCCCTGTAGGCCCAGTCACTCACATAGGCCCCATAGGACCTATAGTCACTTTTACCCCCATAGGCCCCATTGGGCCCATAGGACCCACTGGCCCTGCAGCCCCCCCTGGCTCCACCGG	Repair template to introduce the P506T mutation into iPSCs
Repair_P497H	CATTAGCCACTGAAGCACCTGGCCTGATTCCGAGCTTCACTCCAGGTGTGGGGGTGGGGGTGCTGGGAACCGCTATAGGCCCTGTAGGCCCAGTCACTCACATAGGCCCCATAGGCCCTATAGTCCCTTTTACCCCCATAGGCCCCATTGGGCCCATAGGACCCACTGGCCC	Repair template to introduce the P497H mutation into iPSCs

Name	Primer Sequence 5’-3’	Description
UBQLN2_fwd-1	AAACCCAAGAGCAATGCAGG	Sequencing UBQLN2 in human cells
UBQLN2_rev-1	ATGTCGCCTCCTGTTGCTAT	Sequencing UBQLN2 in human cells
UBQLN2_fwd-2	CGCTAATTATGTCGCCAGCA	Sequencing UBQLN2 in human cells
UBQLN2_Rev-2	AAGAACCCCATTGCGTTGAG	Sequencing UBQLN2 in human cells

2.1.5 Bacterial Strains and Media

XL1 Blue Dh5α *Escherichia coli* (*E.coli)* cells were used for plasmid propagation and cloning. For recombinant protein expression and purification, Rosetta 2 BL21 DE3 *E.coli* cells were used.

2.1.6 Antibodies

Primary antibodies used in this study are listed in the below table. A variety of suitable Alexa fluorophores were used as secondary antibodies for immunofluorescence and HRP-conjugate secondary antibodies for western blotting. All secondary antibodies were used at 1:1000. As TRA-1-60 is a surface stain, block then staining with TRA-1-60 needs to occur before washing and permeabilising to continue the protocol as normal.

Table 2.3 Antibodies list.

Antibody	Species	Concentration	Company
His	Mouse (IgG)	1:5000	BioRad
UBQLN2 S261D	Sheep	1:5000	In House
UBQLN2 (2H9)	Mouse (IgG1)	1:1000	Novus
Nanog	Rabbit	1:800	Cell Signalling
TRA-1-60	Mouse (IgM)	1:100	Santa Cruz
Oct 3/4	Mouse (IgG2b)	1:250	Santa Cruz
Neurofilament H	Chicken	1:1000	Bio legend
Islet 1/2	Mouse (IgG2b)	1:50	DSHB
Caspase 3	Rabbit	1:200	Millipore

2.2 Methods

2.2.1 Computational Modelling Parameters

2.2.1.1 BLAST/PRALINE parameters

The NCBI programme was utilised for BLAST analysis (https://blast.ncbi.nlm.nih.gov/Blast.cgi). The PRR region (residues 487-537) was compared to the non-redundant protein sequence database, with either all organisms included or restricting to human only. The protein-protein BLAST algorithm was used displaying 500 target sequences with a word size of 6 to identify matches. All other parameters were kept as standard recommendations. PRALINE alignment (https://www.ibi.vu.nl/programs/pralinewww/) was carried out using the standard recommended parameters.

2.2.1.2 AlphaFold analysis

The simplified version of AlphaFold v2 1.0 (Jumper *et al.*, 2021) was accessed through the Google Colab page https://colab.research.google.com/github/deepmind/alphafold/blob/main/notebooks/AlphaFold.ipynb. This notebook uses no templates and a selected portion of the BFD database, which generates the potential for a drop in the accuracy of prediction for a small number of targets. The protein sequence of interest was entered as a single sequence for monomeric predictions or three times for homo-trimeric predictions.

2.2.1.3 Chimera X analysis

PDB files generated from AlphaFold analysis were imported into Chimera X. Hydrogen bonds were investigated using the both the default parameters (0.4 Å and 20°) and adjusted to the maximum tolerances (3.5 Å and 155°) (Benco *et al.*, 2001) for distance and angle tolerances respectively using the structural analysis tool.

2.2.2 Microbiology techniques

2.2.2.1 Preparation of chemically competent *E.coli*

A 100 µl aliquot of chemically competent Top-10 *E.coli* cells was grown in 5 ml of LB media supplemented with 10 µg ml^{-1} streptomycin shaking at 185 rpm overnight (~16 hours) at 37ºC. 500 ml of streptomycin supplemented LB was inoculated with 1 ml of this overnight culture and grown at 37ºC whilst shaking to an optical density at 600 nm (OD_{600}) 0.3-0.4. Cultures were transferred to sterile 50 ml falcon tubes and pellet by centrifugation at 1000 g for 10 minutes at 4ºC. The supernatant was discarded and the pellets re-suspended and washed thoroughly with ice-cold 100 mM $MgCl_2$ before being re-pelleted by centrifugation. The cells were resuspended and washed in ice-cold 100 mM $CaCl_2$ before a final centrifugation. Pellets containing chemically competent cells were carefully resuspended in competent cell storage buffer, aliquoted into 50 µl aliquots, snap frozen in liquid N_2 and stored at -80ºC.

2.2.2.2 Preparation of LB plates

LB-Agar (1.5 % (w/v) was autoclaved at 121ºC for 20 minutes and allowed to cool to 55ºC before the addition of kanamycin to a final concentration of 100 µg ml^{-1}. Plates were then poured under sterile conditions.

2.2.2.3 Transformation of competent *E.coli*

A 50 µl aliquot of competent *E.coli* cells was thawed on ice and supplemented with ~100 ng of the relevant pDNA or 5 µl ligation reaction. Suspensions were gently mixed and incubated on ice for 20 minutes, followed by a 45-second heat shock at 42ºC. Cells were returned to ice for 2 minutes before the addition of 450 µl antibiotic-free LB media and recovered at 37ºC for 45 minutes with gentle shaking. Cells were pelleted at 1000 g, 400 µl supernatant removed and replenished with 100 µl fresh antibiotic free media. A suitable volume of transformed cells was then spread on kanamycin-supplemented LB-agar plates and incubated at 37ºC overnight.

2.2.3 Molecular Biology Assays

2.2.3.1 Plasmid DNA Purification and Sequencing

5 ml of LB media supplemented with 100 µg ml^{-1} kanamycin was inoculated with a single colony picked from an *E.coli* transformation plate. This culture was incubated overnight at 37ºC, 185 rpm. Cells were pelleted the next morning by centrifugation at 4000 g for 15 minutes at room temperature. Plasmid DNA was isolated using the QiaPrep® Spin Miniprep Kit (Qiagen) as per the manufacturer's instructions. Purified pDNA concentration was measured using a Nanodrop Spectrophotometer (Thermo) and stored at -20ºC.

Any PCR fragments or plasmid DNA were sent for Sanger sequencing on ABI 3730xl platforms (DNA Sequencing Service, University of Dundee). Sequencing results were analysed using SnapGene and COBALT.

2.2.3.2 PCR

The primers catalogued in Table 2.2 were used to amplify the respective genes outlined in the descriptions. Primers were designed to contain appropriate restriction sites and start and stop codons where necessary. All reactions were carried out in a T100 Thermal Cycler (BioRad). Plasmid DNA was used as a template for all PCR reactions. The PCR reactions contained 1x Polymerase buffer, 200 µM dNTPs, 0.5 µM of each primer (this was increased to 1 µM for Site Directed Mutagenesis), 2 % (v/v) DMSO, 100 ng template pDNA and 0.05 u $µl^{-1}$ polymerase. For low fidelity reactions such as Colony PCR, Taq Polymerase was used. In Site Directed Mutagenesis, Pfu Polymerase (Promega) was utilised whilst Q5 DNA Polymerase (NEB) was used for any other amplifications. Typical thermocycling conditions for the PCR reactions are as follows: Initial denaturation (98˚C for 30 seconds) followed by 35 cycles of denaturation (98˚C for 10 seconds), annealing (50-72˚C for 30 seconds) and extension (72˚C for 30 seconds/kb) and a final extension at 72˚C for 10 minutes. The annealing temperature of ~5ºC below the salt-adjusted melting temperature was generally chosen.

2.2.3.3 Agarose gel electrophoresis

Agarose gel solution (1 % (w/v) Agarose, 1x TAE Buffer, 1x Sybr® Safe) was prepared by dissolving agarose powder in 1x TAE buffer and heating in a microwave. Sybr® Safe was added to a final 1x concentration and the gels poured into a mould containing a loading comb. Once set, the loading comb was removed, and the gel was submerged in 1x TAE buffer within the gel tank. 6x agarose gel loading buffer was added to DNA samples prior to loading into the gel at an appropriate volume. If multiple combs were used on the same gel, the gel was run for a few minutes after each fully loaded row, to prevent the sample escaping out of the well. Electrophoresis was carried out at 100 V for 1 hour to separate the DNA fragments, using either a 100 bp or 1 kb DNA ladder (Promega) to estimate fragment size. Gels were visualised on a UV transilluminator (BioRad).

2.2.3.4 Gel Extraction

If required, fragments were extracted from the gel using the QiaQuick® Gel Extraction Kit (Qiagen) and/or cleaned up using the QiaQuick® PCR Clean Up Kit (Qiagen), following the manufacturer's instructions in both instances. Resulting DNA concentrations were measured on a NanoDrop Spectrophotometer (Thermo) and DNA products stored at -20ºC.

2.2.3.5 Restriction endonuclease digestion

Restriction digests of pDNA or PCR products were set up as outlined in Table 2.4 and incubated for 1 hour at 37ºC to ensure complete digestion. Reactions were cleaned up using the QiaQuick® PCR Clean Up Kit (Qiagen), following the manufacturer's instructions.

Table 2.4 Restriction endonuclease digestion reaction setup.

Component	Stock concentration	Volume (µl)
DNA	Variable	1-25
Restriction enzyme(s)	20 units μl^{-1}	1 each
CutSmart buffer	10x	3
dH_2O	-	0-25
Total		30

2.2.3.6 T4 ligase reactions

Restirction digested DNA (either pDNA or purified PCR fragments) were placed into ligation reactions to allow insertion of the gene of interest downstream of the promoter sequence. Reactions were incubated for 1 hour at 25ºC. The standard reaction setup is outlined in Table 2.5. 5 µl of completed ligation reaction was used to transform competent Dh5α cells as described previously (2.2.2.3). Single colonies were selected and grown overnight for pDNA purification which was subsequently sent for DNA sequencing (see section 2.2.3.1) to confirm correct assembly.

Table 2.5 T4 ligation reaction setup.

Component	Stock Concentration	Volume (µl)
Linearised pDNA	50 ng μl^{-1}	1
Insert DNA	Variable	Variable
T4 DNA ligase buffer	10x	2
T4 DNA ligase	400 units μl^{-1}	1
dH_2O	-	Up to 20
Total		20

2.2.3.7 Colony PCR

Samples were prepared by picking the desired colony from a plate with a P200 tip and mixing with the pre-prepared PCR solution (outlined in section 2.2.3.2). Samples were run on a 1 % agarose gel (section 2.2.3.3) and visualised on a UV transilluminator.

2.2.3.8 Site-Directed Mutagenesis

Overlapping primers for site-directed mutagenesis were designed using Agilent's QuikChange primer design tool. Using 1 ng ul^{-1} of pDNA containing the region of interest to be mutated, forward and reverse primers containing the mutation were used to amplify the mutant plasmid. PFU polymerase (Promega) was used as specified in 2.2.3.2 to amplify desired products. 2 µl of the restriction enzyme DPN1 (Promega) was used for 2 hours at 37°C to digest any remaining parental template. 5 µl of Dpn1 treated reaction was then used to transform Dh5a cells, with successful transformants grown for pDNA isolation and DNA sequencing analysis as outlined previously.

2.2.4 Biochemical Methods

2.2.4.1 Recombinant Protein Expression

Plasmids containing the gene of interest (outlined in Table 2.1) were transformed into BL21 Rosetta 2 competent cells and grown overnight on LB agar plates supplemented with 100 µg/ml kanamycin. Three colonies from each plate were grown in a small scale (5 ml) liquid culture for 4 hours, induced with 0.1 mM Isopropyl ß-D-1-thiogalactopyranoside (IPTG) and incubated overnight at 25ºC to confirm protein expression. Bacterial pellets before and after IPTG induction were separated via SDS-PAGE and stained with Instant Blue to determine which colony expressed the most protein. The highest expressing colony was selected for large scale purifications, where 2 L of inoculated LB were grown at 37ºC to an optical density 600 (OD_{600}) of 3-5. Once this OD was reached, the temperature was reduced to 16ºC for one hour and cultures induced with 0.1 mM IPTG and incubated overnight at 16ºC and 185 rpm shaking. For experiments involving isotopically labelled proteins, the medium was switched to minimal medium (see section

2.1.2) 1 hour prior to induction. Cells were harvested via centrifugation using a JLA8.1000 rotor (Beckman Coulter) at 4000 g and stored at -20ºC until time of use.

2.2.4.2 Recombinant Protein Purification

Cell pellets were resuspended in buffer A (see section 2.1.2), 1 mM phenylmethylsulfonylfluoride (PMSF) and 2 mM $MgCl_2$. The cell solution was lysed by passing through a French Press three times at 750 psi and subsequently clarified by centrifugation at 21000g for 35 minutes at 4ºC to remove any unbroken cells or cellular debris. The resulting supernatant was passed through a 20 G needle and 0.2 µM filter to sheer any DNA and remove any remaining contaminants. 5 mM imidazole was added to the filtered supernatant to prevent non-specific protein binding. The protein solution was loaded onto a column containing Ni-NTA beads pre-equilibrated with equilibration buffer. The beads were then washed once with wash buffers 1, 2, and 3. Remaining bound protein was eluted in elution buffer. The imidazole was removed by three rounds of dialysis into dialysis buffer for downstream use.

2.2.4.3 Protein concentration measurement

Protein concentrations were determined using a Jasco V-550 UV-vis spectrophotometer, accounting for molecular weight and extinction co-efficient predicted by the amino acid sequence, the information of which was obtained using the tool *ProtParam*.

2.2.4.4 Cleavage with TEV

The tobacco etch virus (TEV) protease (kindly provided by the lab of Professor Walden) was used to cleave the His_6 tag from the His_6-PRR-UBA and His_6-UBA only proteins. This was achieved with a 1:1 ratio of TEV protease: protein in solution (150 mM NaCl, 50 mM Tris, 5 % (v/v) Glycerol, pH 7.4) for 48 hours at 4ºC, with the completed reaction passed over an equilibrated Ni-NTA column for clean-up. The products were run on an SDS-PAGE gel and visualised with Instant Blue Coomassie stain.

2.2.4.5 Cleavage with Thrombin

Restriction grade Thrombin was used to cleave the His_6 tag from the HFPRR and HCPRR proteins in NMR buffer (20 mM KPi, 150 mM NaCl, 0.01 % (v/v) NaN_3, pH 7.4). This was achieved with a 3:1 ratio of protein: thrombin for 6 hours at 22 ºC. The completed reaction was passed over an equilibrated SEC SD200i column for clean-up and cleavage verified by SDS-PAGE and Instant blue staining.

2.2.4.6 SDS-PAGE

Bis-Tris SDS PAGE was used to separate proteins by their molecular weight. 11 % or 15 % acrylamide resolving gels were cast (depending on the molecular weight of the protein of interest) in BioRad gel casting cassettes using 1.5 mm glass plates and set at room temperature. After setting, resolving gels were overlaid with stacking gel containing a sample comb and again left to set at room temperature. Samples were prepared in 4x SDS Sample Buffer and boiled at 95ºC for 5 minutes prior to loading into the well. Samples were run alongside the Precision Plus Protein™ All Blue Pre-stained Protein Standards ladder. The gels were run at 100 V for 60-90 minutes or until the dye front escaped the gel in a 1x MOPS running buffer (50 mM MOPS, 50 mM Tris Base, 0.1% SDS, 1 mM EDTA, pH 7.7). Upon completion, the stacking gel was discarded and the gels were either stained with InstantBlue® Coomassie Protein Stain (Abcam) overnight, or transferred onto a 0.4 µM nitrocellulose for western blotting. A table displaying the Bis-Tris SDS-PAGE gel component can be found below.

Table 2.6 Recipe for casting SDS PAGE gels.

Component	Stacking	Resolving		
Acrylamide	4 %	8 %	11 %	15 %
Acryl:Bisacrylamide (40%)	1 ml	2 ml	2.75 ml	3.75 ml
Bis-Tris (1.2 M stock)	3 ml	3 ml	3 ml	3 ml
ddH_2O	6 ml	5 ml	4.25 ml	3.25 ml
APS (10 %)	100 µl	100 µl	100 µl	100 µl
TEMED (100 %)	6 µl	6 µl	6 µl	6 µl
Total	10.16 ml	10.16 ml	10.16 ml	10.16 ml

2.2.4.7 Transfer and western blot analysis

Transfer of the SDS-PAGE gels onto 0.4 µM nitrocellulose membrane was achieved using a wet transfer system (Bio-Rad) in 1x transfer buffer supplemented with 20 % Methanol. Transfers were completed at 70 V for 60 minutes for proteins 50 kDa and below, and 70 V for 90 minutes for proteins >50 kDa. Following transfer, the membrane was blocked in 5 % Marvel skimmed milk in TBST for 1 hour at RT with gentle agitation. The membrane was subsequently incubated in primary antibody (see Table 2.3) overnight at 4ºC with rocking. The membrane was then washed in 1 x TBST at RT for 10 minutes under gentle agitation for a total of three washes. Secondary antibody incubation (see Table 2.3) was then carried out with 5 % Marvel skimmed milk in TBST for 1 hour at RT whilst rocking. A final three TBST washes were carried out prior to imaging of the membrane. Chemiluminescence solution (Pierce) and ChemiDoc XRS (BioRad) machine were used to image the membranes.

2.2.4.8 Clear-Native PAGE

The protocol for Clear-Native PAGE was taken and adapted from Dr Erika Fernandez-Vizara (Fernandez-Vizarra and Zeviani, 2021). The Bio-Rad mini-gel casting system was utilised, and gels were cast at 4°C under constant stirring on a stirring plate using a gradient mixer. 4-16% gradient gels were cast to visualise the various states of the recombinant protein and Table 2.7 describes the reagent volumes for each percentage gel. The gels were run at 4°C in a lower (Anode) buffer consisting of 50 mM bis-Tris/HCl pH 7.0 and upper (Cathode) buffer consisting of 50mM Tricine, 15mM Bis-Tris/HCl pH 7.0. Initially the gel was run at 90V for 30 minutes, at which point the constant was altered to 12mA (limiting 300V) until the dye front reached the bottom of the gel.

Table 2.7 Native PAGE gel recipe.

Component	Stacking 4%	4%	16%
Acrylamide Mix (49% Acryl:Bisacrylamide 48:1.5)	0.25 ml	0.48 ml	1.28 ml
1D Gel Buffer 3x	1 ml	2 ml	1.33 ml
Ultrapure ddH_2O	1.75 ml	3.48 ml	0.56 ml
Glycerol	--	--	0.72 ml
10% APS	25 µl	40 µl	14 µl
TEMED	3 µl	4 µl	3 µl
Total	3.3 ml	6 ml	4 ml

2.2.4.9 Silver Staining

The gel was silver stained following electrophoresis (either SDS-PAGE or Native PAGE) using the Thermo Silver Stain kit (24612/ Pierce™ Silver Stain Kit), following the manufacturer's instructions. Once in stop solution, the gels were imaged using the ChemiDoc XRS (BioRad) under white epiluminescence.

2.2.4.10 Ubiquitin binding analysis (NMR)

Monoubiquitin (Ub) and K48-linked diubiquitin (K48-Ub_2) were provided courtesy of Dr Mark Nakasone of the Beatson Institute. Diubiquitin was distally labelled with ^{15}N. Samples were prepared as outlined in the chapter and binding was monitored through ^{15}N-HSQCs acquisition, as outlined in section 2.2.5.4. The deposited assignments for the structure of Ubiquitin were imported from BMRB 16228 (Ikeya *et al.*, 2009). 200 µM of ^{15}N-HFPRR, 100 µM ^{15}N-Ub, or 50 µM ^{15}N-K48-Ub_2 were used as the observed macromolecule with ligand added to a 2, 5, or 2.7 molar ratio for each respective macromolecule. The observed CSP at any titration point can be described as $\Delta\delta obs = \Delta\delta\max * [L]([L] + Kd)$ where $\Delta\delta\max$ is the difference in chemical shift between the free and fully bound states for that given amide resonance, K_d is the dissociation constant, and [L] is the molar concentration of the free ligand. The measured chemical shifts, scaled by 0.15 in

the nitrogen dimension, were fit to a hyperbolic saturation curve with covariance method (Williamson, 2013) with the following equation:

$$\Delta\delta_{obs} = \Delta\delta_{max}\left\{([P]_t + [L]_t + K_d) - [([P]_t + [L]_t + K_d)^2 - 4[P]_t[L]_t]^{1/2}\right\}/2[P]_t$$

where $\Delta\delta_{obs}$ is the change in the observed shift from the unbound state, $\Delta\delta_{max}$ is the maximum shift change detected on saturation, and total concentrations of ligand and observed protein/macromolecule are $[L]_t$ and $[P]_t$ respectively. A value for K_d was fitted from the equation above using measured values of chemical shift changes at varying concentrations in the CcpNMR v2.4.2 (Vranken *et al.*, 2005) software assuming a single-site binding model. Reported K_d values were mean averages of residue-specific apparent K_d values, where the errors represent the standard deviation of these values.

2.2.4.11 Stress Assays

iPSC-derived MNs were plated in a 24 well plate containing glass coverslips at cell densities of ~30,000 cells per coverslip. To undergo heat-shock, plates were placed in a benchtop incubator at 42°C for 2 hours. Plates were then fixed in 4 % PFA or placed back in 37°C to recover for either 24 or 48 hours. Tunicamycin was used at either 2.5, 5 or 10 µg ml^{-1} to induce endoplasmic reticulum stress. For vehicle control, DMSO alone was used. Cells remained in culture with Tunicamycin for either 24 or 48 hours at 37°C 5 % CO_2.

2.2.4.12 Immunofluorescence

Neurons were cultured at 30,000 cells per glass coverslip. Cells were fixed in 4 % Paraformaldehyde (PFA) and permeabilised in 0.2 % Triton-X-100-PBS at room temperature for 10 minutes. Cells were subsequently blocked in 6 % Goat Serum and 0.2 % Triton-X-100- PBS for 1 hour at room temperature. Primary antibody staining was performed overnight at 4°C and secondary antibody staining performed at room temperature for 1 hour, with 3x PBS washes between stages. As TRA-1-60 is a surface stain, block then staining with TRA-1-60 occurred before washing and permeabilising, continuing the protocol as normal. Nuclei were counterstained with 1 µg ml^{-1} Hoechst after the secondary antibody staining prior to mounting on slides. Coverslips were sealed in place using nail varnish around

the edge. Images were acquired on the Zeiss Observer microscope or on the confocal microscope (see section 2.2.4.13) and images processed using Fiji v1.52i.

2.2.4.13 Confocal microscopy

All experiments were performed using a Zeiss LSM 880 confocal microscope equipped with a 20 x and 40 x objective and a 63 x oil immersion objective. Samples were cultured on glass coverslips and mounted onto slides following immunofluorescence staining (see section 2.2.4.12 and Table 2.3). Image analysis was performed using Fiji v1.52i.

2.2.4.14 Definiens Analysis

To confirm the pluripotency of iPSC colonies and efficiency of the conversion protocol into motor neurons, images were analysed using Definiens Developer XD (Definiens Developer XD®, Munich, Germany). Nuclei were first detected using a fluorescence threshold to separate background pixels from Hoechst-stained objects. Any holes in objects were filled and the objects were smoothened by shrinking and growing each object by 2 pixels. Objects were then classified as nuclei based on their size (excluding small and large objects) and their elliptical fit (only objects with an elliptical fit greater than 0.8 were classified as nuclei). The nuclei were then expanded by 20 pixels and within this area a fluorescence threshold (at 2 times the corresponding background) was applied to identify the pluripotency, neuronal and motor neuron markers listed in Table 3.1.

2.2.4.15 Electrophysiology

Whole-cell current clamp experiments were carried out following a standard protocol outlined in (Johnson et al., 2007). To summarise, sequential steps of increasing current were introduced until action potential response was recorded, expressed as voltage (mV) over time (ms). Experiment was carried out on MNs 3 weeks post dissociation. MNs were 3 weeks post dissociation (~5 weeks total in age).

2.2.5 Biophysical Methods

2.2.5.1 Fast Protein Liquid Chromatography

Fast Protein Liquid Chromatography (FPLC) was employed to both enhance protein purity for downstream applications and as an experiment in its own right. Samples were first concentrated using a Vivaspin 20 ultrafiltration unit (MW 3000-10,000 depending on protein size) (Sartorius) at 3000g at 4ºC. Centrifugation of the concentrated protein at 17,000g for 10 minutes removed any aggregates or precipitates. The AKTA FPLC (GE Healthcare) apparatus was used in conjunction with an SD200i 10_300 size exclusion chromatography (SEC) column. 500 µl of concentrated protein was injected into the pre-equilibrated column and fractionated by approximate molecular weight on the SD200i column. The system was run at 0.5 ml/min to avoid elevated column pressure. Following separation, each fraction was analysed via SDS-PAGE and instant blue Coomassie staining. Two sets of protein standards (Set 1: Aldolase, Ovalbumin, Ribonuclease A; Set 2: Conalbumin, Carbonic Anhydrase, Aprotinin) were run on the column to generate a calibration curve following the manufacturers instructions (GE Healthcare).

2.2.5.2 Circular Dichroism

CD spectra were measured using a J-810 spectropolarimeter, with the support of Ms June Southall of the Structural Biology and Biophysical Characterisation Facility. Bandwidth was set to 1 nm and 20 nm/min continuous scanning speed with a 1 sec response was used throughout the experiments. Far-UV (260-190 nm) CD spectra measurements were taken 0.01 cm pathlength cuvette. All measurements were acquired at 20°C and carried out in 20 mM phosphate buffer and 150 mM NaCl. The concentrations of samples measured were as follows: His_6-UBA (62 µM), Monomeric His_6-PRR-UBA (13 µM), Multimeric His_6-PRR-UBA (6 µM), HFPRR WT (14 µM). Four consecutive scans were averaged to obtain the resulting spectra. The ellipticity signal was buffer corrected and normalised to Mean Residue Ellipticity (deg cm^2 d mol^{-1})(Wallace, BA, & Janes, 2010). The spectra were fitted with the CDSSTR, SELCON3 and CONTIN (Compton and Johnson, 1986; van Stokkum *et al.*, 1990; Sreerama, Venyaminov and Woody, 1999; Sreerama and Woody, 2000) programmes using reference set 7 (Sreerama and Woody, 2000) employed in www.dichroweb.cryst.bbk.ac.uk/ (Whitmore and Wallace, 2004).

2.2.5.3 Analytical Ultracentrifugation

Analytical Ultracentrifugation (AUC) experiments were performed on an Optima XL-I analytical ultracentrifuge (Beckman Coulter, Palo Alto, CA, USA), with the support and guidance of Dr Mads Gabrielsen of the Structural Biology and Biophysical Characterisation Facility. All experiments were completed at 4ºC, Sedimentation Velocity (SV) experiments with a rotor speed of 49,000 rpm and Sedimentation Equilibrium (SE) experiments were completed at 20,000 rpm. In SV experiments, 360 µl of varying concentration of protein were loaded into double centrepieces. 90 µl of the same samples (if applicable) were loaded for SE. Data were acquired with absorbance optics at 235 nm for His_6-PRR-UBA protein and 280nm for HFPRR, with scans acquired every 5 minutes. Data analysis was completed using SEDFIT ((Schuck, 2000). Partial specific volume of His_6-PRR-UBA (0.72580 / 0.72629), buffer density (1.0074 / 1.0092) and viscosity (0.0102329 / 0.0160024) were calculated at 4 °C and 20 °C respectively, using SEDNTERP3 (jphil.mailway.com). Partial specific volume of HFPRR (0.72320 / 0.7300), buffer density (1.0076 / 0.998234) and viscosity (0.01567 / 0.01002) were calculated at 4 °C and 20 °C respectively. Figures were prepared in GUSSI (Brautigam, 2015). To determine the mass of each species from the SV data, the c(s) distribution was converted to a distribution of molar masses (c(M) distribution). Each peak on the distribution plot was integrated to obtain the weight-averaged values for sedimentation coefficient and molecular mass estimations.

2.2.5.4 Surface Plasmon Resonance (SPR)

SPR experimentation was conducted under the supervision of Dr Mark Naksone. SPR binding experiments were performed at 25 °C on a Biacore T200 (GE Healthcare) using a Ni-NTA sensor chip conditioned with $NiCl_2$ according to the manufacturer's instructions. Prior to SPR, each His_{12}-UBQLN2 variant and Ub analyte were exchanged into SPR running buffer (25 mM HEPES, 150 mM NaCl, and 0.005% (v/v) Tween-20) using a 0.5 ml Zebra desalting column. His_{12}-UBQLN2 variants were captured on the Ni-NTA sensor chip and the resonance units (RU) were stable following washing. Serial dilutions of mono-Ub or K48-Ub_2 were applied in two technical replicates. The SPR binding data were processed with BIAevalution (GE Healthcare) and fit to a one-site binding model using Prism (GraphPad).

2.2.5.5 Nuclear Magnetic Resonance Spectroscopy

NMR experiments were carried out on a Bruker AVANCE IIIHD 600 MHz spectrometer equipped with an Ultrashield™ Plus magnet and TCI triple resonance cryoprobe. Standard Bruker pulse programmes were run using Bruker's Topsin™ v3.2 software.

All experiments were conducted using either ^{15}N-labelled or [^{15}N, ^{13}C]-labelled protein produced in minimal media as described in section 2.1.2. All experiments were run at 310 K unless otherwise stated. Protein samples were prepared in NMR buffer (see section 2.1.2) and deuterium oxide (D_2O) added to a final concentration of 5 % in a 600 µl sample. Particulates were removed by centrifugation at 17,000 g in a microcentrifuge prior to pipetting into a 5 mm NMR tube (Wilmad 535-PP-7).

The acquisition parameters for the multidimensional experiments used in this study are outlined in Table 2.8. Data were processed using Bruker's TopspinTM v3.2 and assignment analysis carried out on the CcpNMR analysis v2.4.2 software (Vranken *et al.*, 2005).

Dihedral angle predictions were calculated with the DANGLE algorithm, which was applied to the data from within the CcpNMR v2.4.2 software. Outputs were extracted and plotted in GraphPad Prism v7.

Table 2.8 NMR acquisition parameters. NMR experiments and the corresponding acquisition parameters and references. NC = Nucleus, SW = Spectral Width (ppm), AQ = Acquisition Time (ms), QD = Quadrature Detection Mode: ST = States-TPPI, NUS = Non-Uniform Sampling Amount (%), NS = Number of Scans, Temp = Temperature (K).

Experiment	F3				F2				F1			NUS	NS	Temp	Pulse Programme Reference
	NC	SW	AQ	QD	NC	SW	AQ	Q	NC	SW	AQ				
^{15}N-HSQC	-	-	-	-	^{1}H	15.97	106.91	ST	^{15}N	24.00	87.75	-	2	310	(Mori *et al.*, 1995)
HNCA	^{1}H	14.04	121.65	ST	^{15}N	24.00	24.00	ST	^{13}C	30.00	14.15	25	24	310	(Grzesiek and Bax, 1992; Schleucher, Sattler and Griesinger, 1993; Kay, Xu and Yamazaki, 1994)
HNCACB	^{1}H	15.79	108.13	ST	^{15}N	24.00	24.68	ST	^{13}C	80.00	6.47	28.49	16	310	(Wittekind *et al.*, 1993; Muhandiram and Kay, 1994)
CBCACONH	^{1}H	15.79	108.13	ST	^{15}N	24.00	24.68	ST	^{13}C	80.00	6.47	28.49	16	310	(Grzesiek and Bax, 1992; Muhandiram and Kay, 1994)
HNCO	^{1}H	14.04	121.65	ST	^{15}N	24.00	24.00	ST	^{13}C	14.00	52.09	22.86	8	310	(Grzesiek and Bax, 1992; Schleucher, Sattler and Griesinger, 1993; Kay, Xu and Yamazaki, 1994)
HNCACO	^{1}H	14.04	121.65	ST	^{15}N	24.00	24.00	ST	^{13}C	14.00	52.09	22.86	32	310	(Clubb, Thanabal and Wagner, 1992; Kay, Xu and Yamazaki, 1994)

2.2.6 CRISPR/Cas9 gene editing

2.2.6.1 gRNA design and cloning

The sequences encoding guide RNAs (gRNAs) were designed on Benchling as 20 bp single stranded oligonucleotide donors (ssODN) to introduce the P506T mutation into control iPSC. ssODNs were annealed together, phosphorylated with T4 PNK (following manufacturers guidance) and ligated with T4 ligase (see section 2.2.3.6) into the pSpCas9-2A-Puro plasmid vector cut with BbsI (see section 2.2.3.5). A colony PCR (see section 2.2.3.7) was completed to confirm successful cloning and samples sent for sequencing to further confirm this.

2.2.6.2 T7 endonuclease assay

The pSpCas9 vector containing the sequences encoding for the gRNAs was transfected into HEK293 cells with Lipofectamine 3000 with selection achieved with 3 μg ml^{-1} puromycin treatment for 72 hours. Successfully transfected cells were harvested using the WIZARD® genomic DNA recovery kit (Promega) and gDNA concentration measured on the Nanodrop spectrophotometer. The purified gDNA product used for the T7 endonuclease assay, set up as outlined in (2.2.3.5). Digested products were run on a 2 % agarose gel and visualised on a UV-transilluminator.

2.2.6.3 iPSC transfection

Transfection of iPSCs was achieved with electroporation. The Lonza nucleofection protocol was followed and electroporation executed in a Lonza Nucleofector 2b. Briefly, at ~80 % confluency iPSCs were prepared for transfection. 1-5 $x10^6$ cells, pSpCas9 vector (containing sequences for gRNA but with puromycin cassette removed), zeocin cassette, repair templates and nucleofector R solution were combined and transferred to a certified cuvette. Each sample was processed separately to avoid prolonged storage of cells in solution. The corresponding pre-installed programme was run. Cells were plated as single cells in a 96 well plate pre-coated and pre-warmed with Matrigel (1:60) and recovered at 37ºC.

2.2.7 iPSC culture and MN conversions

2.2.7.1 iPSC maintenance

iPSC colonies were propagated on a layer of Matrigel (1:60) and passaged 1:6 with Dispase Collagenase 1:1 (Gibco) at 80 % confluency. iPSCs were fed every 24 hours in E8 supplemented media (Gibco) and maintained at 37°C at 5 % CO_2.

2.2.7.2 iPSC-derived MN conversion

Spinal motor neurons were differentiated in MN Diff Base media with a variety of supplements following an adapted protocol of a well-established method (Maury et al. 2014; Selvaraj et al. 2017). MNs were differentiated in an agitated (300 rpm) solution, to encourage formation of spheres to mimic the formation of the neural tube. The stages of differentiation and the components required are outlined in Table 2.9.

Table 2.9 Media supplements for MN differentiation.

Stage of differentiation	Components
Day 0-6	20 µM AI, 0.1 µM LDN, 3 µM CHIR
Day 0	10 ng ml^{-1} FGF2, 10 µM RI
Day 2-6	100 nM RA, 500 nM SAG
Day 7-8	100 nM RA, 500 nM SAG, 10 ng ml^{-1} BDNF
Day 9-13	100 nM RA, 500 nM SAG, 10 ng ml^{-1} BDNF, 10 µM DAPT
Day 14-16	100 nM RA, 10 ng ml^{-1} BDNF, 10 µM DAPT, 10 ng ml^{-1} GDNF

2.2.7.3 MN culture

sMNs were cultured in MN-NF media ay 37ºC 5 CO_2 on LMF coated plates or coverslips, with a media change every 48 hours. For the first two days post dissociation, media was supplemented with GluE (1:1000) and U/FDU (1:10000).

3 Generation of an iPSC-derived motor neuron model for ALS-causing UBQLN2 mutation

3.1 Introduction

One of the original aims of this thesis was to generate an ALS-mutant iPSC line through CRISPR/Cas9 engineering. Multiple steps were completed in this process and are outlined below. However, due to the Covid-19 pandemic, full completion of the cell model was not possible. The resulting project changed dramatically and is described in detail in the remaining chapters. Nonetheless, significant effort was made during this aspect of the project which warranted acknowledgement in this thesis. In addition to the experimental work outlined below, a summary of i*n vivo* and *in vitro* neurogenesis was completed alongside a protocol for a systematic review into the role of UBQLN2 in disease (dx.doi.org/10.17504/protocols.io.bgynjxve) during the lockdown period. Due to lab pressures following return from the national lockdown, the latter was not completed but it is the aim of the author to do so upon completion of this thesis. Below is a summary of the work completed prior to the Covid-19 pandemic and subsequent alteration of this PhD project.

3.2 Generation of functional spinal motor neurons

Motor neurons (MNs) were differentiated following an adapted protocol of a well-established method (Maury *et al.*, 2014; Selvaraj, Livesey and Chandran, 2017). Overall, the conversion takes 16 days in suspension culture to efficiently produce ventralised spinal motor neurons. The iPSCs are first converted from an undifferentiated, proliferating state through to neuralisation, with the use of N2/B27 media and Dual SMAD inhibitors: AI and LDN (20 μM and 0.1 μM respectively). The mature nervous system contains two distinct and specialized classes of cells: neurons and glia. Both groups of cells are derived from the same neural precursor. Early Wnt signal activation, using 3 μM Wnt agonist Chir 99021, is crucial for rapid specification of spinal motor neurones from these neural precursor cells. Spinal cord identity is achieved through increased Retinoic Acid (RA) levels (100 nM) from Day 2, which drives the cells away from the default

rostral pathway and towards a caudal fate. The neural spheres are then ventralised with a gradient of Sonic hedgehog (Shh) agonist (500 nM). At Day 9, these motor neuron progenitor (pMN) domain patterned cells are consequently driven towards a motor neuron fate with the use of γ-secretase inhibitor DAPT (10 μM), which acts to increase neurogenesis and decrease gliagenesis through upregulation of expression of pro-neural genes and inhibition of Notch. Finally, the motor neurons are dissociated and cultured as a monolayer on Laminin, Matrigel and Fibronectin (LMF) coated dishes and left to mature for a further 3 weeks before use in experiments.

3.3 Quality Control of iPSC Cultures

Three different iPSC lines (CS02, CS25 18n2, CS25 18n6) from control (WT) individuals were selected for conversion to MNs. These lines had previously been used for oligodendrocyte conversion, but not MN conversion in this lab. Therefore, the efficiency of the MN conversion protocol had to be established before selecting the line for CRISPR/Cas9 mutation. Furthermore, as these lines had been held as iPSCs for a length of time, the genomic integrity and pluripotency of the cultures also needed validating prior to progression of CRISPR/Cas9.

3.3.1 Definiens Analysis of iPSC colonies and iPSC-derived MN

In order to confirm the pluripotency of iPSC colonies and efficiency of the conversion protocol into MNs, quality control analysis was first carried out. This was achieved through immunofluorescence staining followed by analysis with Definiens XD software (supported by Dr James Longden). The Hoechst-positive nuclei were identified based on fluorescence, contrast, size, and shape. Within this cellular population, the presence of iPSC or MN markers were then assessed (Table 3.1).

Table 3.1 iPSC quality control markers. A summary of the immunostaining markers used to assess the pluripotency and differentiation levels and their expression pattern. Markers are listed in increasing indication of differentiation.

Marker	Location	Function
Hoechst	Nuclear	Nuclear marker
Nanog	Nuclear	Pluripotency marker
Oct3/4	Nuclear	Pluripotency marker
TRA-1-60	Surface	Pluripotency marker
Neurofilament H	Cytoplasmic	Neuronal marker
Islet 1/2	Nuclear	Motor neuron marker

3.3.1.1 iPSC Quality Control

In order to assess the pluripotency of the iPSC colonies, a panel of three pluripotency markers was assessed (Table 3.1; Figure 3.1A). The number of cells and area of each marker, normalised to the nuclei threshold, were counted, and the mean and standard deviation calculated. All iPSC lines yielded similar positive results with regards to high expression of the three transcription factors representing pluripotency markers used in this panel (Figure 3.1B and C).

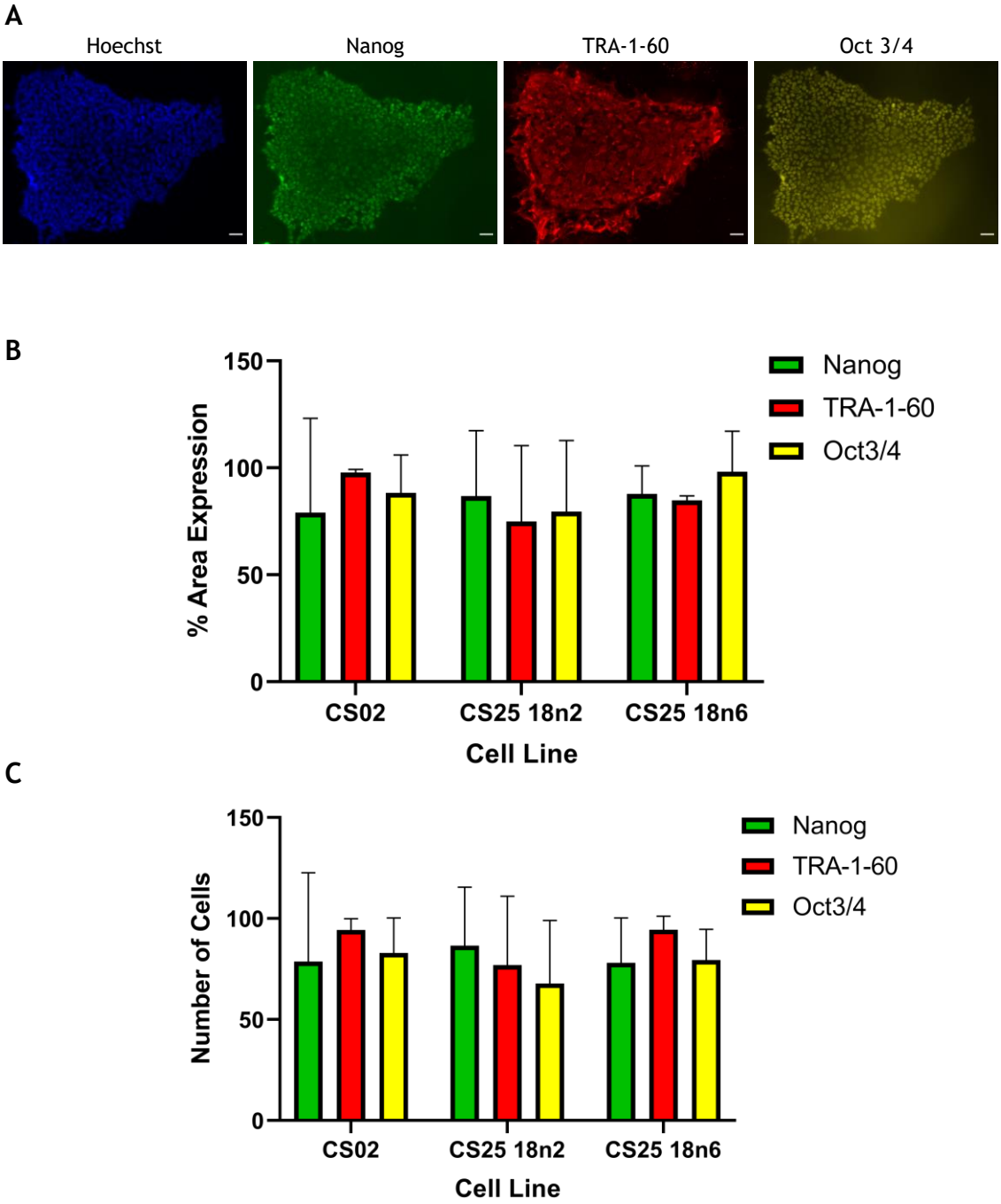


Figure 3.1 iPSC quality control using immunofluorescence and Definiens analysis. Immunostaining was carried out on all three control cell lines for Hoechst (Blue), Nanog (Green), TRA-1-60 (Red) and Oct3/4 (Yellow). A representative image from the CS25 18n6 line is shown (A) where the scale bar represents 50 µm. The percentage of area expressed (B) or number of cells (C) (normalised to Hoechst) expressing each pluripotency marker is shown graphically for each of the 3 control lines.

Chromosomal abnormalities can occur when holding cells in an unstable state, such as long-term cultivation of iPSCs (Vaz *et al.*, 2021). Karyotyping with the Aneuploidy BACs-on-Beads assay assesses the genomic integrity of the chromosomes at predictable locations susceptible to abnormalities in long-term

iPSC cultures. These results concluded that of the three lines in culture, only the CS25 18n6 line was karyotypically normal.

3.3.1.2 MN Quality Control

A slightly different approach was taken with the MN quality control, as motor neurons had not previously been generated from these control cell lines in our laboratory. The number and percentage of all neuronal cells (Neurofilament positive) from the whole cell population described as above was first calculated. Then, of these neuronal cells, the number and percentage expressing motor neurone marker Islet1/2 was calculated (Figure 3.2). This gives a better estimation of how successful the entire conversion protocol was. Analysis was carried out at both day 7 and day 12 post-dissociation to ensure complete differentiation with no residual neuronal precursor cells (NPCs) present. This analysis is simply qualitative as there is no accepted threshold of what a successful conversion is. Nonetheless, a large proportion of the neurons produced were indeed motor neurons, suggesting the protocol used to generate MNs from iPSCs was successful and reproducible.

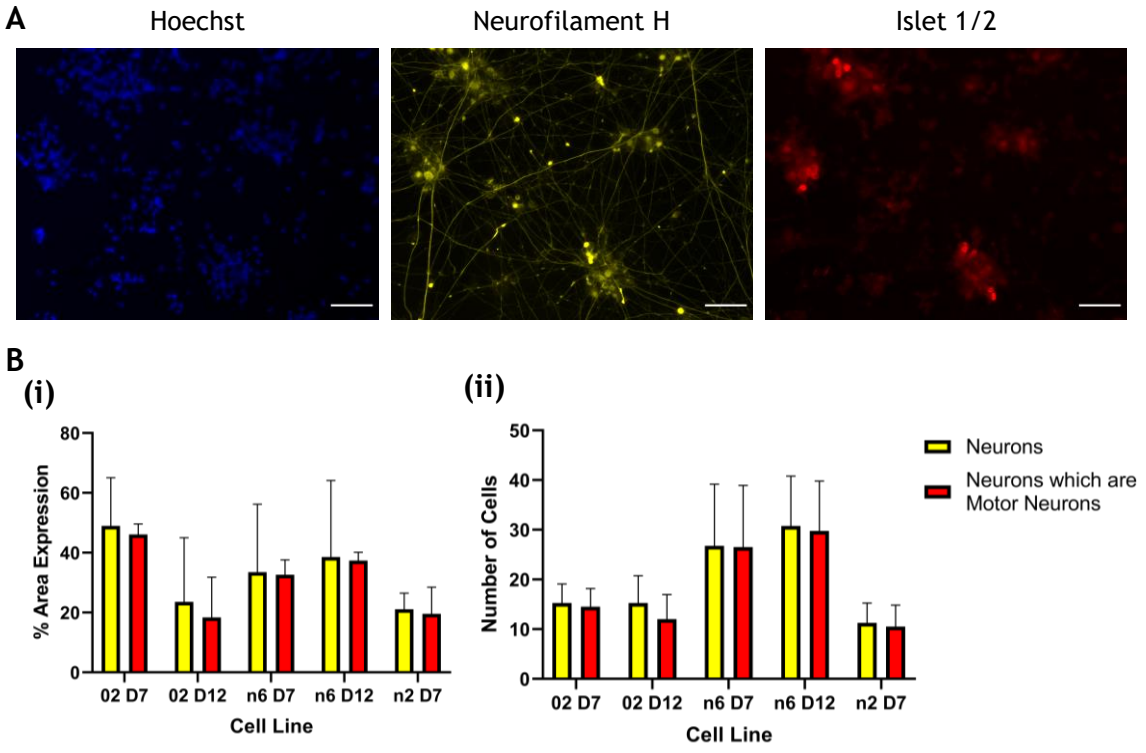


Figure 3.2 Quality control of iPSC-derived MNs. (A) Representative image of the immunostaining of cell line CS02 at day 12 (scale bar represents 50 µm) used in the definiens analysis with results shown in (B). Data is expressed as a percentage of cell area (i) or the number of cells (ii) expressing either Neurofilament H (neuronal marker, yellow) or Islet 1/2 (motor neuron marker, red).

Together, the Definiens analysis combined with the Karyotyping analysis dictates that the CS25 18n6 line should be used from here on in; this line had the highest motor neuron enrichment with a karyotypically normal background.

3.3.2 Verifying the maturity of the iPSC-derived MNs

Monitoring the electrophysiological condition of MNs *in vitro* is one of the most thorough techniques developed to assess their maturity. Voltage gated K^+ and Na^+ ion channel currents capable of generating induced action potentials (APs) are developed in the early stages of iPSC-derived MN differentiation. However, these induced APs develop overtime, from slow depolarization and repolarization to sharper APs as Na^+ and K^+ currents increase (Johnson et al., 2007).

To determine whether the iPSC-derived MNs produced in this protocol were mature and functioning neurons, whole-cell current-clamp recordings were performed. Figure 3.3A displays the electrophysiological trace of CS25 18n6 neuron (n=1) at three weeks post-dissociation (work complete by Dr. Matt

Livesey). Sequential current steps were injected into a single neuron, with depolarizing peaks indicating action potential firing. Furthermore, the train of APs without diminished amplitude (Figure 3.3B) is characteristic of mature neurons *in vitro* (Johnson et al., 2007). It can therefore be concluded from these data that the differentiated MNs generated using the above outlined protocol can indeed fire action potentials in response to depolarizing current. Thus, these iPSC-derived MNs are functional neuronal cell models.

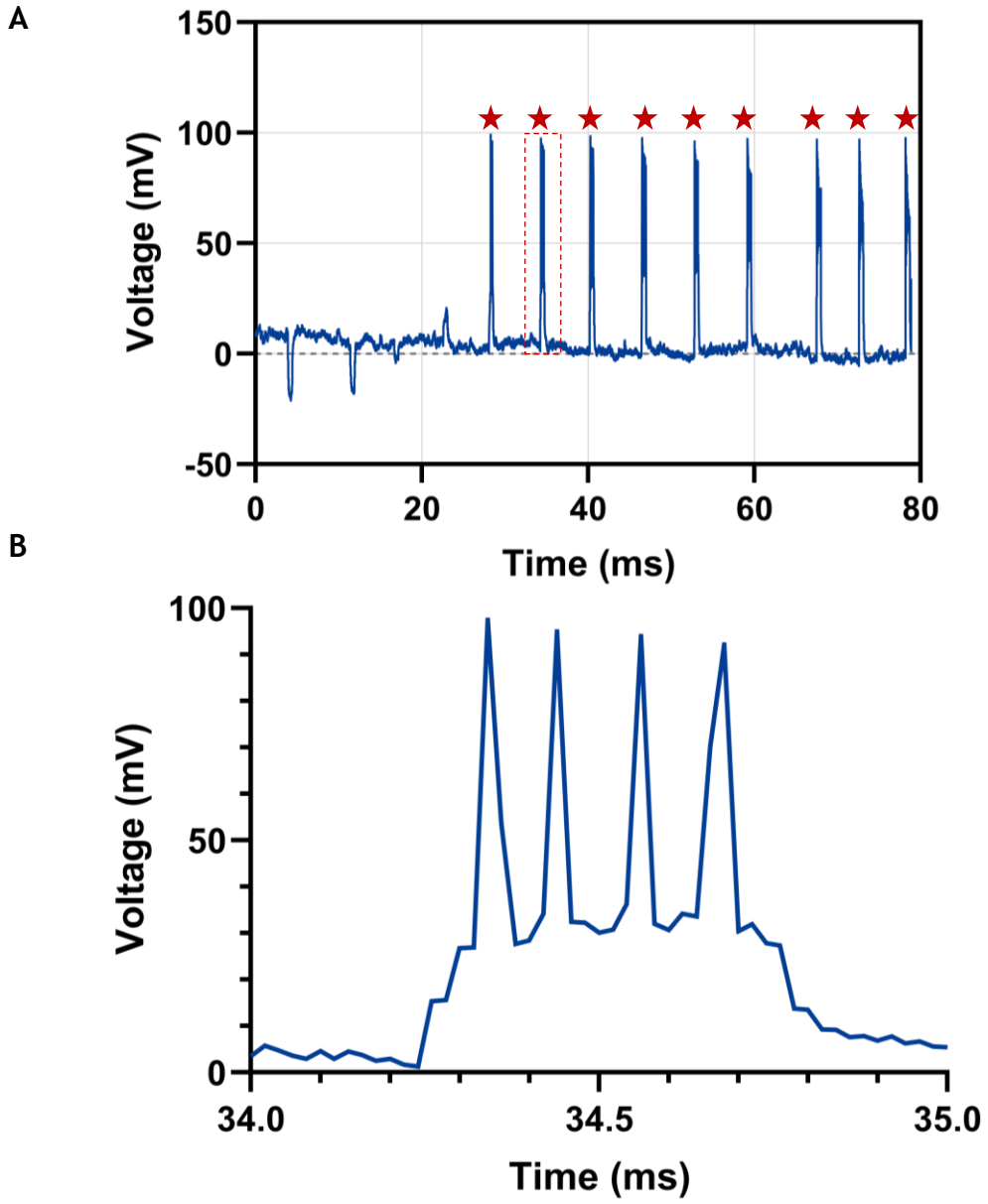


Figure 3.3 iPSC-derived MNs fire action potentials. (A) Current injection whole-cell recordings of an iPSC-derived MN from the control line CS25 18n6. Red star highlights successful action potential firing with the action potential highlighted within the red box enhanced in (B) showing successive action potential induction. Action potentials displayed as change in membrane voltage (mV) over time (ms).

3.4 Stress Assay Development

The general function of UBQLN2 in the cell and its contribution to ALS development remains elusive. To test different stress pathways, a series of stress assays were developed. Heat shock is a global stressor (Vogel, Dux and Wiessner, 1997) whilst tunicamycin is an ER stressor (Wu *et al.*, 2018). Once optimised, these assays were to be used in conjunction with western blot and qPCR analysis to determine the effects of UBQLN2 mutations on the neurons ability to handle stress.

3.4.1 Heat shock stress

Preliminary experiments to optimise experimental conditions for the heat shock response assay in iPSC-derived MNs were conducted. There was no evidence in the literature of previous attempts to stress iPSC-derived MNs with heat and monitor their recovery post-stress. Based on the work in primary rat neuronal cultures (Vogel, Dux and Wiessner, 1997), an initial starting point of 42ºC for 30 minutes was used to stress the neurons, followed by 24- and 48-hour recovery timepoints.

Previous studies have reported increased levels of ubiquitin following stress in Hek293 cells (Hjerpe *et al.*, 2016) and WT cells recovered whilst UBQLN2 mutants did not. A similar pattern was expected in UBQLN2 mutant iPSC-derived MN lines, which may possibly lead to increased neuronal death. First, immunofluorescence of UBQLN2 was optimised. iPSC-derived MNs express UBQLN2 and this is easily visualised in the neurons (Figure 3.4A). The level of this signal appears unchanged throughout the stress, demonstrating iPSC-derived MNs can withstand global stress (Figure 3.4B). This stress level could be increased to investigate whether any phenotypes become apparent with increased stress, alongside staining for ubiquitin or caspase 3 to monitor protein degradation and cell viability.

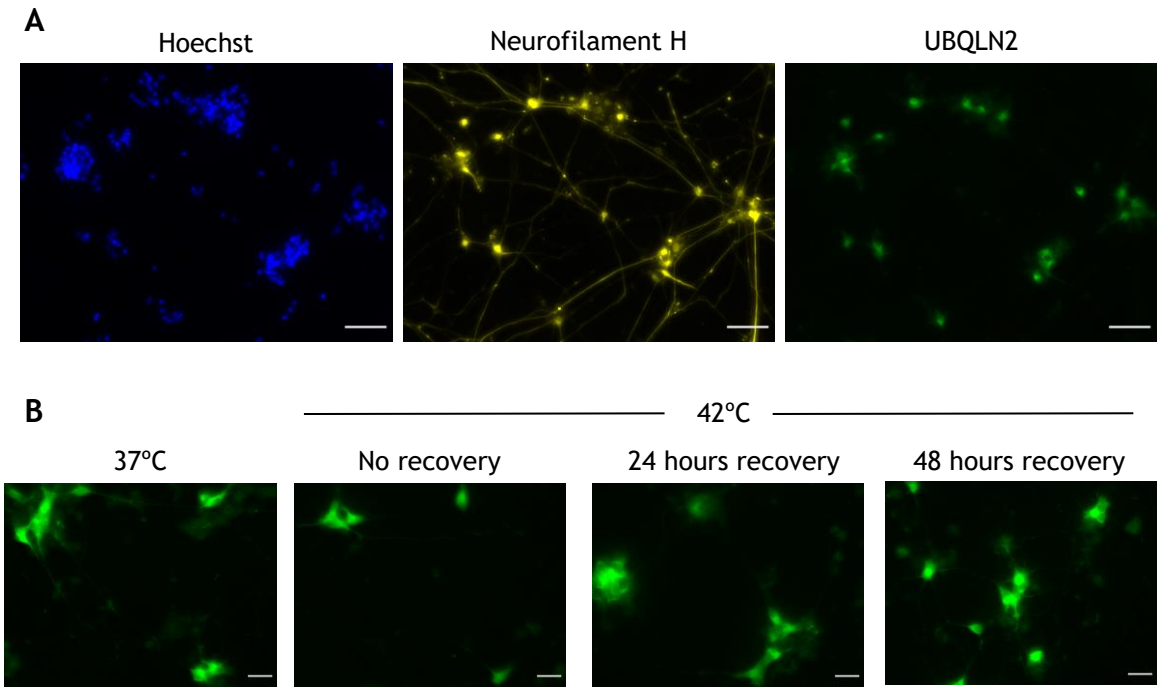


Figure 3.4 iPSC-derived MNs can withstand heat stress. (A) Example image of the staining used to confirm the cells under investigation are living neurons (nuclear stain, Hoechst, blue; neuronal marker, Neurofilament H, yellow; UBQLN2, green). Scale bars represent 50 µm. (B) iPSC-derived MNs express UBQLN2 and can withstand insult from heat shock.

3.4.2 ER stress

There is evidence for a role of UBQLN2 in stress granule formation following endoplasmic reticulum (ER) stress (Alexander *et al.*, 2018; Dao *et al.*, 2018). The impacts of ER stress on UBQLN2 behaviour in the iPSC-derived MNs was established using an increasing concentration gradient of tunicamycin, a known ER stressor, over 48 hours. Tunicamycin acts to inhibit the DPAGT1 enzyme in the first step of glycoprotein biosynthesis. Thus, an accumulation of misfolded proteins and subsequent ER stress occurs with Tunicamycin treatment (Wu *et al.*, 2018). Therefore, Tunicamycin treatment was expected to cause the movement of UBQLN2 into foci formations. These would be further explored as possible recruitment to stress granules (Alexander *et al.*, 2018; Dao *et al.*, 2018).

A pattern of increasing puncta was observed when treating the cells with tunicamycin for 48 hours (Figure 3.5). However, this pattern was not statistically significant. Increasing the sample size in future experiments may have revealed a significant change not detectable in this smaller preliminary experiment. Nonetheless, there was no significant change observed in UBQLN2 localization

following ER stress in WT iPSC-derived MNs. Furthermore, the MNs were able to withstand this stress, which would provide a good baseline, upon which stress levels can be increased.

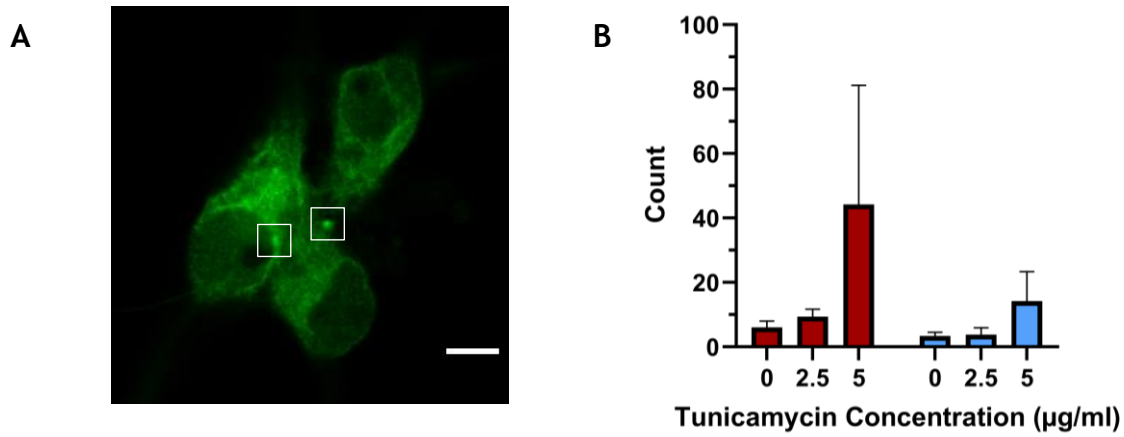


Figure 3.5 iPSC-derived MNs can withstand ER stress. (A) Example of the UBQLN2 foci counted in this experiment (white boxes) with UBQLN2 visualised in green and the scale bar representing 20 µm. (B) Graphical representation of the overall number of foci counted across all cells (red) and the number of cells containing foci (blue).

3.5 Generating Mutant Lines

In order to compare the effects of a disease-linked UBQLN2 mutation with a wild-type control, the CRISPR/Cas9 gene editing system was utilised to introduce ALS-patient UBQLN2 mutations into control iPSCs prior to MN conversion. The UBQLN2^{P506T} and UBQLN2^{P497H} mutants were chosen due to the broad neurodegenerative phenotype in both mice (Le *et al.*, 2016) and humans (Deng *et al.*, 2011).

Two guide RNAs (gRNAs) per mutation were designed with Benchling to introduce the mutations UBQLN2^{P506T} (a/b) or UBQLN2^{P497H} (c/d) with the lowest possible off target effects whilst ensuring a suitable PAM sequence was located in the guide (Table 3.2). The sequences encoding the gRNAs were cloned into the pSpCas9 vector. A colony PCR and subsequent sanger sequencing confirmed correct cloning into the vector, ensuring the gRNAs and other surrounding DNA sequence were intact and unmutated. The correct colonies were then amplified, and DNA extracted using a Maxiprep technique.

Table 3.2 gRNA design. gRNAs used in CRISPR/Cas9 gene editing to target and cut WT human UBQLN2. PAM sites are in capital underlined.

Name	Code	Wild Type gRNA Sequence 5'->3'
P506T_gRNA1_TOP	a_top	CACCgtagtccctttaccccat
P506T_gRNA1_Bottom	a_bottom	AAACatgggggtaaaagggactac
P506T_gRNA2_TOP	b_top	CACCgggggtaaaagggactata
P506T_gRNA2_Bottom	b_bottom	AAACtatagtccctttaccccc
P497H_gRNA1_TOP	c_top	CACCgtaggcccagtcacccccat
P497H_gRNA1_Bottom	c_bottom	AAACatgggggtgactgggcctac
P497H_gRNA2_TOP	d_top	CACCgtcacccccataggccccat
P497H_gRNA2_Bottom	d_bottom	AAACatggggcctatgggggtgac

Test transfections were carried out in Hek293 cells using Lipofectamine. The efficiency of the Cas9 nuclease was confirmed using a T7 endonuclease assay (Figure 3.6); T7 endonuclease detects mismatches from DNA repair and digests at these points, thus indicating whether the DNA has been cut at the correct locations by Cas9.

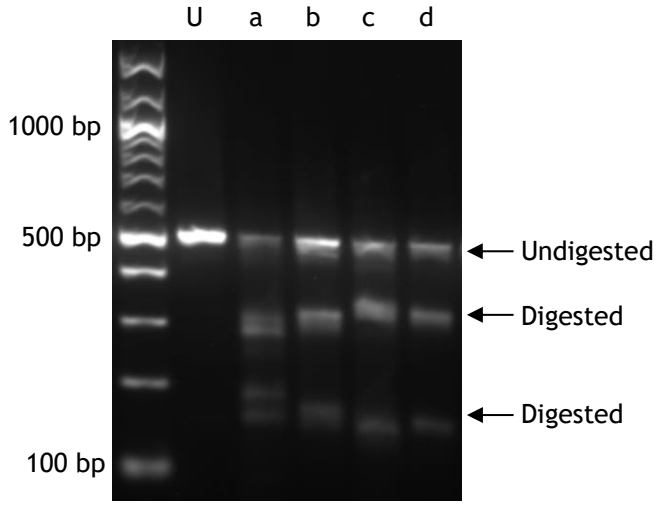


Figure 3.6 T7 endonuclease assay confirms efficiency in cutting by Cas9. A T7 endonuclease assay demonstrating Cas9 has cut the DNA in the desired locations in samples a and b (P497H gRNA) and c and d (P506T gRNA) but no cutting occurred in the transfected cells.

A panel of stress assays had been developed and optimised (Section 3.4) for use in experiments with the iPSC-derived motor neurons. One of these stressors not yet developed but intended for use was puromycin stress, to increase aggregate formation within the neurons (Hjerpe *et al.*, 2016). Therefore, the decision was made to remove the puromycin selectivity cassette from the vector; the possibility of integration of the puromycin-resistance cassette into the genome and its possible presence during the stress assay may have cast doubt on any data collected. EcoR1 restriction sites flanking the puromycin cassette were targeted with an EcoR1 restriction digest and the digestions run on a gel for verification (Figure 3.7). Successfully digested samples were ligated and transformed into competent cells. Miniprep cultures were set up and DNA extracted, with removal of the puromycin cassette subsequently confirmed by restriction digest and sequencing of re-ligated constructs.

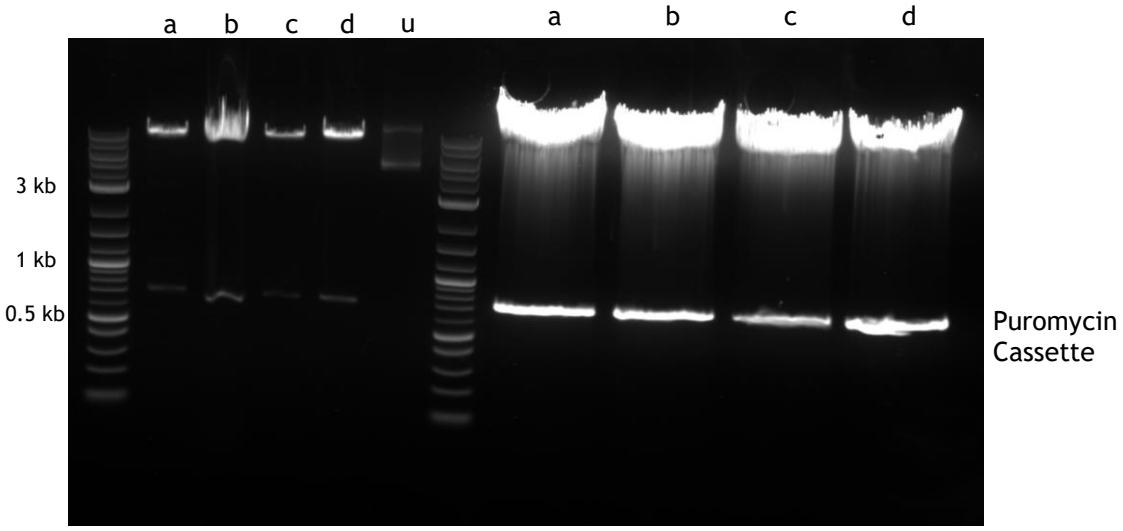


Figure 3.7 Removal of the puromycin selectivity cassette. Successful removal of the puromycin selectivity cassette from all vectors with EcoR1 restriction digest. Samples a-d underwent the restriction digest process whilst "u" is the undigested control.

Repair templates for the CRISPR/Cas9 gene editing were synthesised with the intention to introduce the desired mutation (P506T). 200 bp ssODNs were generated (Table 3.3) with 3 aims in mind:

1. To generate the desired ALS-causing mutation.

2. To mutate the PAM site or seed region to prevent continuous binding and cutting of the Cas9 after repair has occurred.

3. To introduce or lose a restriction enzyme site unique to the amplified region of DNA.

Table 3.3 Repair template strategy to introduce ALS mutations into UBQLN2. Repair template design strategy highlighting the ALS-causing mutation to be inserted, alongside the PAM mutation required to prevent continued cutting by Cas9 and the restriction site (RE) alteration as a consequence.

Repair Template	ALS mutation	PAM mutation	RE site
P506T	CCT → A<u>C</u>T	G502G (GGC → GG<u>A</u>)	Intro Adh1
P497H	CCC → C<u>A</u>C	T496T (ACC → AC<u>T</u>)	Loss Hph1

The resulting vector was co-transfected using electroporation as a triple transfection into control iPSCs (line CS25 18n6), along with a zeocin-resistance cassette and the repair templates designed to introduce the mutations. Transfected iPSCs were plated as single cells and individual colonies were picked and cultured for screening. Unfortunately, due to the Covid-19 pandemic, the screening could not continue, and lines had to be destroyed due to national lockdown. As a result, the PhD project was adapted, and focus altered to investigate the PRR region of UBQLN2 in its structure and function. The remainder of this thesis details the advances made in this investigation.

4 Characterisation of the biophysical properties of the PRR Region and its (potential) trimeric state

4.1 Introduction

The proline-rich-repeat (PRR) region of Ubiquilin 2 (UBQLN2) has historically been called the Pxx domain. This study challenges this nomenclature on two grounds: First, the word *domain* suggests a defined structure, of which there is no evidence of in the PRR region in current literature. Second, in viewing the amino acid triplets as P-X-X, a bias occurs in the way the sequence is viewed. I propose this bias has prevented a collagen-like amino acid sequence being detected and explored further until now (Figure 4.1).

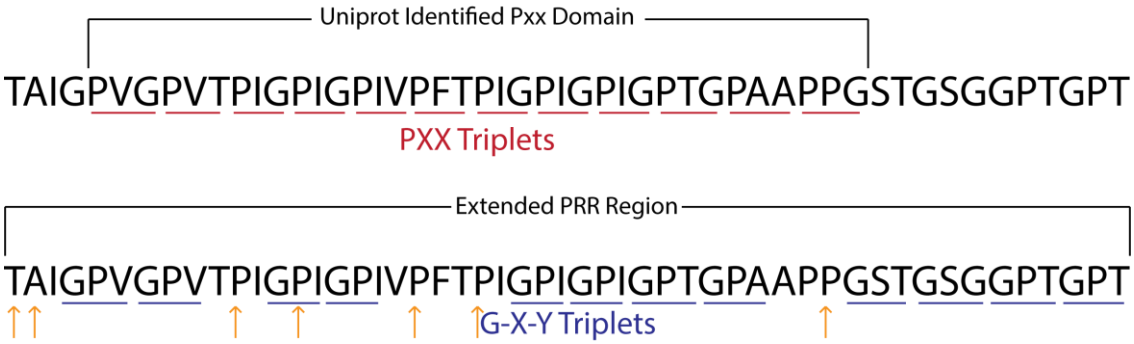


Figure 4.1 Evidence of a collagen-like amino acid sequence in the PRR region of UBQLN2. Single letter amino acid code of the PRR region of UBQLN2, with the classical Pxx domain identified by UniProt vs the extended PRR region under investigation in this study. The P-X-X triplets are highlighted in red whilst the collagen-like G-X-Y triplets highlighted in blue. The orange arrows denote locations of ALS-causing mutations, the majority of which fall outside the G-X-Y triplets.

The poly-L-proline type II helix (referred to as PPII or polyproline-II from here on in) is markedly different to the other two major structures of folded protein, α -helix and ß-structures. The helix formed is extended (3.1 Å per residue compared to 1.5 Å in the α-helix) with a 3-fold rotational symmetry with three residues per turn (Adzhubei, Sternberg and Makarov, 2013). The high content of prolines within the PRR region suggest a secondary structure similar to the polyproline-II helix. Furthermore, the polyproline-II helix is an established component of the collagen triple helix structure (Ramachandran and Kartha, 1954).

Collagen-like sequences are characterised by $(Gly\text{-}Xaa\text{-}Yaa)_n$ triplet repeats ranging in length from 9 to <450 residues across all species, with varying levels of interruptions (Brodsky and Persikov, 2005; Yu *et al.*, 2014). These repeats form triple-helical coils in which Glycine is the only amino acid small enough to accommodate the tight coil formed. In mammalian collagen sequences, the Y position is usually occupied by a Proline residue, which is often hydroxylated for stability (Prockop, 2004). The sequence observed in UBQLN2's PRR region does not have Prolines in the Y position. Instead, Proline often occupies the X position. The occurrence of Gly-Pro-Xaa repeats aligns the amino acid sequence of the PRR region more closely to bacterial collagen-like proteins (Rasmussen, Jacobsson and Björck, 2003). The body of evidence around bacterial collagen-like proteins has been expanding in recent years (Lukomski *et al.*, 2000; Rasmussen, Jacobsson and Björck, 2003; Paterson *et al.*, 2008; Duncan *et al.*, 2011; Pizarro-Guajardo *et al.*, 2014; Yu *et al.*, 2014; Bachert *et al.*, 2015; Zhao *et al.*, 2015) and investigations into their mechanisms of stability have begun to be elucidated (Lukomski *et al.*, 2000; Rasmussen, Jacobsson and Björck, 2003; Bhowmick and Fields, 2013; Bachert *et al.*, 2015).

This chapter provides evidence that the PRR region of UBQLN2 is capable of oligomerizing and investigates the secondary structure of the species identified. Furthermore, it explores potential structural homology of the PRR region with bacterial collagen-like proteins and proposes alternative methods of predicting structure based on these new assumptions.

4.2 Protein truncations of UBQLN2 can be stably expressed and purified

The UBL domain of UBQLN2 has been found to interact with the UBA domain (Nguyen, Puthenveetil and Vinogradova, 2017), whilst the STI1-II domain interacts with various C-terminal regions of UBQLN2, including the PRR region (Dao *et al.*, 2018). A reductionist model was adopted in this study to investigate the contribution of the PRR-region to structure and function, without the interaction and possible interference of the aforementioned regions. This comprised of a truncation of UBQLN2 expressing a C-terminal portion of the protein, containing only the PRR region (residues 487-537), UBA domain (residues 581-624), and the

intervening linker. Controls of the UBA domain alone and the linker sequence with the UBA domain (residues 538-624) were also created to ascertain the impact of the PRR specifically (Figure 4.2A). A His_6 tag was added N-terminally followed by a TEV cleavage site to allow removal of the His_6-tag for downstream applications.

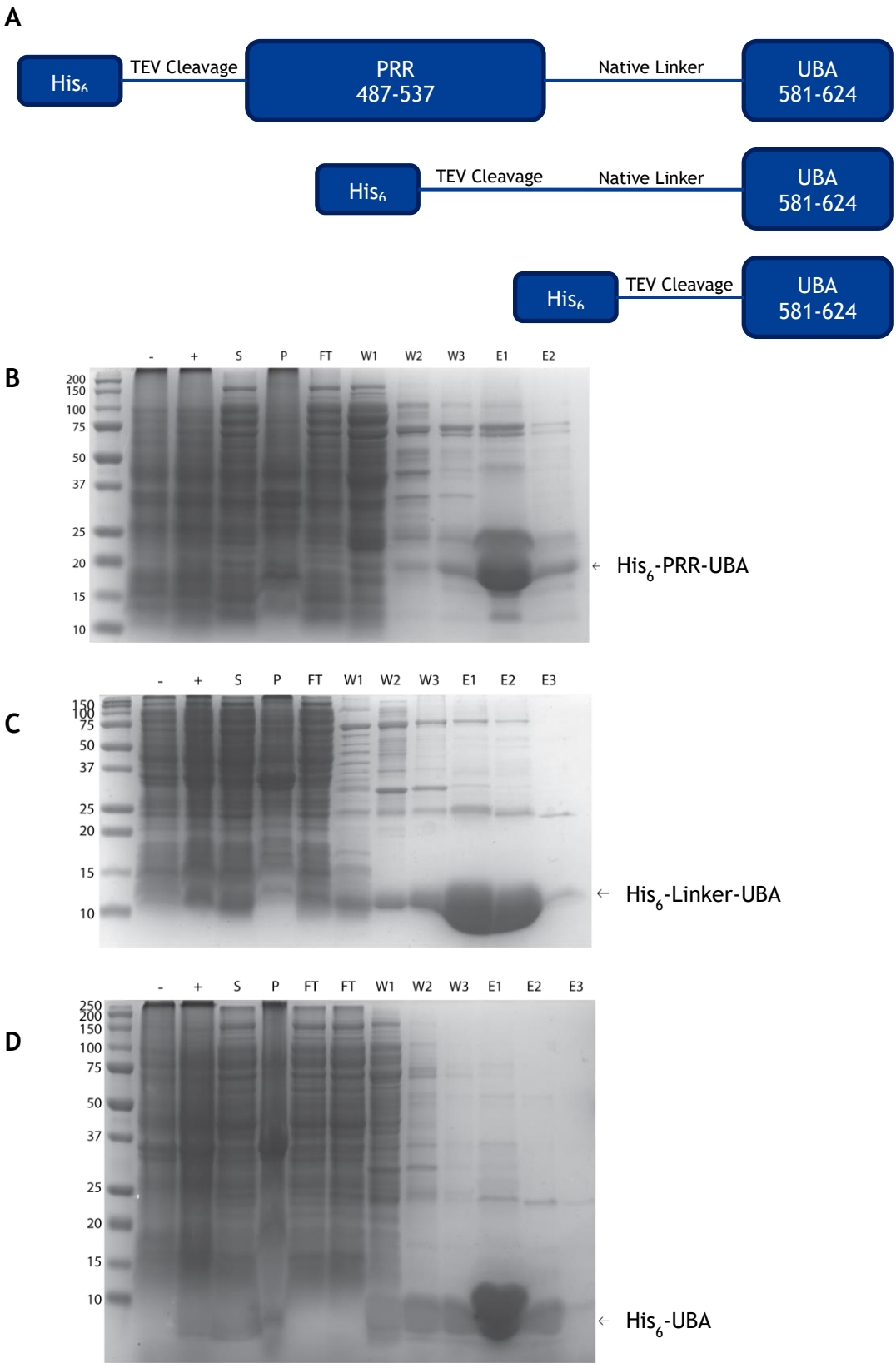


Figure 4.2 C-Terminal domains of UBQLN2 and their expression and purification. A representation of the various truncations of UBQLN2 used in this study (A) followed by the successful purifications of the PRR-UBA (B), Linker-UBA (C) or UBA only (D) proteins, analysed with SDS-PAGE and Instant Blue Coomassie staining. Notations as follows: Addition of 0.1 mM IPTG is denoted by +/-, supernatant (S) or pellet (P) fractions following lysis and clarification, flowthrough (FT) of the supernatant after adding to the Ni-NTA column, 3 washes (W1/W2/W3) and 3 elutions (E1/E2/E3). Note the schematics are not drawn representative to actual contribution in the amino acid chain.

Once designed, the constructs were expressed in *Escherichia coli* Rosetta™ 2(DE3) cells and recombinant proteins purified (Figure 4.2B-D) using the Ni-NTA affinity purification protocol outlined in section 2.2.4.1 and 2.2.4.2.

Maximising the proportion of soluble protein from each expression was achieved by optimising the concentration of IPTG used to induce protein synthesis (Figure 4.3A) and the subsequent overnight incubation temperatures (Figure 4.3B). A single overnight bacterial culture was split into four followed by induction of His_6-PRR-UBA protein synthesis using varying concentrations of IPTG at 25°C, followed by overnight incubation. Cells were lysed using 5x 10 second on/off sonication at 25 Hz and clarified into soluble or insoluble (pellet fraction) by centrifugation for 35 minutes at 31000 g. Figure 4.3A shows the range of IPTG concentrations tested and the proportional soluble: pellet protein following induction. As a result of producing the lowest insoluble protein proportion, an IPTG concentration of 0.1 mM was used in all further expressions. A similar method was applied to assess the impact of temperature at induction and in the subsequent incubation period. Samples were induced at varying temperatures and incubated at either 16°C overnight, 25°C overnight or 37°C for 4 hours. Figure 4.3B demonstrates the effect of temperature at the point of induction and the subsequent overnight incubation on the yield of His_6-PRR-UBA. Again, due to the lowest proportion of insoluble protein produced, 16°C induction temperature with overnight incubation was used in all subsequent expressions.

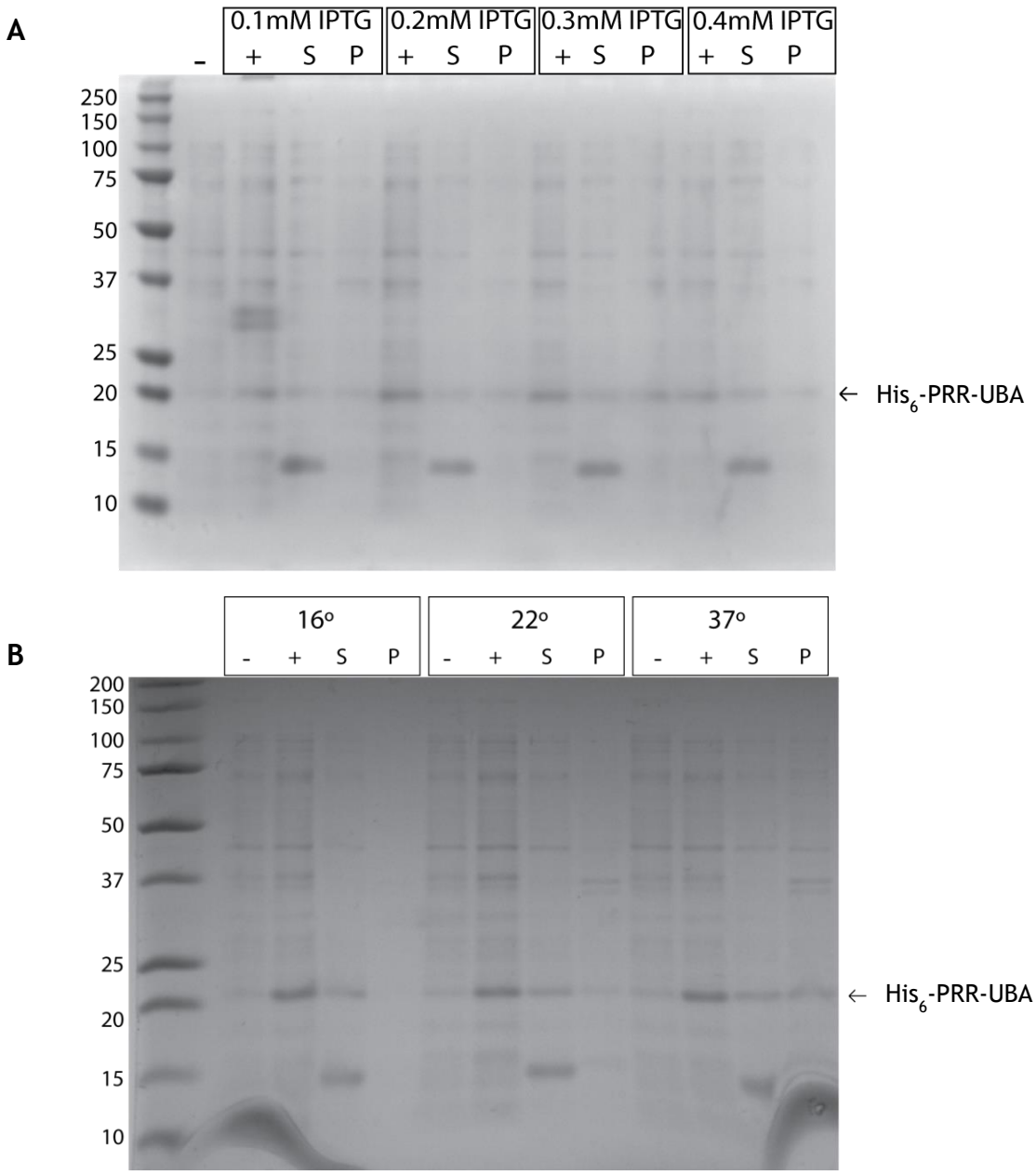


Figure 4.3 Optimisation of soluble protein expression in *Escherichia coli*. SDS-PAGE analysis of soluble protein expression in varying IPTG induction conditions (A) with overnight incubation at 25°C or overnight incubation temperatures (B) following induction with 0.1 mM IPTG. IPTG induction denoted by -/+ and soluble supernatant fraction (S) or insoluble pellet (P) fraction following lysis and clarification by centrifugation.

Cutting at the TEV cleavage site to remove the His_6 purification tag was optimised using the His_6-UBA only protein (residues 581-624) for simplicity, due to its distance from the TEV protease control band on an SDS PAGE gel. 15 µg of purified His_6-UBA only protein was incubated at a 1:1 ratio with TEV protease at either 4˚C for 48 hours or 30˚C for 1 hour and run on an 11 % SDS PAGE gel followed by Instant Blue Coomassie staining (Figure 4.4). Due to the low proportion of aromatic residues downstream of the TEV cleavage site in all the constructs, and thus the

lack of Coomassie dye binding, complete removal of the His_6 tag by TEV protease was established by the loss of signal in the Instant Blue Coomassie stain. 48-hour incubation at 4°C resulted in almost complete removal of the tag while one hour incubation at 30°C led to a large proportion of protein retaining the tag (Figure 4.4).

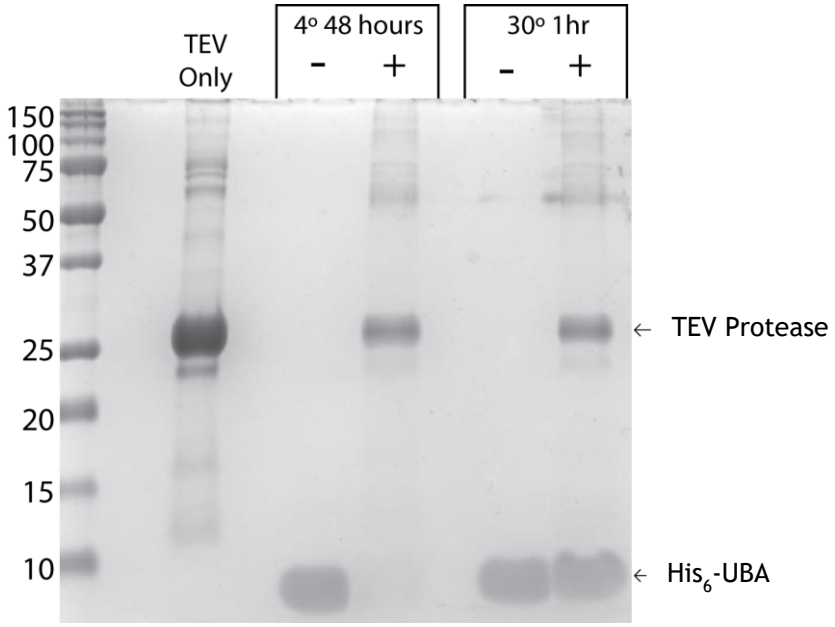


Figure 4.4 TEV Protease cleavage conditions. Cutting of the His_6-UBA with TEV protease was assessed by SDS-PAGE analysis and Instant Blue Coomassie staining. The loss of Instant Blue Coomassie stain following incubation with TEV protease was used to determine removal of the proteins N-terminal tag. -/+ denotes the addition of TEV protease to His_6-UBA protein at a 1:1 ratio.

Large quantities (~60 mg/ml) of recombinant protein were expressed and purified following optimisation of protocols. Furthermore, specific cleavage of the His_6 tag was possible. In order to establish if the recombinant protein was able to multimerize, as had previously been established in the full length protein (Hjerpe *et al.*, 2016), investigations into its oligomeric state were required.

4.3 Investigating the oligomeric profile of the His_6-PRR-UBA protein

4.3.1 Assessment of the homogeneity of oligomeric states and estimation of molecular weight by FPLC and Native PAGE analysis

UBQLN2 has previously been shown to oligomerize into a dimer and a trimer (Hjerpe *et al.*, 2016), as well as undergo liquid-liquid phase separation (LLPS)(Dao *et al.*, 2018). Identification of the bacterial collagen-like sequence residing within the PRR region suggests a possible role for the PRR region in forming a collagen triple helix. Fast Protein Liquid Chromatography (FPLC) using a size exclusion column was employed to investigate the oligomeric profile of the His_6-PRR-UBA protein, with the identity of the peaks confirmed by SDS-PAGE.

Initially, 400 µl of 6 mg/ml protein was run on an SD200i 10_300 column via an AKTA purifier. The appearance of multiple peaks indicates numerous, differentially migrating species within the recombinant protein sample (Figure 4.5A). This trace was compared with protein standards run on the same column (see section 2.2.5.1) to give an estimation of the molecular weight (MW) of the various multimeric species. The first, potentially multimeric, peak eluting at 9.1 ml is estimated to have a MW of 766 kDa. The later peak, eluting at 15 ml, is estimated to have a MW of approximately 42 kDa. As a comparison, purified His_6-UBA only was run in the same way on the same column and yielded a single peak with a MW estimation of 12 kDa (Figure 4.5B). The Linker-UBA control protein aggregated, and so was disregarded from future experiments. The sequences of both recombinant proteins were inputted into *ProtParam*, which predicted a MW of 16 kDa for a monomeric form of the His_6-PRR-UBA protein and 7 kDa for the His_6-UBA protein.

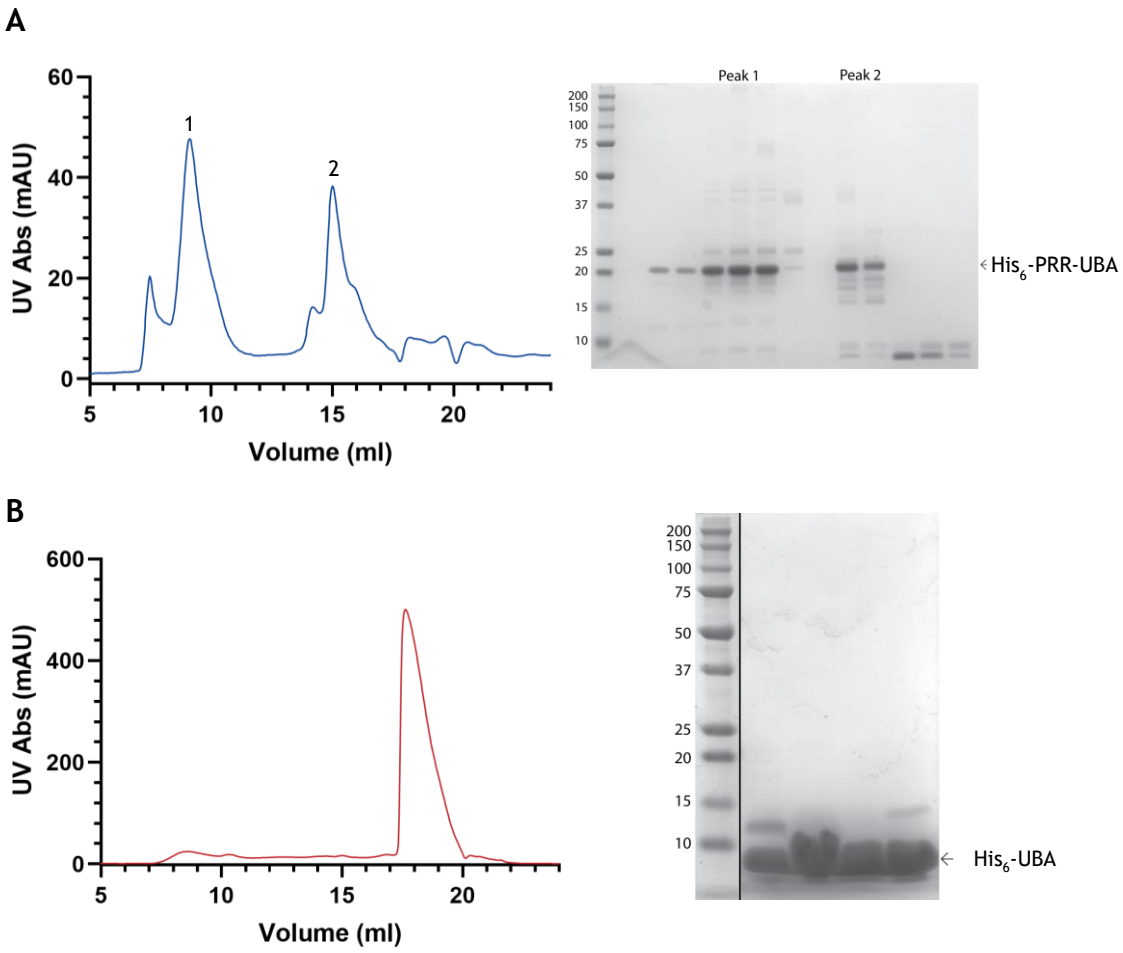


Figure 4.5 Analysis of oligomeric states using FPLC. FPLC was performed using a SEC column (SD200i) followed by subsequent SDS-PAGE analysis and Coomassie stain of His_6-PRR-UBA (A) and His_6-UBA only (B) peaks. Splice line identified by solid black line in (B) Coomassie image. Samples of 10 mg/ml in 500 µl volume were loaded.

Multimerization in the His_6-PRR-UBA protein corroborates previous findings of oligomerization of full length UBQLN2 (Hjerpe *et al.*, 2016). Furthermore, the lack of oligomerization in the His_6-UBA protein suggests any multimerization seen is likely due to the presence of the PRR region.

Whilst the predicted MW of all proteins tested appear larger than predicted at first glance (Peak 1= 766 kDa, Peak 2= 42 kDa vs 16 kDa predicted, UBA only= 12 kDa vs 7 kDa predicted), it is worth noting the mechanism by which FPLC and Native PAGE operates, relying on the assumption that the protein in question is globular. Whilst the structure of the PRR region remains elusive, previous NMR investigations (Dao *et al.*, 2018) and the current hypothesis of this study would assume that the PRR region would not form a globular unit. Therefore, migration

through the column will not correspond accurately with the protein standards used in the calibration, so projections of MW should be made with caution.

Stable collagen-like triple helical domains have shown resistance to denaturation with SDS (Tasab, Jenkinson and Bulleid, 2002). The stability of the multimer and monomer peaks of the His_6-PRR-UBA protein was therefore assessed by Native PAGE in the absence and presence of 2 % (w/v) SDS and stained using *Invitrogen* Silver Stain kit (Figure 4.6A). To confirm the identity of the bands as the His_6-PRR-UBA, and to ascertain if any resolution between the various higher order species could be obtained, a Native PAGE and western blot using the *BioRad* Anti-His antibody was performed with all elutions contributing to the peak (Figure 4.6B). The multimeric peak (Peak 1, 766 kDa from Figure 4.5A) displays a level of resistance to SDS denaturation where the monomer (Peak 2, 42 kDa from Figure 4.5A) does not. This suggests the possibility that the protein in Peak 1 is forming a stable, multimeric structure within the PRR-UBA protein.

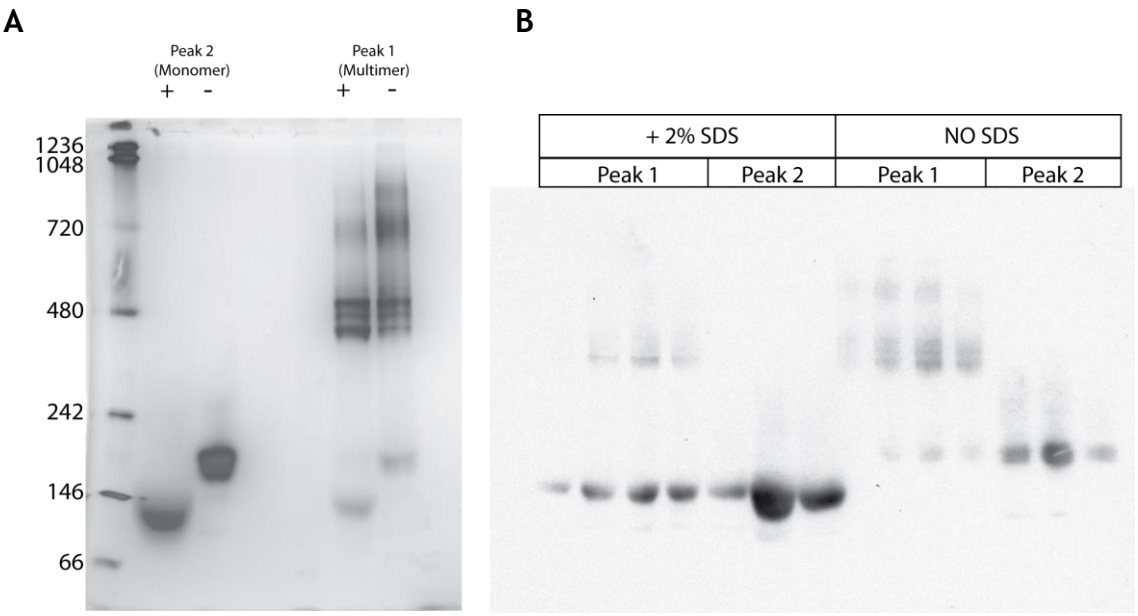


Figure 4.6 The PRR-UBA multimer demonstrates resistance to SDS. Native PAGE analysis followed by silver stain (A) or western blot (B) with anti-His antibody in the absence or presence of 2 % SDS (+/-). Initially, the elution at the highest point in the peak was assessed (A). All fractions contributing to the peak were subsequently blotted for His in the absence and presence of 2 % SDS (B).

4.3.2 Assessment of the homogeneity of oligomeric states and estimation of molecular weight by Sedimentation Velocity

Due to the potential drawbacks of estimating MW with FPLC or Native PAGE (see section 4.3.1), Analytical Ultracentrifugation (AUC) was employed. AUC is performed in free solution, removing the complications of matrix interaction or surface binding potentially experienced in FPLC and Native PAGE.

AUC is a technique employed to contribute to the understanding of the biophysical properties of macromolecules in solution. Measurement of the hydrodynamic sedimentation properties of molecules, Sedimentation Velocity (SV), can inform on a molecule's shape, mass and interactions with itself or other components.

The composition of the two peaks separated by FPLC (Figure 4.5A) of the His_6-PRR-UBA protein were investigated using the technique of SV. The experimental data, residuals and fit of the data as a c(s) plot are displayed in Figure 4.7. Sedimentation coefficients (S) were normalised to $S_{20,w}$ values using the parameters outlined in section 2.2.5.3. Analysis of the multimeric Peak 1 reveals three species with $S_{20,w}$ values of 1.2 S, 3.2 S and 5.3 S respectively (Figure 4.7A). This corroborates the evidence of multiple multimeric species seen in the previous experiments (Figure 4.5; Figure 4.6). To estimate the mass of each species, each peak on the distribution plot was integrated to obtain the weight-averaged values for sedimentation coefficient and molecular mass. This produced molecular weight estimations of the species as 6 kDa, 35 kDa, and 76 kDa respectively. Analysis of Peak 2, the assumed monomer, produced a result suggesting a single species with an $S_{20,w}$ value of 1.7 (Figure 4.7B). Based on the $S_{20,w}$ value, the estimate the molecular weight of this species is in the region of 14 kDa, which is comparable with the predicted value for monomeric His_6-PRR-UBA generated from ProtParam (16 kDa).

From the data generated, it is difficult to elucidate the order of multimerization occurring in the various species in Peak 1. The species of lowest S value (1.2 S or 6 kDa) is likely low molecular weight contamination. Based on the experimental MW estimation of the monomeric His_6-PRR-UBA as 14 kDa (Figure 4.7B), a dimeric or trimeric structure would produce a protein with a MW in the region of 28 kDa

or 42 kDa respectively. Instead, intermediate values of 35 kDa and 76 kDa respectively are observed. Similar to previous MW estimations (see section 4.3.1), caution should be exerted when drawing conclusions from this data. The analysis method of c(s) distribution employed here depends upon a number of assumptions. Firstly, a constant shape and equal frictional ratio (f/f_0) for all species is assumed in order to generate a scaling relationship between the sedimentation coefficient and the diffusion coefficient, of which the subsequent analysis is simulated from. Secondly, the c(s) method assumes the sample is of a non-interacting nature. In using a model-free approach such as c(s), precision in the sedimentation coefficients, and therefore mass of each species, may be lost. The low MW contamination observed in Peak 1 could be accounted for if it were possible to apply a hybrid c(s)-discrete species model approach to this sample. However, as there is no model for the His_6-PRR-UBA protein, the c(s) distribution analysis provides the best resolution and sensitivity (Cole *et al.*, 2008). Therefore, the best fit of the data under these experimental conditions suggests a monomer and two larger tumbling species are contained within these FPLC elutions, where the difference in the MW estimation between predicted and measured may be caused by a distortion in the slower tumbling nature of a non-globular protein, as is hypothesised.

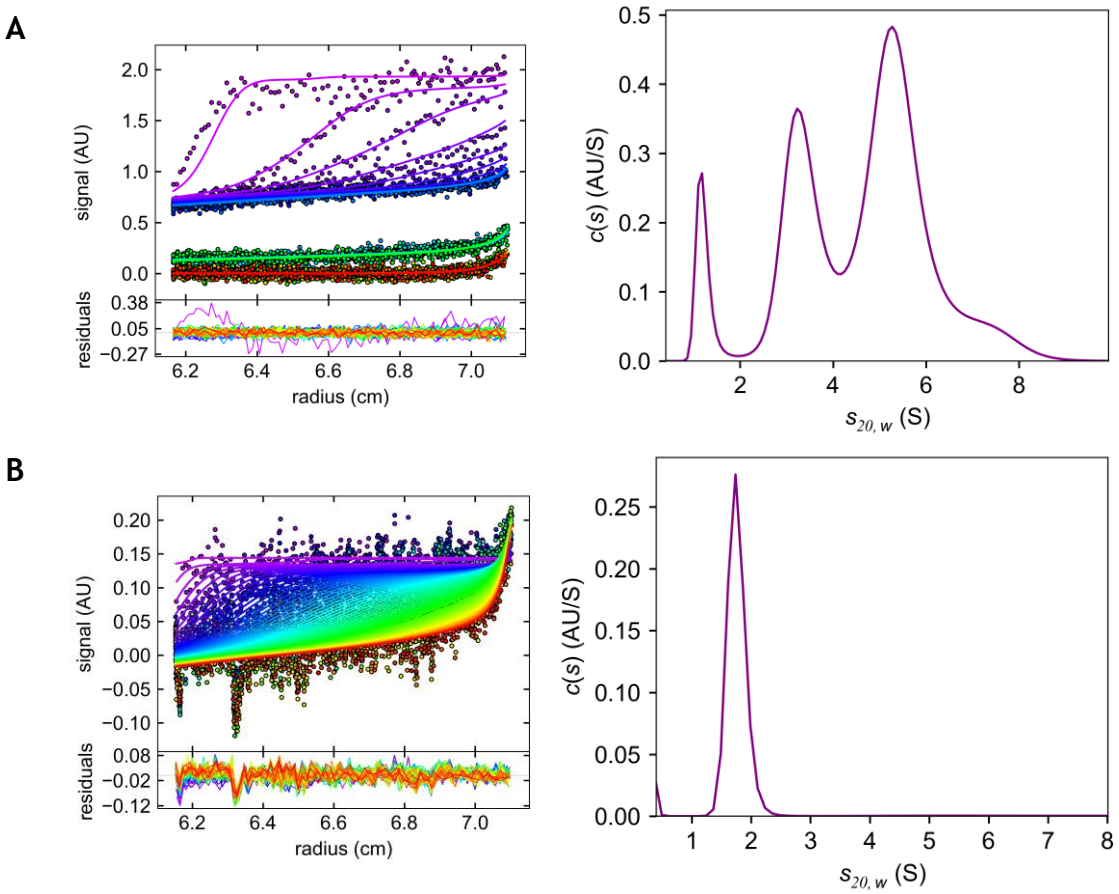
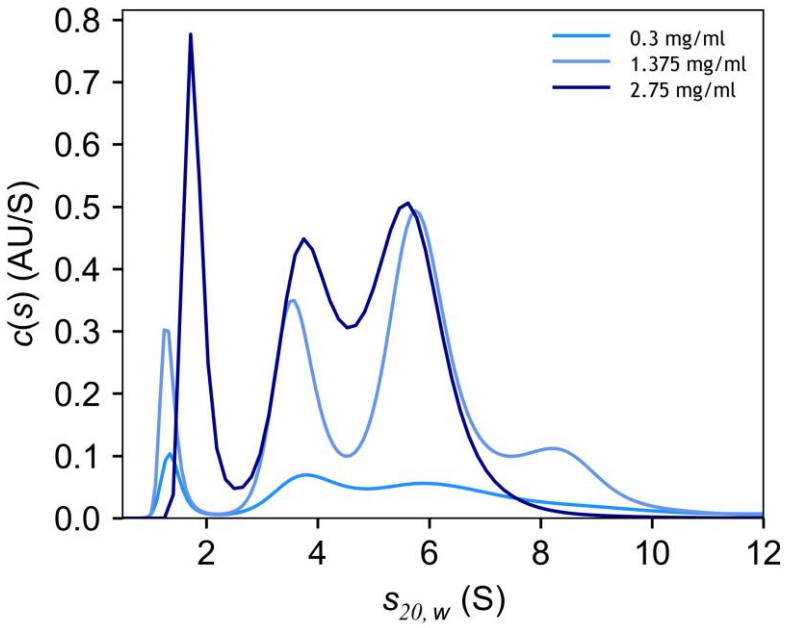


Figure 4.7 Analytical Ultracentrifugation reveals multiple species of His_6-PRR-UBA. Sedimentation Velocity analysis of the purified His_6-PRR-UBA protein. Experimental data is displayed in the left-hand panels alongside the c(s) fit in the right-hand panels. Suggests multiple species in the first FPLC elution (A) compared with a singular species in the second elution of the FPLC (B).

Previous reports have suggested a role for concentration in the trimerization of full length UBQLN2, where trimerization occurred at concentrations above 1 mg/ml (Hjerpe *et al.*, 2016). Therefore, the effect of concentration on multimerization of the His_6-PRR-UBA protein was subsequently investigated, using protein from peak 1. Sedimentation Velocity with six cells was set up to analyse His_6-PRR-UBA peak 1 protein at 0.3, 0.6, 1.375, 2.75 or 5.5 mg/ml concentrations. Unfortunately, due to leaking of sample within the cell only concentrations of 0.3, 1.375 or 2.75 mg/ml were analysed. Sedimentation coefficients (S) were normalised to $S_{20,w}$ values using the parameters outlined in section 2.2.5.3. The comparative c(s) distribution plots are displayed in Figure 4.8A, with the integrated peak values as a proportion of the total signal, along with the weighted S-value, are displayed in Figure 4.8B. The similarity in height distribution between the species and the unchanging $S_{20,w}$ values suggest no role for concentration on the equilibrium of oligomerization, at these experimentally tested values.

A



B

	Peak 1		Peak 2		Peak 3	
Concentration	% Total Signal	$S_{20,w}$ Value (S)	% Total Signal	$S_{20,w}$ Value (S)	% Total Signal	$S_{20,w}$ Value (S)
2.75 mg/ml	18 %	1.8	29 %	3.3	41 %	5.5
1.375 mg/ml	5 %	1.3	16 %	3.6	32 %	5.9
0.3 mg/ml	9 %	1.4	19 %	3.8	30 %	6.3

Figure 4.8 Effect of concentration on oligomerization of His_6-PRR-UBA. SV experiments were performed on the His_6-PRR-UBA protein from the first FPLC elution peak (see Figure 4.5) at either 0.3 mg/ml, 1.375 mg/ml, or 2.75 mg/ml concentrations. The fit of the experimental data as c(s) distributions are shown in (A). Each peak in (A) was integrated and the proportion of the total signal, along with its weighted $S_{20,w}$ value, is shown in (B).

4.4 Investigating the secondary structure of the PRR region

UV Circular Dichroism (CD) is an absorbance spectroscopy technique employed to determine the secondary structure characteristics of proteins and peptides. This is achieved through measurement of the differentially absorbed polarized light by optically chiral compounds. Far-UV (190-260 nm) measurements can provide information on the amide backbone, as the exhibited ellipticity of the peptide bond changes based on the local conformation of the molecule.

The amino acid sequence of the PRR region suggests a polyproline-II helix secondary structure may be possible. Previous reports however have determined the PRR region to be largely disordered (Hjerpe *et al.*, 2016; Dao *et al.*, 2018). To elucidate the propensity for secondary structure formation in the PRR region of the recombinant protein, far-UV CD analysis was carried out using the purified, recombinant, His_6-PRR-UBA used in previous experiments. The monomeric protein was run alongside the His_6-UBA only control. (Figure 4.9).

Far-UV CD spectra of the His_6-UBA protein displays good secondary structure composed largely of α-helices, identified by the characteristic negative peaks at 208 nm and 220 nm. The presence of the PRR region causes a change in the observed CD spectra from the UBA alone (Figure 4.9A). However, the nature of this difference is unclear as proteins with a strong secondary structure (as in the UBA domain) contain the characteristic intense negative peaks at 208 nm and 220 nm which can make visual identification of other structural information more difficult. Therefore, the spectra relating to the His_6-UBA protein was mathematically subtracted from the His_6-PRR-UBA spectra, with the remaining spectra plotted in Figure 4.9B. Following mathematical extraction, the resulting spectra of the PRR-linker region suggests a secondary structure similar to a Polyproline-II (PPII) helix due to the presence of a large negative peak at ~195 nm and a small positive peak at ~220 nm (Lopes *et al.*, 2014). This extracted spectrum was submitted to the DichroWeb server (http://dichroweb.cryst.bbk.ac.uk), which projected the contribution of the PRR-linker to be largely disordered in nature. However, the NRMSD values, which determine the goodness-of-fit, were >0.2. This suggests the analysis failed and any secondary structure is unlikely to be correct. None of the reference sets in the Dichroweb server contain reference spectra for collagen or polyproline-II structures. Collagen and polyproline-II reference spectra have since been established (Lopes *et al.*, 2014) and the addition of these to mainstream servers such as Dichroweb will prove valuable to further investigations.

Polyproline-II helices are common constituents of collagen triple helices (Ramachandran and Kartha, 1954). It was hypothesised that UBQLN2 may form trimeric structures via its PRR region, and these structures could resemble a collagen-like triple helix due to the polyproline-II sequences observed in the amino acid residues and previous CD data (Figure 4.9B). The effects of multimerization

on the secondary structure were subsequently investigated, using the multimeric His_6-PRR-UBA protein from peak one of the FPLC purification. The equivalent mathematical extraction of the PRR-linker was carried out and showed a similar pattern to the monomeric protein (Figure 4.9C). However, visually discerning between collagen triple helices and polyproline-II structures is not possible, as red-shift in peak locations from collagen to polyproline-II are just a few nm different (Lopes *et al.*, 2014). Therefore, further investigation into the structure of the PRR region is required, which is pursued later in this study.

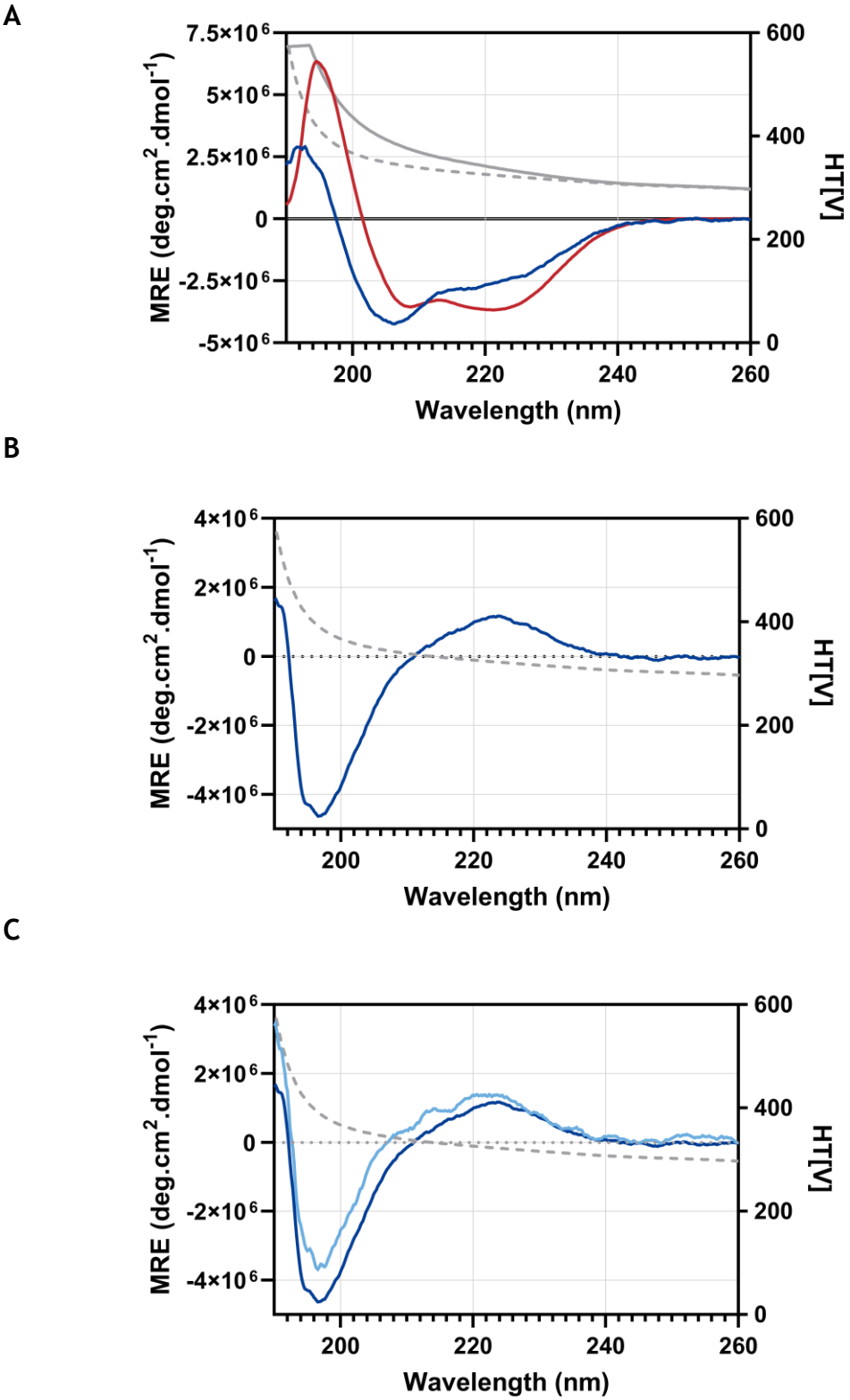


Figure 4.9 Mathematically extracting the spectra of the PRR region reveals a polyproline-II helical structure. (A) Far-UV CD spectra of the His_6-UBA protein (red) in comparison with the His_6-PRR-UBA protein (dark blue). Far-UV CD spectra of the mathematically extracted monomeric (dark blue)(B/C) compared with the multimeric (light blue)(C) PRR-linker region. The corresponding HT curves (grey) show the wavelength cut-off values for the His_6-UBA (solid), monomeric (dashed) and multimeric (dotted) are at around 193 nm, which does not affect measurement of the negative peak.

4.5 Computationally modelling the predicted collagen triple helix structure

The emergence of *AlphaFold* as a tool for predicting protein structure has piqued the interest of many in the biophysical field, allowing quick estimations of protein structure without the long and expensive methods required by the likes of NMR spectroscopy or X-Ray Crystallography (Jumper *et al.*, 2021). However, *AlphaFold* does have its limitations. By calculating structure based on a monomeric form of the protein, some oligomeric structures such as the triple helix formed by Collagen-1, are incorrectly predicted by the programme (Figure 4.10A). *AlphaFold* predicts the structure of monomeric UBQLN2 and its PRR region to be largely disordered, with exceptions at the well-characterised UBL and UBA domains (Figure 4.10B). The prediction of unstructured domains is not uncommon for collagenous proteins (Figure 4.10A) or proteins requiring multimerization for structure and function (Evans *et al.*, 2021), both of which are hypothesised in this study. This would suggest that whilst *AlphaFold* is a leap forwards in structural prediction, it still possesses limitations in deducing structure of known multimeric complexes with the current software. Initially, this study is attempting to find a set of reasonable constraints in which to model the PRR region computationally, using previously solved structures.

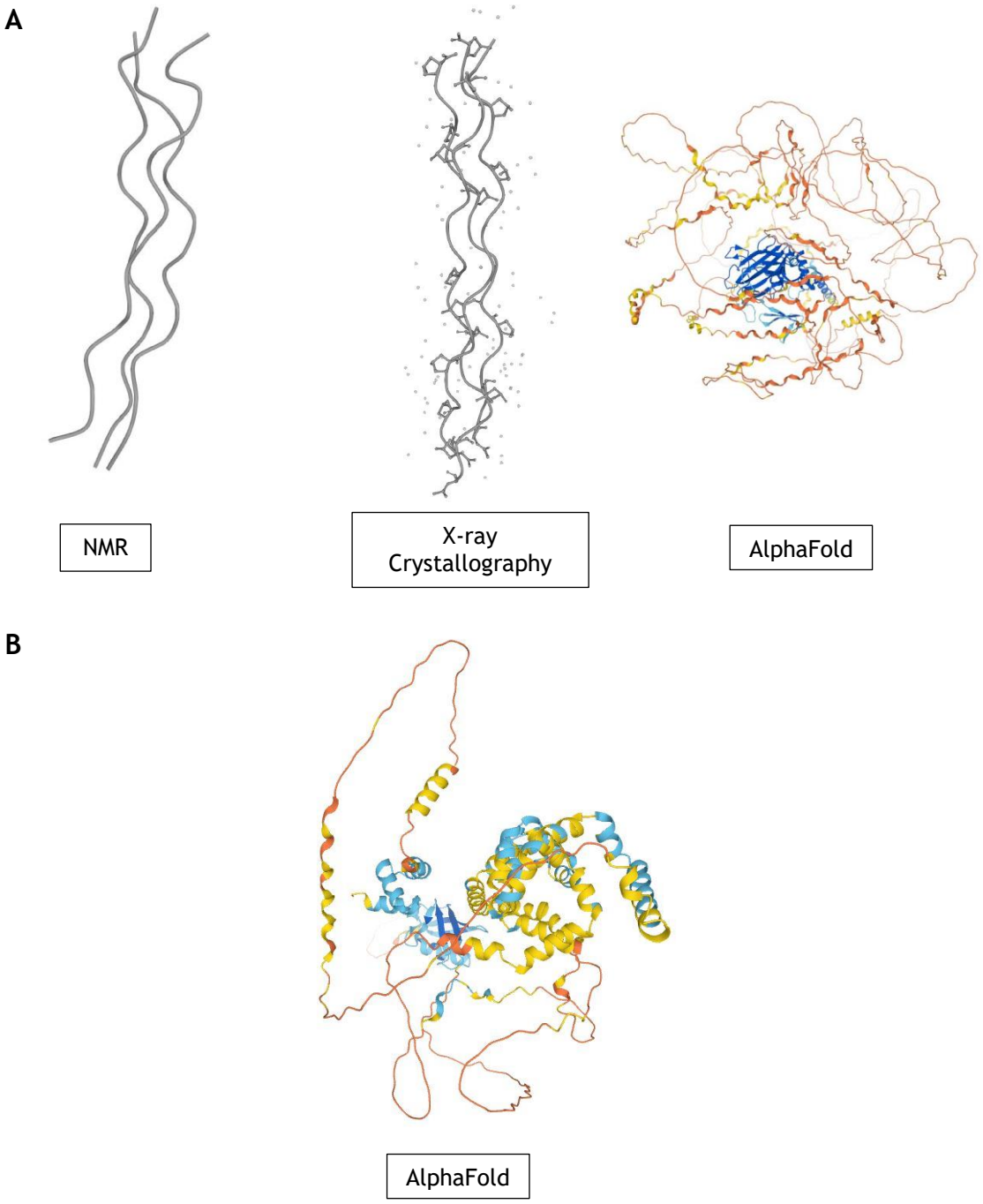


Figure 4.10 AlphaFold incorrectly predicts the structure of Collagen-Iα. Comparison of different structural determinations of the protein Collagen-Iα (A). Experimental methods, either NMR spectroscopy or X-ray crystallography, determine the structure of Collagen-I as a triple helix. In comparison, AlphaFold incorrectly predicts a largely disordered protein. (B) UBQLN2 structure prediction from AlphaFold. PDB files obtained from RCSB or AlphaFold databases respectively (Collagen 1 PDB: 2LLP; 1Q7D, AF-P02452-F1; UBQLN2 AF-Q9UHD9-F1).

4.5.1 Homology Modelling of the PRR region of UBQLN2

To develop a template for which the structure and potential oligomerization of the PRR region could be based upon, proteins of similarity were first identified. Sequence similarities between the PRR region and other proteins were initially

identified using BLAST analysis (see section 2.2.1.1 for parameters used). Hits for UBQLN2 were removed manually, and the remaining hits identified in Table 4.1. These demonstrate numerous examples of sequence similarity between the PRR region of human UBQLN2 and bacterial collagen-like proteins, with almost 60% sequence identity with a query cover >90%. Interestingly, when constraining the BLAST analysis to human alone, one human collagen triple helix hit is produced, Prostate collagen triple helix (PCOTH) (Anazawa *et al.*, 2005). With full query cover and over 40% identity, this is another example of the possibility of the PRR region of UBQLN2 to form collagen-like triple helices.

Table 4.1 BLAST analysis identifies homology between PRR region of UBQLN2 and bacterial collagen-like proteins.

Description	Species	Query Cover	E-Value	Percent Identity
Collagen-like protein	*Glaesserella parasuis*	98%	5e-04	58%
Collagen triple helix	*Clostridium sp*	100%	0.001	59.18%
Collagen-like protein	*Bacillus thuringiensis*	96%	0.006	57.89%
Collagen triple helix repeat-containing protein	*Clostridium aceticum*	98%	0.010	60%
Collagen-like protein	*Clostridiacae bacterium*	94%	0.016	58.33%
Collagen-like protein	*Enterocloster asparagiformis*	96%	0.026	51.02%

To establish sequence homology and alignment, the hits of interest from the BLAST analysis (Table 4.1) were then analysed using the multiple sequence alignment tool PRALINE (see section 2.2.1.1 for parameters). Numerous residues are well conserved between the aligned sequences of the PRR region and bacterial collagen-like proteins (Figure 4.11A). Furthermore, when assessing the homology

between the PRR region and the human collagen protein PCOTH, numerous examples of fully conserved residues are highlighted in red (Figure 4.11B). Together, these data provide the rationale that the PRR region may contribute to the trimerization of UBQLN2 *in vivo*.

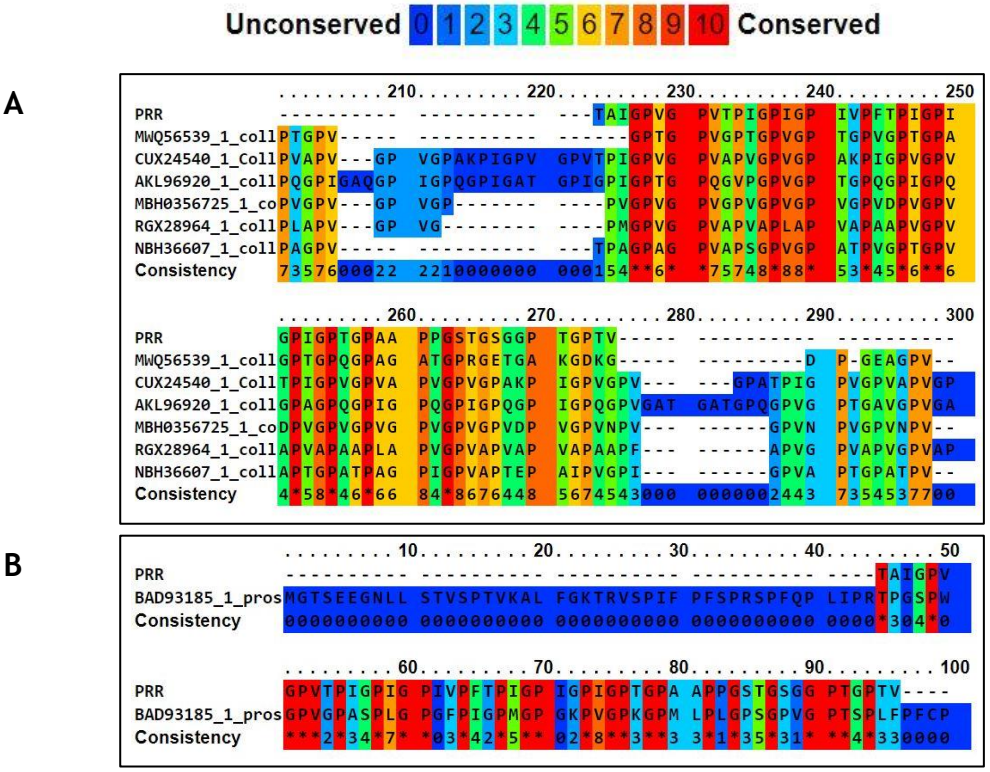


Figure 4.11 Amino acid conservation between the PRR region of UBQLN2 and BLAST hits. BLAST was used to identify hits of similar sequence to the PRR region of UBQLN2. PRALINE multiple sequence alignment was then run using the recommended constraints. Alignment of the PRR region to bacterial collagen-like proteins is displayed in (A) whilst the alignment of the PRR region to the human BLAST hit PCOTH is shown in (B). Regions of conservation are identified with a colour scale ranging from unconserved (dark blue) to conserved (red).

Although PCOTH and numerous bacterial collagen-like proteins hits were obtained in the BLAST analysis, none of these had currently solved structures which could be used as a template for modelling the PRR region.

4.5.2 AlphaFold Multimer

AlphaFold can predict individual protein chains *in silico* by combining amino acid sequence information with multiple sequence alignment (MSA) and homologous structures (Jumper *et al.*, 2021). Following release of the original *AlphaFold* software, a prototype release of *AlphaFold-Multimer* occurred, which allows *in silico* prediction of multi-chain protein complexes (Evans *et al.*, 2021). Previous approaches to multimeric structure prediction have centred around template-

based modelling or free docking (Chen, Li and Weng, 2003; Guerler, Govindarajoo and Zhang, 2013; Kozakov *et al.*, 2017). In *AlphaFold-Multimer,* a deep learning model is trained specifically on oligomeric data, resulting in significant improvement of the accuracy of multimer prediction.

A restricted, free-access version *AlphaFold-Multimer* was employed to predict the structure of the PRR region of UBQLN2. Modelling the PRR region as a trimer did not produce a model that would appear to lend itself to a triple helical structure (Figure 4.12A). However, despite its improvements upon the earlier *AlphaFold* programme, *AlphaFold-Multimer* only accurately predicts 34% of homomeric interfaces (Evans *et al.*, 2021). Therefore, whilst a useful tool to employ, the data generated cannot be trusted until the structure is verified experimentally.

It is understood that numerous collagen triple helical proteins require an independent, non-collagenous trimerization site to overcome the slow kinetics of collagen folding (Engel and Prockop, 1991). In combination with the data displayed previously in this study, it could be possible that such nucleation site necessary for trimerization was lost in the original design of the truncated protein. The natural trimerization domain of T4 fibritin, named foldon, has been used by others to bioengineer triple helical structures (S. P. Boudko *et al.*, 2002; Kaur *et al.*, 2015) and has been shown to function in a similar way to the non-collagenous globular nucleation domains (S. Boudko *et al.*, 2002). Adding foldon to the N-terminus of the PRR-UBA sequence and modelling in *AlphaFold-Multimer* generated a prediction suggesting the PRR region has a propensity to form a triple helical structure when bioengineered as a trimer (Figure 4.12B).

Collagen triple helices contain a pattern of interhelical hydrogen bonding at each cross-section of the helix core, often mediated by the Glycine residues occupying the tight turn (Bella *et al.*, 1994). The hydrogen bonding between the strands of the *AlphaFold-Multimer* prediction was investigated. No evidence of hydrogen bonding was evident from the *AlphaFold-Multimer* prediction when analysed in Chimera X. However, these tools are only predictions with limitations already outlined previously (Figure 4.10), and so results should be concluded with caution.

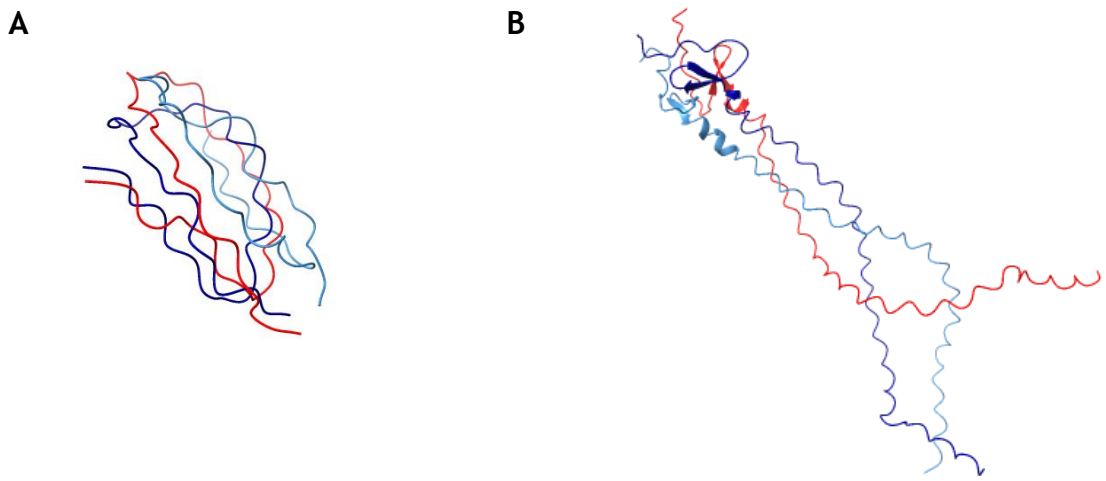


Figure 4.12 AlphaFold-Multimer predicts a propensity of the PRR region to form a helix if existing as a trimer. AlphaFold-Multimer predictions of the PRR region alone (A) or the PRR region with the Foldon N-terminally (B). Each monomeric unit is identified in a separate colour, whether red, dark blue or light blue.

4.6 Chapter Discussion

The role of the PRR region in UBQLN2's function as a ubiquitin receptor, or any other mechanism, has remained elusive for many years. Whilst numerous papers have attempted to infer the role of the PRR region through deletion (Hjerpe *et al.*, 2016) or mutation experiments (Hjerpe *et al.*, 2016; Teyssou *et al.*, 2017; Dao *et al.*, 2019), none have directly attempted to characterise the structure and infer the function of this region until now. In fact, research by Dao et al., 2018 used the UniProt-defined Pxx-UBA as a non-oligomerizing control in the investigation of LLPS. This chapter has found contradicting evidence that the PRR-UBA region can form oligomers. Further investigation into the LLPS properties of ALS-mutated UBQLN2 revealed that whilst some mutations increased self-assembly, numerous did not and in all cases this oligomerization was reversible (Dao *et al.*, 2019). This would suggest mechanisms outside of LLPS play a role in disease development. Whilst often overlooked due to its predicted disorder, it is fair to presume the PRR region holds a vital role in the function of UBQLN2 due to the presence of numerous disease-causing mutations located within this region (Deng *et al.*, 2011; Teyssou *et al.*, 2017). This chapter has focussed on identifying whether recombinant expression of the PRR region and UBA domains of UBQLN2 is possible in isolation, the conditions best suited to this expression in *E. coli* and the possibility of multimeric species forming due to the presence of the PRR region.

The results presented in section 4.2 demonstrate the viability of expressing and purifying the PRR-UBA protein from *E. coli* as a reproducible method for prolonged study. Ample protein is reliably produced following the protocol developed, allowing many avenues of investigation to occur, both within this study and in future investigation.

Initial investigation into the oligomeric states of the PRR-UBA protein corroborated existing evidence that UBQLN2 can form multimeric structures (Hjerpe *et al.*, 2016; Dao *et al.*, 2018). Previous reports have assigned this oligomerization to residues outside of the PRR region (Alexander *et al.*, 2018; Dao *et al.*, 2018) or have not commented on the necessity of the PRR region to accomplish multimerization (Hjerpe *et al.*, 2016). The result in Figure 4.5 and Figure 4.6 demonstrates the possibility of multimerization of the PRR-UBA protein, likely driven by the PRR region. Whilst limitations regarding the mechanism by which FPLC and Native PAGE operate have been discussed (see section 4.3.1), this chapter has revealed numerous, differentially migrating species, the larger of which demonstrate a resistance to SDS denaturation. This SDS resistance could be suggestive of a stable multimeric protein (Tasab, Jenkinson and Bulleid, 2002; Papanikolopoulou, van Raaij and Mitraki, 2008) or potential aggregation. Whilst the large MW estimation from the FPLC may support this possibility, the results from the AUC experiments do not. If these SDS resistant species were in fact aggregation, clean peaks of consistent S values would not be produced. Instead, upward inflections at 100 S would be observed which did not occur in this study. Therefore, the SDS resistant species of His_6-PRR-UBA are likely higher order species, as opposed to simple aggregation.

To provide increased clarity on the oligomeric profile of these species, Sedimentation Velocity was employed. This revealed two distinct species and low MW contamination within the first peak of the FPLC (Figure 4.7). Previous studies have reported full-length UBQLN2 dimerization and trimerization occur in a concentration-dependant manner, but no higher order oligomer was observed (Hjerpe *et al.*, 2016). Whilst limited in the range of concentrations tested, the investigations into the effects of concentration on oligomerization of the His_6-PRR-UBA revealed no role for concentration in multimerization, a contrast to current reports. This would suggest a region outside of the PRR-UBA governs the concentration-dependant manner of oligomerization, possibly the STI1-II domain

as reported by Dao et al., 2018. However, the results in this study do corroborate the notion that no higher order species or aggregations were formed (Hjerpe *et al.*, 2016). AUC is an extremely beneficial tool for characterising solutions of macromolecules in their native state under biologically relevant conditions. Using three optical systems for measurement (absorbance, interference and fluorescence), precise and selective observation of sedimentation can occur in real time. However, there are still a number of assumptions required in order to apply AUC to a previously uncharacterised protein. The assumptions involve the shape and size of the molecule (see section 4.3.2) which may impact on the validity of the results generated. Despite these caveats, all results presented in this chapter point to the possibility of multimerization directed by the PRR region, which warranted further investigation.

The amino acid sequence of the PRR region (Figure 4.1), would predict a secondary structure similar to that of a polyproline-II helix, despite the consistent prediction of disorder in computational modelling (Hjerpe *et al.*, 2016; Dao *et al.*, 2018). Polyproline-II helices are common components of collagen triple helices (Ramachandran and Kartha, 1954). Historically, it has been challenging to distinguish polyproline-II and collagen structures from that of disordered or unfolded proteins (Woody, 1992). However, work by Lopes et al., 2014 has identified two diagnostic features in distinguishing polyproline-II structures from disordered regions. These are the presence of a large, negative peak at ~195 nm and a small, positive peak at ~220 nm, the latter of which is not present in unfolded or disordered proteins. When the contribution of the PRR region is mathematically extracted from the far-UV spectra the resulting spectra is characteristic of a polyproline-II helix. This result contradicts numerous reports that the PRR region displays no ordered structure (Hjerpe *et al.*, 2016; Dao *et al.*, 2018). These misinterpretations in previous studies are not surprising; there is increasing evidence that the contribution of unfolded states is significantly smaller than what is often predicted and assumed, due to the presence of unidentified polyproline-II structures (Rucker and Creamer, 2002; Lopes *et al.*, 2014). Few reference sets for computationally predicting or analysing the structure of “disordered” regions include reference spectra for polyproline-II helices. Instead, if the structure does not conform to α-helical or ß-strand, it is termed “random coil” or “disordered”. Polyproline-II helices are present in ~2 % of all proteins

(Adzhubei and Sternberg, 1993; Stapley and Creamer, 1999; Jha *et al.*, 2005; Berisio *et al.*, 2006). At an individual amino acid level, the conformation to a polyproline-II helix is comparable to that of a ß-strand conformation (Adzhubei *et al.*, 1987). This data challenges the notion that "disordered" regions have randomly distributed amino acid residues in the sterically permitted regions of the Ramachandran plot. Instead, some of these "disordered" regions contain polyproline-II helices which were previously indistinguishable from genuine disordered regions. The results generated in Figure 4.9 support this argument.

In expanding the region of interest from the pre-defined Pxx domain to the PRR region, a bacterial collagen-like amino acid sequence was revealed (Figure 4.1). Homology modelling established sequence similarity between the PRR region and numerous bacterial collagen-like proteins. Less is known about bacterial collagen-like proteins in comparison to mammalian, but emerging evidence suggests differential stabilisation mechanisms are in play, (Rasmussen, Jacobsson and Björck, 2003; Mohs *et al.*, 2007; Berisio *et al.*, 2009). There are other examples in eukaryotes of collagen stabilisation occurring as a result of mechanisms other than hydroxylation of the second proline in the triplet, such as stabilization of the cuticle collagen of the vent worm species via Threonine glycosylation (Mann *et al.*, 1996) or hydroxylysine glycosylation of Collagen 1 in vertebrates (Yamauchi and Sricholpech, 2012). Therefore, it is possible that the PRR region could form a triple helix, despite not possessing hydroxyprolines within its sequence for stabilisation.

Combining the data presented in this chapter, a revised hypothesis was proposed: UBQLN2 trimerizes via its polyproline-II helices contained within the PRR region to form a collagen like triple helix. This may improve binding affinity of ubiquitin, confer specificity of the UBA domains for particular linkage lengths or types, or improve efficiency through the degradation system.

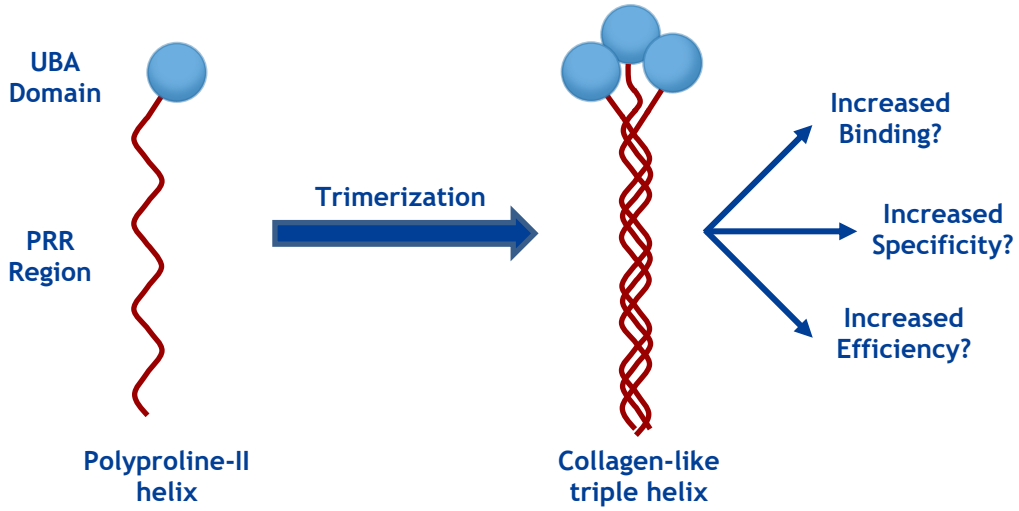


Figure 4.13 A refined hypothesis for the trimerization of UBQLN2. Schematic depicting the polyproline-II helix of the PRR region forming a collagen like triple helix structure and the potential consequences of this structure being formed.

Numerous collagen-like proteins and collagen triple helices are flanked by non-collagenous nucleation domains (Engel and Prockop, 1991; S. P. Boudko *et al.*, 2002). The formation of the collagen triple helix alone is relatively slow due to the high activation energy required (80kJ/mol) (Reimer *et al.*, 1998). As such, numerous trimerization regions have been identified which progress this folding event at a faster rate (Bächinger et al., 1980; Dölz, Engel and Kühn, 1988; Boutaud et al., 2000; Engel and Bächinger, 2000). Initially, the decision was made to remove previously established oligomerization sites (Alexander *et al.*, 2018; Dao *et al.*, 2018) from the protein of interest, as no role was identified for the PRR region in oligomerization in these reports. It is possible however, that in truncating the start of the protein tightly to the PRR region, a nucleation site was lost. This loss of nucleation site may have prevented *AlphaFold-Multimer* from predicting a structure for the PRR region. Therefore, the trimerization domain of the bacteriophage T4 fibritin, named Foldon, was included N-terminally to the PRR region. With the addition of a non-collagenous nucleation site, the PRR region displayed a propensity to form a triple helix structure (Figure 4.12B). To truly ascertain the structure of the trimeric PRR region, further biophysical investigation is required, which this study progresses on to.

Initial evidence of multimerization via the PRR region has been presented in this chapter through the use of FPLC, Native PAGE and AUC analysis. Furthermore, a

previously unreported secondary structure for the PRR-linker region as a polyproline-II helix has been determined, which is a common component of the collagen triple helix. Homology modelling revealed sequence similarity to numerous bacterial collagen-like helices and a human collagen protein PCOTH. It was therefore hypothesised that the multimerization observed earlier could be due to the individual polyproline-II helices forming a collagen-like trimer via the PRR region. However, distinguishing between polyproline-II and collagen triple helix CD spectra is near impossible. Therefore, other structural methods were required. Solution Nuclear Magnetic Resonance (NMR) spectroscopy was the chosen technique to employ as it allows determination of structural characteristics in modifiable environments as well as determining binding affinities of ligands. To account for the possible loss of a non-collagenous nucleation site and to ensure reliable isolation of trimeric protein, it was decided an obligate trimer should be engineered for downstream investigations.

5 Engineering an Obligate Trimer

5.1 Designing an Obligate Trimer

5.1.1 Rationale

Data collected in the previous chapter has led to the evolution of a hypothesis where UBQLN2 trimerizes via its polyproline-II helix located in the PRR region to form a more structured collagen-like triple helix. A previous study has revealed dimerization and trimerization of full-length WT UBQLN2 in a concentration dependant manner (Hjerpe *et al.*, 2016), but no higher order species were observed. The inability to gain enough resolution on the SEC column to separate the multimeric species and identify their true size (Figure 4.5) was identified to be problematic in ongoing experiments. To gain an understanding of the structure of the PRR region, NMR Spectroscopy was to be employed. However, a pure, singular species of high concentration is required for this technique. Given previous studies did not identify numerous large species, it is possible that in designing the truncated PRR-UBA a nucleation site was lost. There are multiple reports of collagen-like triple helices being flanked by non-collagenous domains often acting as a trimerization site (Bächinger et al., 1980; Boutaud et al., 2000; Snellman et al., 2000). It has already been reported that the STI1-II region, N-terminal to the PRR region, acts as a director of liquid-liquid phase separation (LLPS) and oligomerization (Alexander *et al.*, 2018; Dao *et al.*, 2018). This region was originally removed to prevent any LLPS capabilities interfering with assessment of oligomerization via the collagen-like sequence. Due to the high sequence identity shared between the PRR region and numerous bacterial collagen-like proteins (see Table 4.1), it was decided the hypothesis of trimerization via the PRR region should be investigated further. To accommodate the loss of a potential nucleation site, engineering of an obligate trimer by adding a constitutive trimerization domain N-terminally to the PRR region was identified as the next step in elucidating whether the PRR region adopts a structure not yet reported in the literature. This chapter will outline the considerations made when designing an obligate trimer and identify the successes and limitations of investigating proteins in such a way.

5.1.2 Design Strategy

Fibritin is an elongated, trimeric protein of the bacteriophage T4 head with structural, chaperone and sensory functions (Conley and Wood, 1975; Coombs and Eiserling, 1977; Terzaghi, Terzaghi and Coombs, 1979). Fibritin is composed of three key regions: an N-terminal anchor domain, a central coiled-coil region, and a C-terminal globular domain (Tao *et al.*, 1997). Within the C-terminus lies a trimeric ß-sheet propellor consisting of monomeric ß-hairpin segments. It is this ß-propellor which is necessary for correct folding of the trimeric protein (Letarov *et al.*, 1999), leading to its naming as foldon.

The trimerization domain of T4 fibritin, named foldon, was modelled in *AlphaFold-Multimer* as a nucleation domain for the PRR region (Figure 4.12). This modelling revealed a propensity of the PRR region to form a triple helix if existing as a homotrimer. There are numerous reports of utilising foldon to successfully engineer trimeric protein structures (Frank *et al.*, 2001; S. P. Boudko *et al.*, 2002; Meier *et al.*, 2004; Papanikolopoulou, van Raaij and Mitraki, 2008). Residues G457 to L483 were utilised as the trimerization domain in this study as an isolated trimeric foldon structure is sufficiently formed with these residues (Güthe *et al.*, 2004).

Foldon is naturally located at the C-terminus of Fibritin, a protein of the bacteriophage T4 (Letarov *et al.*, 1999). A number of collagen proteins trimerize via a C-terminal nucleation domain (Bächinger et al., 1980; Dölz, Engel and Kühn, 1988; S. Boudko et al., 2002). Therefore, protein engineering of collagens often attaches foldon at the C-terminus of the protein of interest (Frank *et al.*, 2001; S. Boudko *et al.*, 2002; Bhardwaj *et al.*, 2008), but triple helix formation can be nucleated from either end (Frank *et al.*, 2003). Due to the presence and interest of the C-terminal UBA domain of UBQLN2 in downstream applications, the addition of foldon at the C-terminus was undesirable. N-terminal attachment of foldon to collagen triple helices is uncommon, with one current example of N-terminal nucleation of collagen fragments existing in the literature (Frank *et al.*, 2003). When using foldon to engineer other protein fibres, it has been reported the structures formed are less stable and lack SDS resistance in comparison to C-terminally attached foldon chimeric proteins (Papanikolopoulou *et al.*, 2004;

Bhardwaj *et al.*, 2008). However, these N-terminal chimeric proteins did not have an extended linker region included in the design strategy (Papanikolopoulou *et al.*, 2004), as was the case for the comparable C-terminal protein, or contained foldon domains at both the N- and C-termini (Bhardwaj *et al.*, 2008). These studies did not pursue investigations into the N-terminally attached foldon further. Therefore, it is not possible to conclude whether the lack of stability and SDS resistance was due to a loss of trimeric folding entirely or instead production of more sensitive trimers. It is also unclear whether increasing the linker region or attaching foldon to N-terminally folding collagen structures would provide the necessary flexibility for foldon to act as the registration motif. Nonetheless, upon investigation into the PDB structure of foldon, it was concluded the protein chain ends would lie within close enough proximity to form a trimer. The resulting His_6-foldon-PRR-UBA protein was termed HFPRR in subsequent experiments for simplicity.

A second obligate trimer was designed utilising a different trimerization domain. CC-pII is composed of a coiled coil structure in which the hydrophobic surface formed by the a and d residues is designed to give rise to a specific oligomerization state. In the case of CC-pII, isoleucine (I) occupies both the 1st and 2nd hydrophobic position in the *h-x-x-h-x-x-x* heptad repeat (where *h* is a hydrophobic residue), giving rise to a trimeric state (Harbury *et al.*, 1993; Fletcher *et al.*, 2012). CC-pII can function as a trimerization domain for bacterial collagen at either the N- or C-termini (Yoshizumi *et al.*, 2011). N-terminal trimerization was preferable in design but would potentially be more difficult to discern between the structures of the coiled-coil region and any triple helices formed in downstream investigations. The His_6-CCpII-PRR-UBA obligate trimer was named similarly to its counterpart using foldon, as HCPRR. A schematic outlining the design strategy for both obligate trimers is displayed in Figure 5.1.

Similar to Chapter 4, the classical recombinant bacterial expression and purification system was employed with a His_6- tag to allow column purification of the expressed protein of interest from the lysate (see section 2.2.4.2). In contrast to Chapter 4, a recognition site for the serine protease Thrombin was used in place of the viral TEV protease. This change was made as despite the high specificity of cleavage by TEV, the turnover rate is ~100 fold lower than Thrombin (Waugh, 2011), resulting in higher amounts of protease required. As mentioned previously,

the desired final downstream experimental technique was NMR spectroscopy, which requires high concentrations of >95% pure isotopically labelled protein. As Thrombin did not cleave the His_6PRR-UBA non-specifically, it was decided this would be the most efficient way to ensure quick and full cleavage.

Collagen helices fold with a leading, middle, and lagging strand to finally form the stable collagen triple helix (Fallas *et al.*, 2012), with some disorder occurring at the termini (Schumacher, Mizuno and Bächinger, 2006). A linker of variable length was designed to connect the trimerization domain to the PRR region, allowing flexibility in the folding of the region, thus allowing the correct register to occur. The design strategy for both obligate trimers involved nesting restriction enzyme sites within the DNA sequence to allow alteration of the linker length (Figure 5.1B). At its full length, the flexible linker is eight amino acids in length (GSGTGTGS). Cutting with KpnI would produce a linker of six amino acids (GSGTGS) whilst cutting with BamH1 would produce just two (GS). Restriction enzymes encoding flexible amino acids Glycine, Serine and Threonine were chosen as these amino acids are classical linker choices to provide the maximum flexibility (Karplus and Schulz, 1985) to allow correct folding of the strands, should a triple helix form.

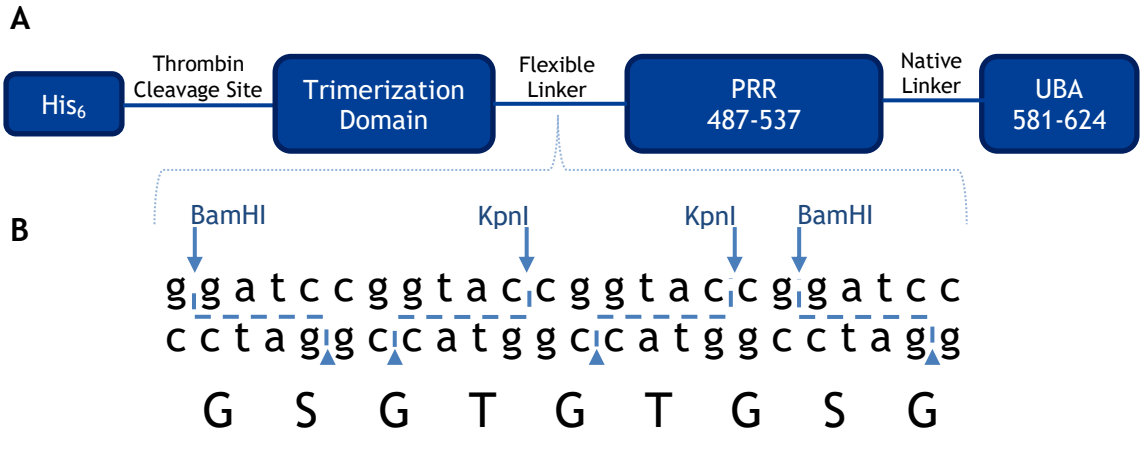


Figure 5.1 Protein map of the engineered obligate trimers. (A) Schematic of the engineered obligate trimer proteins, where the trimerization domain correlates to either foldon or CCpII. (B) DNA and amino acid sequence of the flexible linker between the trimerization domain and the PRR region. The nested restriction enzymes required to alter the length of the linker if desired are also annotated. Note the schematics are not drawn representative to actual contribution in the amino acid chain.

5.1.3 Proof of Concept

The optimised expression conditions identified in section 4.2 were applied to the new proteins. Excellent expression (Figure 5.2A) and subsequent purification (Figure 5.2B) of both the HFPRR and HCPRR recombinant proteins was achieved. Thrombin cleavage was optimised for time (3 hours, 6 hours or overnight) and ratio of Protein:Thrombin (3:1, 1:1 or 1:2) with 3 µg sample protein (Figure 5.2C) at 22ºC.

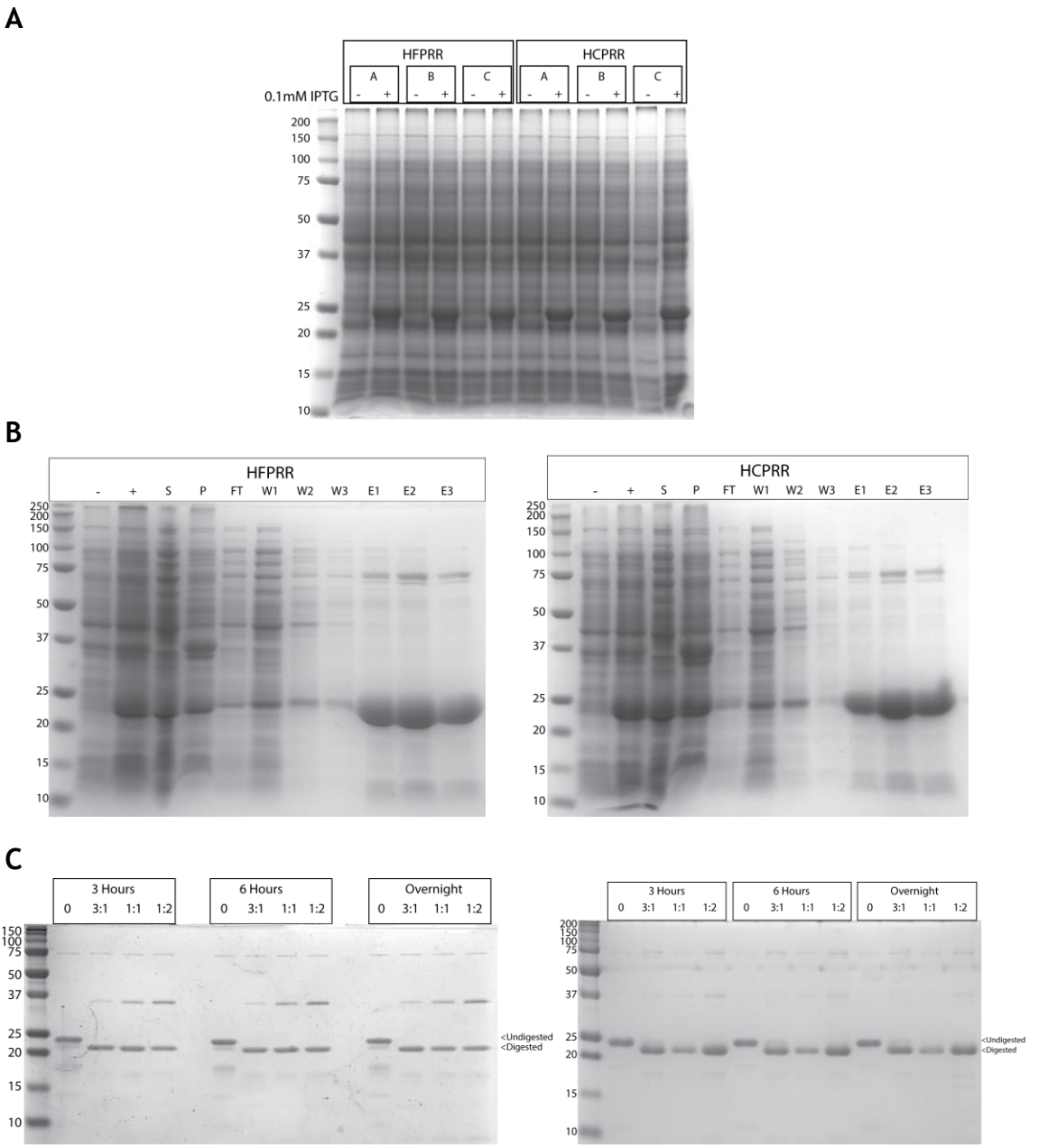


Figure 5.2 Expression, purification and cleavage of the obligate trimers is sufficient. Sufficient expression (A), purification (B), and cleavage with thrombin (C) of the HFPRR (Left hand panel) or HCPRR (right hand panel) assessed by SDS PAGE analysis and Coomassie Instant Blue staining. Notations for (B) are as follows: Notations as follows: Addition of 0.1mM IPTG is denoted by +/-, supernatant (S) or pellet (P) fractions following lysis and clarification, flowthrough (FT) of the supernatant after adding to the Ni-NTA column, 3 washes (W1/W2/W3) and 3 elutions (E1/E2/E3). Thrombin cleavage (C) was tested with 3 ug protein and 0.001 units (3:1), 0.003 units (1:1), or 0.006 units (1:2) for varying lengths of time.

The addition of a trimerization domain should produce a singular species if the design strategy was successful. To confirm the number of species present in the new proteins, FPLC using a size exclusion column (SD200i) was employed. FPLC traces demonstrate both HCPRR and HFPRR recombinant proteins produce a singular species of approximately 167 kDa and 144 kDa respectively (Figure 5.3A). *ProtParam* predicts a MW of the monomeric proteins as 19 kDa for both HCPRR

and HFPRR. Therefore, a trimeric protein would have a predicted MW in the region of 57 kDa. Whilst these proteins are larger in experimentally estimated MW, the same drawbacks as outlined in section 4.3.1 apply, where estimations of MW are based upon calibration of globular proteins through the matrix. As such, FPLC alone cannot be used to determine MW. Nonetheless, this result indicates a singular species was successfully produced for both obligate trimers.

SDS denaturing has been used as an indicator of stability in collagen (Tasab, Jenkinson and Bulleid, 2002) and foldon trimeric structures (Papanikolopoulou *et al.*, 2004; Bhardwaj *et al.*, 2008). The stability of the obligate trimer species in 2 % (w/v) SDS was investigated with Native PAGE analysis and Coomassie Instant Blue Staining. An attempt was made to compare the obligate trimers to the multimeric His_6-PRR-UBA protein from section 4.2. However, the protein had degraded beyond comparable levels. Figure 5.3B demonstrates the structures formed by HFPRR or HCPRR are resistant to SDS denaturing but not 2 M Urea or boiling, suggesting stable trimeric structures are formed.

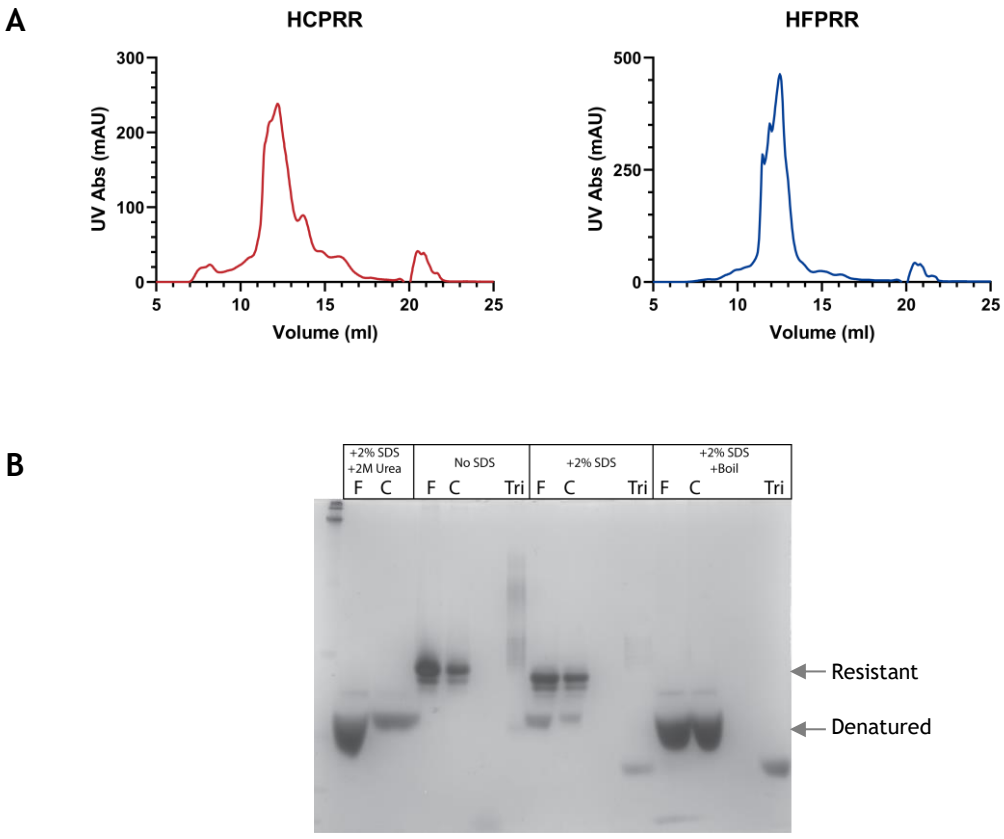


Figure 5.3 The obligate trimers produce singular species with SDS resistance. (A) FPLC was performed using a SEC column (SD200i) of the HCPRR (red) or HFPRR (blue) proteins. (B) Native PAGE analysis comparing the HFPRR (F) and HCPRR (C) resistance to various denaturing techniques. Tri denotes the His_6-PRR-UBA multimeric protein identified in Figure 4.5, which degraded prior to experimentation.

For efficiency, only one obligate trimer was investigated further. Both HFPRR and HCPRR recombinant proteins expressed, purified, and cleaved well (Figure 5.2A). Furthermore, both obligate trimers produce singular species of similar sizes which demonstrate a resistance to SDS denaturing (Figure 5.2). As mentioned previously, it would be more difficult to discern between the coiled coil structure of CCpII and any triple helical structures of the PRR region should they form. In comparison, the ß-propeller structure of foldon and the published NMR assignments ensure simpler data analysis in downstream experiments. Therefore, HFPRR was chosen as the protein to pursue in subsequent investigations.

5.2 Investigating the biophysical properties of the Obligate Trimer

5.2.1 Circular Dichroism

The addition of foldon to previously engineered proteins had little impact on their structure and function (Frank *et al.*, 2001; S. P. Boudko *et al.*, 2002). To establish whether foldon had any impact on the secondary structure of the PRR-UBA regions, Circular Dichroism was again employed.

Far-UV CD spectra were measured from the HFPRR recombinant protein (conditions outlined in section 2.2.5.2), with the His_6-UBA contribution mathematically removed using the spectra from Figure 4.9A. The resulting spectra of the foldon-Linker-PRR-Linker are displayed in Figure 5.4, compared with the spectra of the PRR-Linker measured in Figure 4.9B. The increased size and slight red shift of the positive peak ($1x10^6$ vs $1.2x10^6$ $deg.cm^2.mol^{-1}$) is likely due to foldon contribution (Frank *et al.*, 2001; Papanikolopoulou *et al.*, 2004). Intriguingly, the negative peak in the region of 195 nm has shifted toward 193 nm and sees a larger drop in value (from $-4.6x10^6$ to $-5.6x10^6$). A blue shift with a larger negative peak is more characteristic of collagen spectra in comparison to polyproline-II helix spectra (Lopes *et al.*, 2014). This data may suggest the polyproline-II helices identified in the monomeric PRR region (Figure 4.9) could be forming a collagen triple helical structure when expressed as an obligate trimer. However, it is not possible to confirm with certainty whether the spectra indicate polyproline-II helices or a more structured collagen triple helix when existing as a trimer. As the majority of available servers do not include polyproline-II or collagen triple helices in their reference spectra, it was also not possible to submit these spectra for analysis. Therefore, different experimentation is required to further elucidate the structure of the trimeric PRR region. Nonetheless, the mathematically extracted spectra in Figure 5.4 suggests the presence of the foldon as a trimerization domain does not perturb the polyproline-II secondary structure elucidated previously in section 4.4. Furthermore, it suggests a possible development of the structure from a precursory polyproline-II helix into a collagen triple helix.

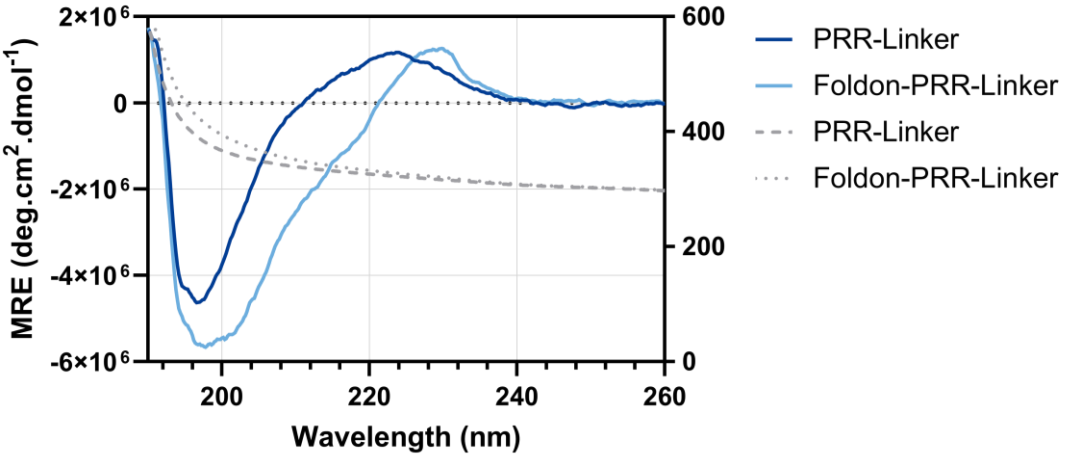


Figure 5.4 The presence of foldon does not impact the secondary structure of the PRR region. Far-UV CD spectra of the mathematically extracted foldon-PRR-Linker contribution of the HFPRR recombinant protein (light blue) in comparison to the previously determined (Figure 4.9B) PRR-Linker spectra (dark blue). The corresponding HT curves (grey) show the wavelength cut-off values for the HFPRR (dotted) and PRR-Linker (dashed) at around 193 nm, which does not affect measurement of the negative peak.

5.2.2 AUC Analysis of the Obligate Trimer

Previous results in this study (Figure 4.5) suggested two distinct species His_6-PRR-UBA when conducting FPLC analysis. Subsequent AUC analysis with Sedimentation Velocity (SV) experiments revealed the first of these peaks contained multiple species (Figure 4.7). To confirm the number of species present in the sample and investigate its oligomeric profile, SV was again employed.

SV experiments were run on HFPRR recombinant protein. A singular species with an $S_{20,w}$ value of 2.2 S was identified (Figure 5.5A). Estimation of molecular weight from this would produce a value of approximately 65 kDa, which is comparable with the 3x MW predicted by ProtParam (MW=19 kDa; 3MW=57 kDa). Sedimentation Velocity experiments rely on measurement and prediction of a number of forces. Broadly speaking, the centrifugal force is countered by the buoyancy of the molecule in the solvent to produce a net force. These forces are governed by relationships between the mass of the solvent and particle, the partial specific volume of the particle and the density of the solvent. A further consideration, however, is the frictional force as the particle moves through the solvent, which is affected by the macromolecule's shape (Cole *et al.*, 2008). If previously established, frictional ratio can be an inputted parameter, fixed to aid

in the fitting of the experimental data. However, in the case of UBQLN2 and its PRR region, no such information or model exists yet. Therefore, the frictional ratio can be calculated using a non-linear regression during analysis, with the experimental best fit frictional ratio becoming an output of the experiment. A frictional ratio of 2.4 was calculated in Sedfit for the HFPRR protein. This would suggest the species is asymmetrical; in the case of a globular protein the frictional ratio would be ~1.2 (Zhao *et al.*, 2013). This asymmetry could infer an elongation of the protein (Chaton and Herr, 2015). This supports the hypothesis that the PRR region forms a rod-like extension on the UBA domain, which may be more structured if a collagen-like triple helix occurs when expressed as a trimer.

Whilst a useful tool, sedimentation velocity has its drawbacks. Namely, the estimation of MW is dependent on the shape of the macromolecule in question. With proteins where the structure has already been determined, this dependency on shape can be accounted for using modelled data analysis (see section 4.3.2). Sedimentation Equilibrium (SE) involves identifying the centrifugal speed at which the sedimentation flux of the molecule is equal to that of diffusional flux, resulting in equilibrium concentration distribution of macromolecules (Cole *et al.*, 2008). Ultimately, SE allows shape-independent measurement of the molecular weight.

The MW of HFPRR was measured in the experimental conditions outlined in section 2.2.5.3 using SE with the experimental data shown in Figure 5.5B. SE data fitted with a single species model indicated the presence of a species with a mass of 53 kDa, with a root mean squared deviation (R.M.S.D) value of 0.01. This indicates the experimentally determined mass is of reasonable accuracy.

Together, the results generated in the AUC experiments confirm a singular species of a potentially elongated nature. SV estimates the apparent mass of this species as 65 kDa, where the more accurate SE estimation places MW at 53 kDa. This apparent MW is comparable with 3x the *ProtParam* predicted MW (57 kDa). It is therefore concluded that a successful obligate trimer using foldon as the trimerization domain has been recombinantly produced. Finally, AUC results suggest an elongation in the structure, which supports the current hypothesis described in section 5.1.

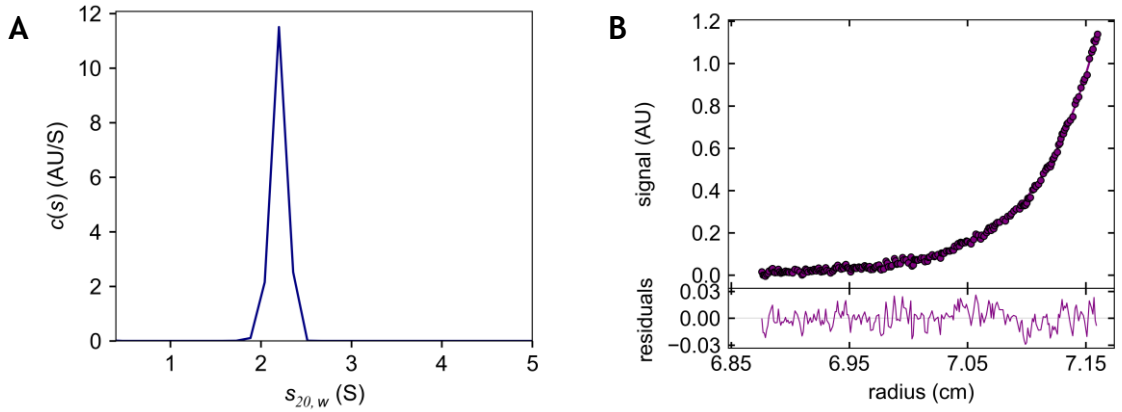


Figure 5.5 AUC analysis of HFPRR. SV experiments (A) suggest a singular species when using the foldon as an obligate trimer. (B) The single run fit to SE data using a single species model estimates the MW of HFPRR as 53 kDa. The experimental data is displayed in the top panel with the plotted residuals in the bottom panel.

5.3 Chapter Discussion

The choice between using foldon and CCpII as the nucleation domain was a difficult one; CCpII has already been established as an N-terminal trimerization domain for bacterial collagen (Yoshizumi *et al.*, 2011). However, no NMR information was available. During NMR spectroscopy experiments, β-sheet rich regions have better chemical shift dispersion than α -helical rich structures, particularly if these structures are repetitive as in the case of CCpII. As a result, the signals from foldon will be spread across the ^{1}H-dimension of the HSQC, where the signals of a coiled-coil will centre around the middle. This central region is also the location of signals arising from polyproline and collagen-like structures. Therefore, the lack of chemical shift dispersion in the ^{1}H-dimension could obscure the signals collected from the PRR region of interest. It was therefore concluded that foldon would be the nucleation domain of choice.

Stabilisation of trimers and triple helices using foldon as the nucleation event is a well-reported technique (Frank *et al.*, 2001; S. Boudko *et al.*, 2002; Papanikolopoulou *et al.*, 2004; Bhardwaj *et al.*, 2008). Expression of the isolated foldon domain produces a stable trimeric structure (Frank *et al.*, 2001). However, N-terminal attachment of foldon in protein engineering did not yield stable trimeric proteins (Papanikolopoulou *et al.*, 2004; Bhardwaj *et al.*, 2008), whilst attachment at the C-terminus did (Frank *et al.*, 2001; S. Boudko *et al.*, 2002; Güthe *et al.*, 2004). Previous reports used foldon conjugated at the N-terminus to

stabilise or nucleate already established trimeric proteins (Frank *et al.*, 2003; Tsuji, Iwamoto and Shintani, 2014; Melendez *et al.*, 2018). This study is the first to generate an obligate trimer with foldon at the N-terminus for investigative, rather than stabilisation purposes. This success is likely due to the longer linker length between foldon and the PRR region employed in this study; previous work attempting to add foldon to the N-terminus had short linker lengths and therefore less flexibility in the conformations possible (Papanikolopoulou *et al.*, 2004; Bhardwaj *et al.*, 2008). In designing a linker of variable lengths (Figure 5.1), this study achieved improved flexibility in the possible protein conformations, allowing stable trimerization to occur. Furthermore, the nesting of restriction sites allows manipulation of this linker length in future studies, such as potentially introducing a cleavage site to test the overall stability of the structure formed once the nucleation event is removed.

Due to the polyproline-II nature of the PRR region, it was hypothesised that UBQLN2 may trimerize via this region to form a collagen-like triple helix. Numerous experiments were conducted to establish whether a shift from polyproline-II to collagen triple helix had occurred.

SV experimentation not only confirmed the presence of a singular species at a reasonable mass, but it also produced an intriguing result regarding the shape of the molecule. The high frictional ratio calculated during the analysis is suggestive of an elongated protein (Zhao *et al.*, 2013). The current model for the trimerization hypothesis involves a rod-like structure formed from the three PRR regions as a triple helix, with a bouquet of UBA domains presented at the end. This result would support the notion that an elongated rod-like structure was formed during trimerization of HFPRR. As both foldon and the UBA domain are globular, it can be assumed this high frictional ratio is largely due to the presence of the PRR region, potentially forming a collagen-like helical structure. This result is quite striking and certainly warrants further investigations into the structure of the PRR region as a trimer. Sedimentation Equilibrium was employed to determine the mass of HFPRR more accurately. The experimental mass derived from SE measurements is comparable to the *ProtParam* predicted size (53 kDa vs 57 kDa respectively), further indicating a trimeric structure has been successfully formed using the foldon domain at the N-terminus. Finally, visual interpretation of the CD data suggests the possible existence of a collagen-like triple helix, due to the blue

shift and increased size in peak magnitude at 193nm. However, no conclusion can be drawn with true certainty in this regard, and so more detailed structural investigations are required.

The data reported in this chapter demonstrates a novel achievement in engineering a stable, trimeric protein utilising foldon at the N-terminus. This protein can now be investigated in more depth, with the aims of elucidating the structure and inferring the function of the PRR region from this.

6 Investigating the structure of the PRR Region

6.1 Introduction

The PRR region of UBQLN2 has been determined to be largely disordered by numerous papers (Hjerpe *et al.*, 2016; Dao *et al.*, 2018). However, the PRR region does bear similarity to sequences known to oligomerize, such as elastin- and collagen-like proteins. The published studies of the PRR region investigate only forms initially purified as monomers. To ascertain whether the PRR region is capable of a triple helical secondary structure if expressed as a multimer, the obligate trimer HFPRR protein was engineered (see Chapter 5). This demonstrated a propensity to form a polyproline-II helix (see Figure 5.4), a common pre-cursor of collagen triple helices (Shoulders & Raines, 2009).

In order to further determine whether the formation of regular secondary structure in the PRR region could be detected, solution Nuclear Magnetic Resonance (NMR) spectroscopy was employed. Since the first detected chemical shift in 1950 (Proctor and Yu, 1950), NMR spectroscopy has proven an invaluable tool for the biophysical investigation of molecules. Allowing observation of the molecule at an atomic level in a non-destructive process with minimal sample preparation, NMR spectroscopy is a vital technique in determining molecular structure of proteins. NMR spectroscopy allows investigation of the composition of atomic groups within a molecule, molecular dynamics, and quantitative analysis such as proportions of compounds in a mixture. Whilst a valuable technique used in this study, explaining the complexities of NMR spectroscopy is not one of the objectives of this thesis. The particulars are outlined exceptionally well in various reviews, such as Bax & Clore, 2019. In general, when nuclei possessing a magnetic moment, such as hydrogen (^{1}H), nitrogen (^{15}N), and carbon (^{13}C) are placed into a strong magnetic field, the nucleus will precess. Irradiation of the sample with radio waves at the same frequency as this precession frequency changes the magnetic field experienced by the atom in the molecule. This can create coherence which results in a measurable resonance frequency. Individual chemical groups have characteristic resonance frequencies which aid in the determination of molecular structure, function and dynamics.

Using triple resonance NMR techniques, the signals originating from the foldon and UBA domains were identified and their backbone resonances assigned, whilst the PRR and linker regions presented larger challenges. Nonetheless, some information was garnered from these regions, and the results are discussed in this chapter.

6.2 Initial optimisation of conditions

The His_6 tag used for purification was cleaved from the HFPRR protein with thrombin and the resulting FPRR protein separated out by gel filtration. A series of ^{15}N-HSQCs were acquired over a range of temperatures (from 278 K to 310 K) to determine the optimum temperature giving good dispersion and resolution of the amide signals. 310 K was deemed to be most appropriate, due to the better intensity and reduced width of the peaks originating from the folded regions of the foldon and UBA domains (Figure 6.1).

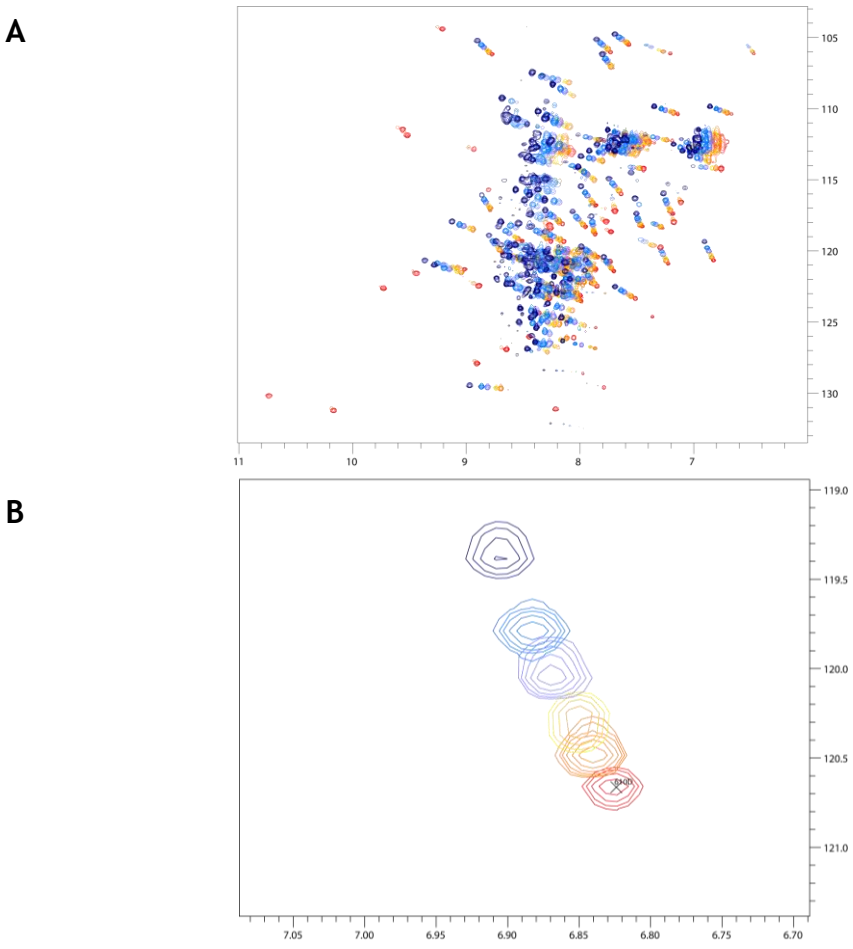


Figure 6.1 Increased temperature improves spectral signals from folded regions. A series of ^{15}N-HSQC at temperatures from 278 K (dark blue) to 310 K (red) at full spectral width (A) and zoomed into a residue of the UBA domain, Asp610 (B). Foldon peaks can be visualised in the left-hand side of (A) at only 310 K (red) whilst peaks of the UBA domain became sharper at higher temperatures (B, red). Spectra were collected in 20 mM KPi, 150 mM NaCl, pH 7.4.

6.3 NMR Assignment of Residues

6.3.1 Assignment Strategy

To elucidate the protein structure of FPRR, each amide resonance had to be mapped to a specific residue in the protein sequence. This was achieved using triple resonance (^{1}H, ^{15}N, and ^{13}C) experiments to measure the chemical shifts of the associated alpha, beta, and carbonyl carbons. The chemical shifts of these carbons fall within ranges that are indicative of particular residue types. The experimental strategy was designed using complementary experiments (found in methods Table 2.8) which revealed the chemical shifts of parent (i) and preceding (i-1) residues. For example, in the HNCACB experiment, the amide chemical shifts correspond to both the i and i-1 residues' Cα and Cβ carbons. In comparison, the CBCA(CO)NH experiment correlates the same chemical shifts for only the preceding (i-1) residue. Using the complementary datasets in tandem, amide links were inferred, and sequences of chemical shifts were generated. These sequences of chemical shifts were correlated with the known protein sequence. This aided in distinguishing between overlapping residues and joining fragmented stretches of sequences where resonances were not observed, such as in overlapping chemical shift regions or proline residues. The completed backbone assignment could then be used in subsequent investigations to infer structure and function of the FPRR protein.

6.3.2 Backbone Assignment

Visual inspection of the ^{15}N-HSQC spectrum indicated sufficient, well-dispersed cross-peaks to account for the residues in the expected globular foldon and UBA domains. A large number of poorly dispersed cross-peaks were also identified, which was expected for the PRR and linker-UBA regions. Peaks were picked in CCPN analysis v2.4.2 software and assigned to resonance objects that were grouped into spin systems. Sequential links were established via analysis of the carbon chemical shifts from the CBCA(CO)NH, HNCACB, HNCA, HNCO and HN(CA)CO (see Table 2.8 for parameters used). Provisional assignments to numerous short stretches of sequence were made based on the chemical shift information of the Cα, Cβ and carbonyl carbons. Prolines in the i-1 position were

identified manually through their distinctive chemical shifts observed in the CBCA(CO)NH and HNCACB spectra. Numerous short stretches were stitched together, until the assignments of the foldon and UBA domains were complete (Figure 6.2). The assigned chemical shifts can be found in Appendix I. Due to the proline rich (>25 %) and repetitive nature of the PRR region, full assignment of this region was not possible with the experiments conducted in this study. Chemical shifts were assigned to almost all residues in the foldon and UBA domains of the FPRR protein, with only residues 451 and 458 missing from foldon and 577, 621, and 622 missing from UBA domain. These residues could not be assigned due to poor resolution between peaks in repeated and flexible regions. Whilst the information gathered from the backbone assignment alone does not provide structural information itself, it provided insights into local chemical environments and provided further information regarding whether the various regions of FPRR are ordered.

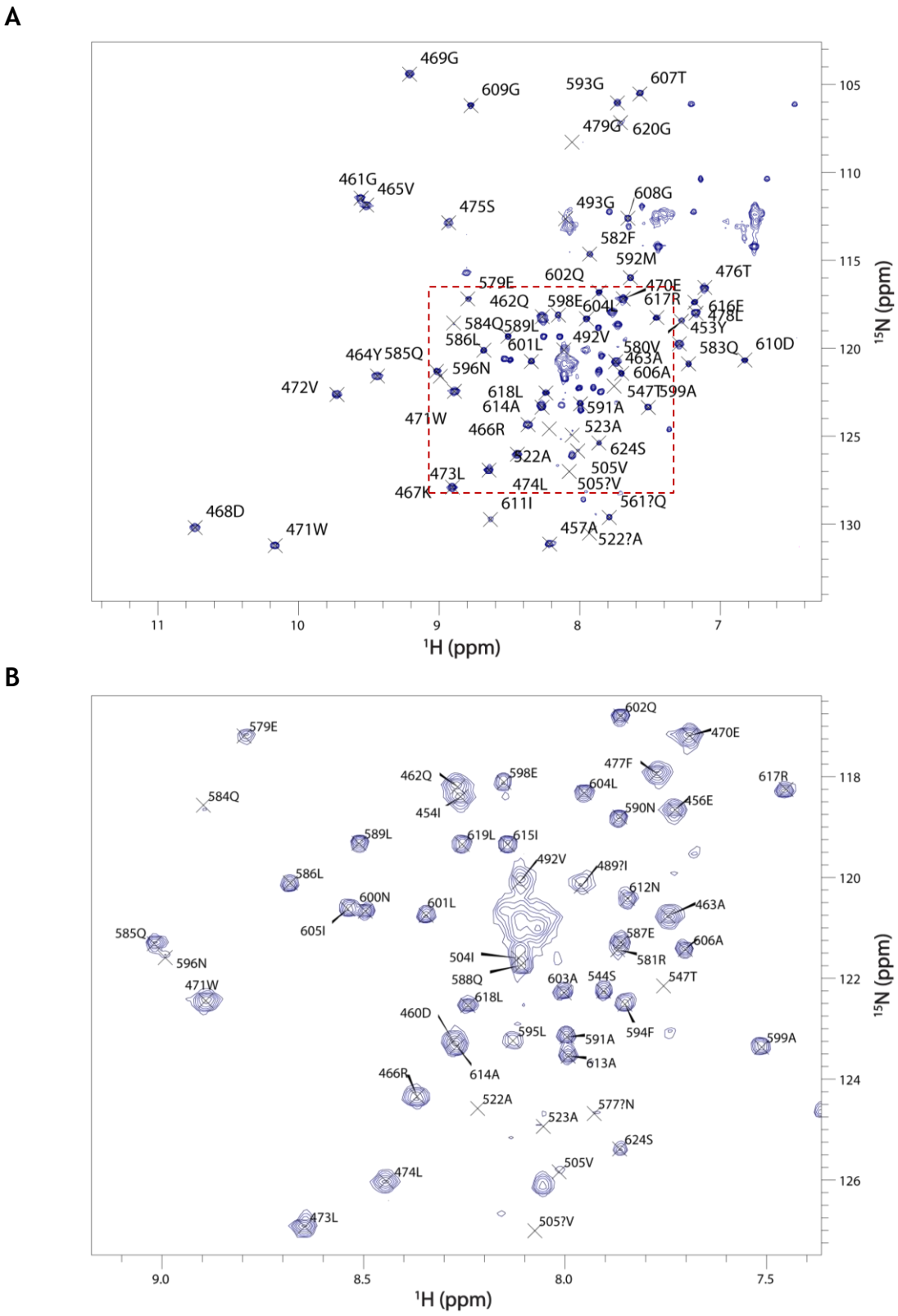


Figure 6.2 NMR backbone assignments of FPRR. ^{15}N-HSQC spectra of FPRR protein. (A) Full spectral width ^{15}N-HSQC with some central assignments not shown for clarity. Dashed red box identifies zoomed in area for (B), with all assignments labelled. Residues whose label includes a "?" denote a tentative assignment, indicating either an alternate conformation or that more information was required for conclusive assignment. Spectra were collected in 10 mM KPi, 75 mM NaCl, pH 7.4 at 310 K.

6.3.3 Verification of foldon folding and trimerization

A number of studies have engineered an obligate trimer using foldon as the N-terminal trimerization domain (Frank *et al.*, 2003; Papanikolopoulou *et al.*, 2004; Bhardwaj *et al.*, 2008; Tsuji, Iwamoto and Shintani, 2014; Melendez *et al.*, 2018) but some experienced limited success for various reasons (outlined in section 5.1.1).

Chapter 5 has already provided evidence for the successful production of a novel engineered trimeric protein, using foldon as a trimerization domain at the N-terminus of the PRR-UBA regions of UBQLN2. The structure of this obligate trimer was verified spectroscopically following backbone assignment of the FPRR protein. It has previously been established that several residues have distinct chemical shift differences between the monomeric and trimeric forms of foldon (Meier *et al.*, 2004). Comparisons between the ^{15}N-HSQC generated in this study (Figure 6.3B) and ^{15}N-HSQC produced for the monomeric and trimeric foldon (Figure 6.3A) were made. Visual analysis determined the chemical shifts relating to the foldon residues correlated with the trimeric, but not the monomeric, foldon. Therefore, foldon has here been successfully engineered as an N-terminal trimerization domain.

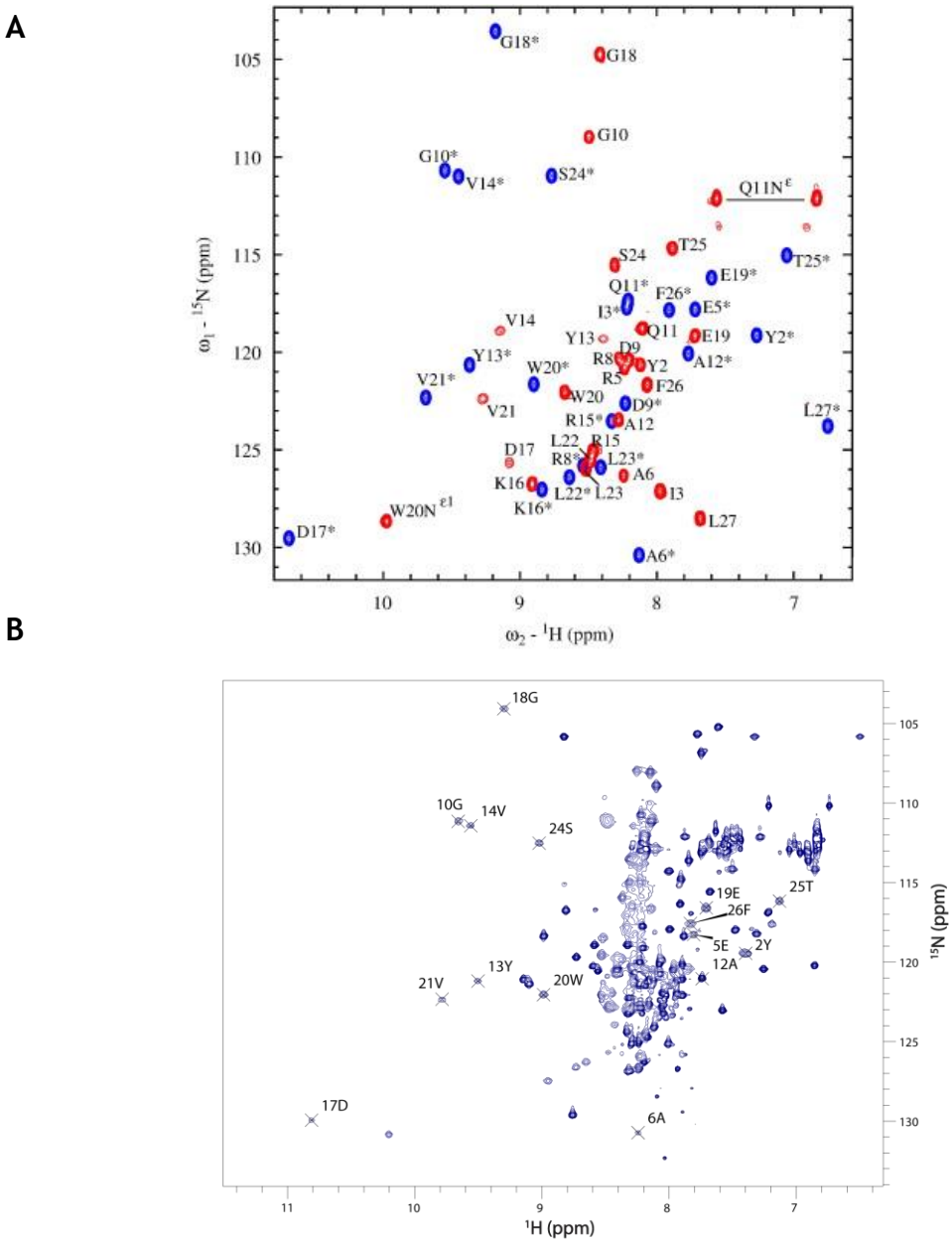


Figure 6.3 Foldon as an N-terminal trimerization domain is a folded, trimeric structure in FPRR. (A) ^{15}N-HSQC (generated in Meier et al., 2004) identifying the chemical shift changes between monomeric (red) and trimeric (blue) foldon. (B) ^{15}N-HSQC generated in this study with foldon annotations adjusted to match the comparison with (A). The peaks assigned match those of the trimeric (A, blue) but not monomeric (A, red). Spectra were measured under similar conditions of 10 mM KPi, pH7, 297 K (A) and 20 mM KPi, pH 7.4, 298 K (B).

6.3.4 Secondary structure predictions of FPRR

The majority of secondary structure calculation programmes predict the PRR region to be largely disordered. Evidence has been presented in this study which suggests a polyproline-II helix conformation is adopted (Figure 4.9). Dihedral Angles from Global Likelihood Measurements (DANGLE) estimates dihedral angles by matching chemical shift data and sequence query to an existing structural

database of known phi/psi angles and chemical shifts (Cheung *et al.*, 2010). Using a Bayesian method, the range of likely phi (ϕ) /psi (ψ) angles may be inferred to produce an estimate of the angle and a related tolerance level. DANGLE predictions were executed on the entire FPRR protein, searching the database with each sequence of 5 amino acids, using experimental chemical shifts in the predictions where assignments have been made. The projected psi (ψ) and phi (ϕ) angles have been visualised as a Ramachandran plot to allow comparison with other publications. The UBA domain (Figure 6.4C) contains mostly right-handed α-helical residues, identified by their location at approximately -60º in the phi and psi planes. This is concordant with published structures of the UBA domain of UBQLN2 (Dao *et al.*, 2018).

DANGLE was able to calculate predicted phi/psi angles for 37 out of the 51 residues composing the PRR region (Figure 6.4A). Dihedral angles for the remaining 14 residues could not be calculated as the prediction generated too many possible conformations. Glycine residues caused the most difficulty for the programme, comprising half of all unpredicted residue angles. Work by Bhattacharjee & Bansal, 2005 has produced a collection of psi/phi angles for various collagen-triple helices, composed of differing collagen fibres and oligopeptide repeats. Using this as an indication of the range of angles adopted by residues in a collagen triple helix, >50 % of the residues in the PRR region for which dihedral angles could be predicted have a structure similar to a collagen triple helix. (Figure 6.4B). Some residues had two possible predicted conformations, with the most likely conformations determined by the software and plotted. Of these, three residues (F507, P500, and G526) contained possible conformations lying in the collagen triple helix range. Furthermore, DANGLE uses experimental chemical shift data, where available, in its prediction of dihedral angles. There were only nine residues of the PRR region able to be assigned in this dataset, however six of these were located in the collagen triple helix region of the Ramachandran plot. This further strengthens the secondary structure prediction for these residues in the PRR region. In contrast, the residues of the linker region between the PRR region and UBA domain were largely dispersed throughout the Ramachandran plot, suggesting a flexible or disordered region. Together, this data suggests a structure similar to the collagen triple helix is adopted by over 50 % of the residues in the PRR region of the FPRR protein.

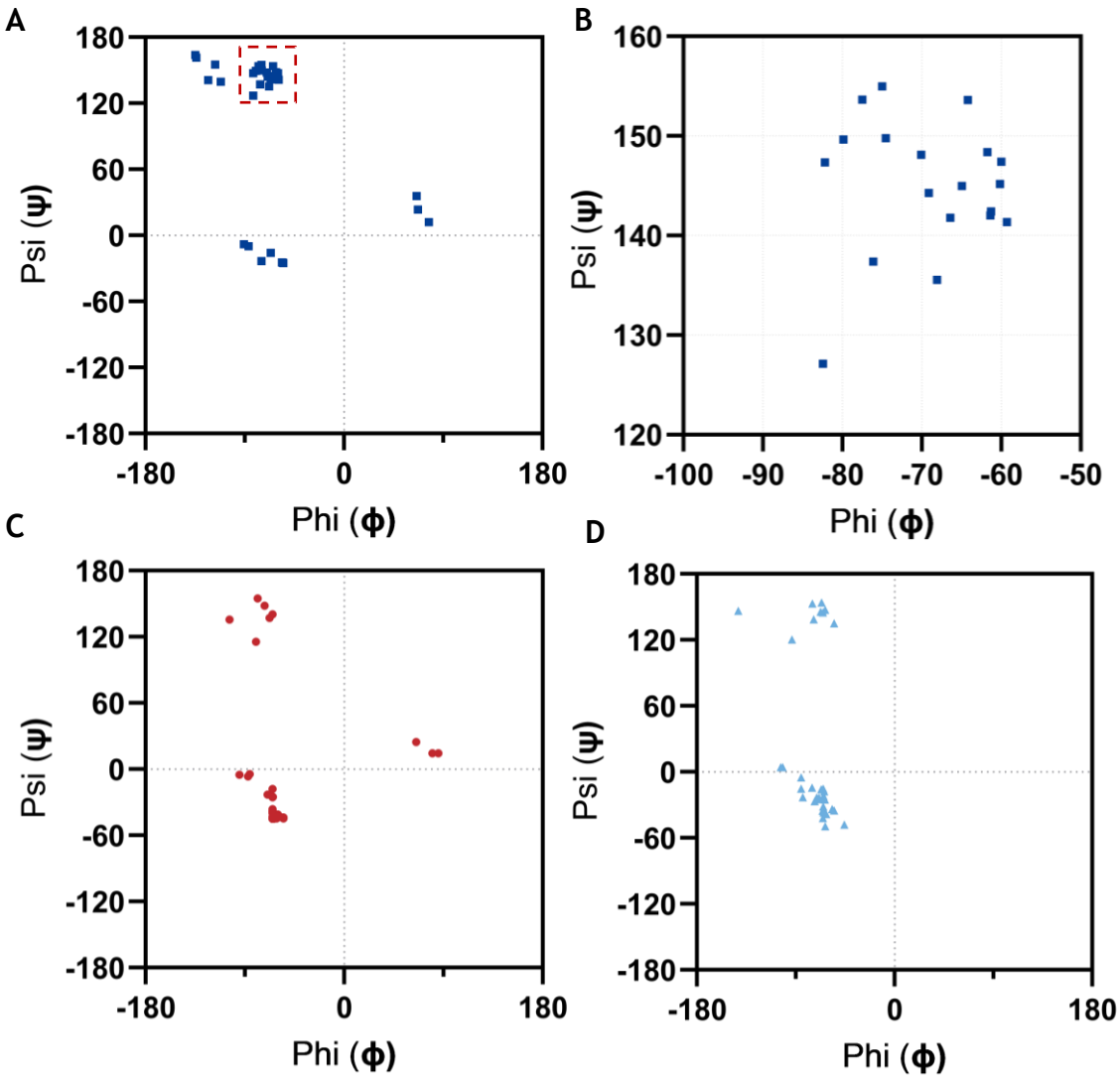


Figure 6.4 DANGLE secondary structure predictions. DANGLE dihedral angle predictions displayed as Ramachandran plots for the 3 key regions of the FPRR: PRR region (A), UBA domain (C), and linker (D). (B) is an enhanced view of the collagen triple helix region of the Ramachandran plot (Bhattacharjee and Bansal, 2005) for the PRR region, identified by the red dashed box.

6.3.5 Investigations into the UBA domains of FPRR

Complete backbone assignment of the PRR region was not achieved due to the proline rich and repetitive nature of the sequence. Therefore, it was not possible to discern a structure for this region. However, if a trimeric structure formed in this region, it was hypothesised this may bring together multiple UBA domains to confer ubiquitin chain specificity or affinity (Raasi and Pickart, 2003). This interaction may be visible as chemical shift perturbations (CSPs) of the residues in the UBA domain. In order to determine whether any structural interaction was occurring between the multiple UBA domains, chemical shift changes in the presence and absence of the trimeric PRR region were analysed. ^{15}N-HSQCs of the

His_6-Linker-UBA (monomeric) and FPRR (trimeric) proteins were acquired. The assignments established previously (section 6.3.2) were transferred to the corresponding HSQC peaks and any visible shift changes noted. Overall, the peaks could be superimposed extremely well (Figure 6.5A) with only minor shift changes noted. A small number of peaks exhibited larger shift changes, but as these were previously unassigned, they can likely be attributed to the flexible, disordered linker. To express the result numerically, chemical shift distances were analysed, and the resulting values plotted in Figure 6.5B. The largest shift change observed was 0.018 ppm, a value which is characterised as insignificant in published shift distance analyses (Williamson, 2013).

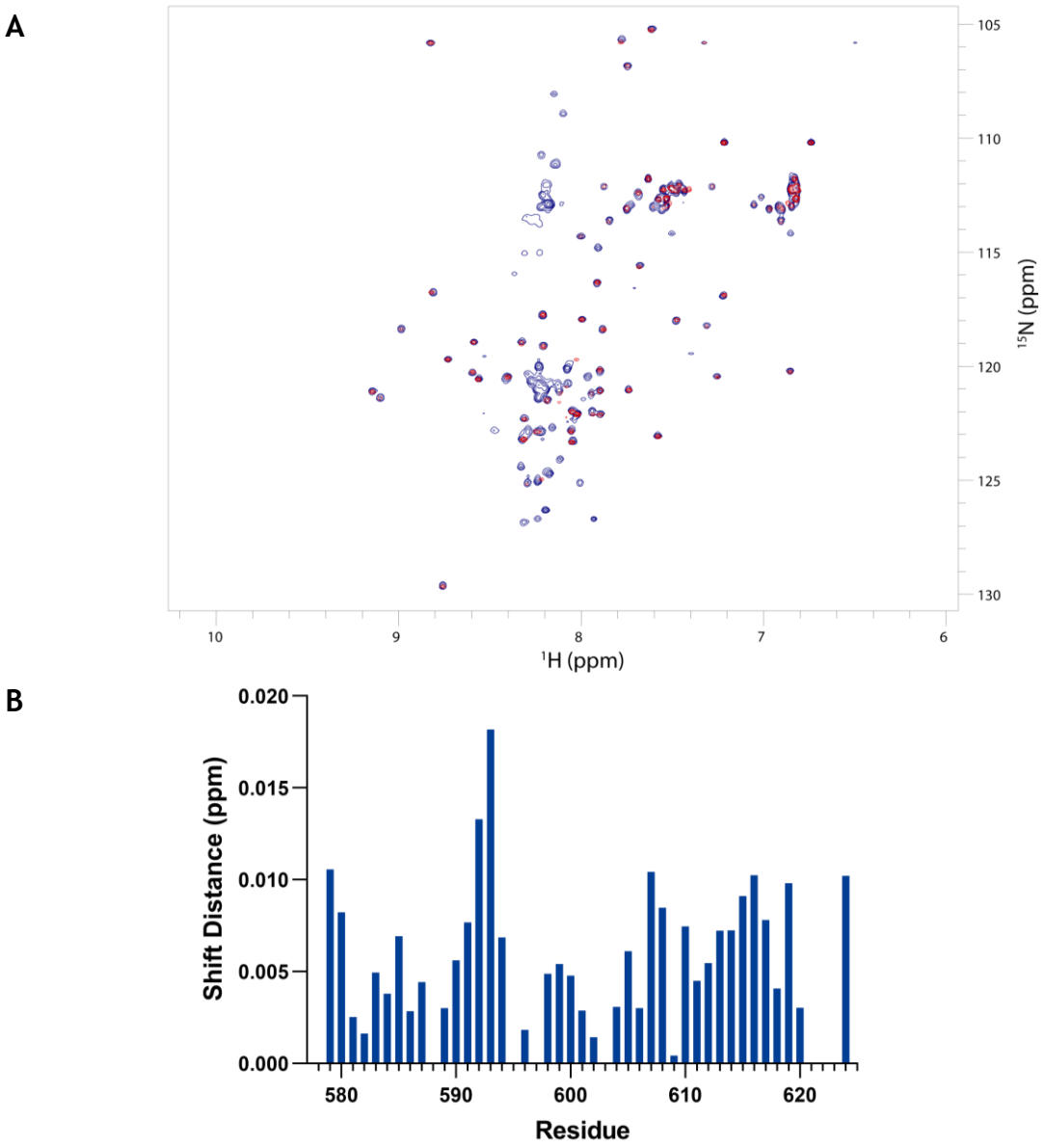


Figure 6.5 The UBA domain structure is unaltered in FPRR. (A) ^{15}N-HSQC of FPRR (blue) and His_6-TEV-Linker-UBA (red). Overlapping peaks demonstrate no change in the structure of the UBA domains, and therefore suggests no interaction between the UBA domains when expressed as a trimer. (B) A bar chart visualising the shift distance (ppm) for each assigned residue of the UBA domain. Shifts in the nitrogen dimension were normalised by a factor of 0.15. Spectra were acquired in 20 mM KPi, 150 mM NaCl, pH 7.4 at 298 K.

Together, this data suggests that the PRR region of FPRR may become structured when expressed as an obligate trimer. However, the lack of CSPs observed between monomeric and trimeric UBA domains demonstrates no interaction between the UBA domains when expressed as a trimer.

6.4 Chapter Discussion

Backbone assignments were successfully completed for the foldon and UBA domains of FPRR. This was especially crucial for the UBA domain, as whilst numerous studies have published a structure solved by NMR (Zhang, Raasi and Fushman, 2008; Dao *et al.*, 2018), none had deposited their assignments into the BMRB databank. Therefore, any downstream investigations into the UBA domain, such as binding analyses, would require assigned spectra for full understanding. Furthermore, this assignment allowed verification that foldon was correctly folded and existed as a trimer. Whilst previous studies have successfully created an obligate trimer using N-terminally conjugated foldon as a method of nucleation or stabilisation of already established trimeric proteins (Frank *et al.*, 2003; Tsuji, Iwamoto and Shintani, 2014; Melendez *et al.*, 2018), this study has successfully engineered an obligate trimer of a novel protein construct for purely investigative purposes.

During the backbone assignment, two issues were encountered. Assignments could not be completed for the PRR region and linker region, due to poor signal dispersion and/or missing amide signals. This likely reflects their repetitive nature or projected disorder respectively. The resonances from the disordered regions could likely be discerned in future studies by acquiring the triple resonance data at lower temperatures in order to slow solvent exchange, thus making the peaks from these regions more intense and distinct. Furthermore, proline residues can create two hurdles. First, proline residues are often located in proline rich regions as repetitive units (Theillet *et al.*, 2013), as is the case in the PRR region of UBQLN2. This repetition can cause overlap in the acquired spectra, thus distinguishing between the residues in these repeats can be very difficult. Secondly, proline lacks the amide proton when part of the polypeptide chain. Therefore, proline cannot be detected in $^{1}H_{N}$ detected experiments, even if the temperature has been optimised to reduce exchange broadening. In contrast, ^{13}C detected experiments can provide distinct residue characteristics when paired with ^{15}N nuclei (Bermel *et al.*, 2006; Hsu, Bertoncini and Dobson, 2009; Lopez *et al.*, 2016; Schiavina *et al.*, 2019) as the ^{13}C nuclear spins have a larger chemical shift dispersion (Dyson and Wright, 2001), despite possessing lower intrinsic sensitivity. A set of 3D carbonyl-carbon detected experiments have been

developed (Felli and Pierattelli, 2014), which reveal information on proline residues which would otherwise be unobtainable. Future NMR investigations into the PRR region of FPRR would benefit from these experiments, requiring fresh optimisation and potentially a higher field magnet or different probe.

Secondary structure predictions were calculated with DANGLE, a tool which uses a database of known structures and chemical shifts alongside sequence query to predict dihedral angles of each residue in the sequence (Cheung *et al.*, 2010). DANGLE is not the only tool capable of such predictions, but it is arguably the best for proteins where non-typical structures may be forming. TALOS (Cornilescu, Delaglio and Bax, 1999) and PREDITOR (Berjanskii, Neal and Wishart, 2006; Neal *et al.*, 2006) are two other dihedral angle prediction tools, using similar methods of tripeptide fragment searching, with homologous protein structure information also supplied in PREDITOR (Neal *et al.*, 2006). Using Ramachandran plots, both methods analyse the backbone conformations of the 10 closest matching fragments, deriving psi and phi from the mean values of hits within the major cluster. By ignoring the contribution from outliers, TALOS and PREDITOR bias the final predictions towards regions of Ramachandran space that are commonly populated (Cheung *et al.*, 2010). Therefore, inaccurate predictions are commonly made for non-canonical structures such as 3_{10} helices and conformations with positive values of phi (such as glycine). This inaccuracy is largely due to the methods inability to appropriately handle glycine or residues preceding prolines. In the case of the PRR region, this residue is often one and the same: a glycine precedes a proline in the $(Gly\text{-}Pro\text{-}Xaa)_n$ repeat. The DANGLE algorithm accounts for the different distributions these residues have in Ramachandran space by explicitly considering the population distributions expected for the different residue types, where TALOS and PREDITOR do not (Cheung *et al.*, 2010). Together, the DANGLE algorithm was the clear choice of secondary structure prediction tool when considering the complexities of the FPRR protein. Ramachandran plots are often mistakenly used to refine a structural prediction, as opposed to verifying it. By using DANGLE as opposed to other prediction algorithms, the bias generated in the final prediction is removed allowing more reliable secondary structure predictions to be generated. Whilst the data collected in this study falls within the same region as collagen triple-helix dihedral angles, it should be noted that this region overlaps severely with polyproline-II helices on the Ramachandran plot:

Polyproline-II helices populate the region of Ramachandran space at around -75º (ϕ) and 145º (ψ) (Adzhubei *et al.*, 1987; Siligardi and Drake, 1995b; Adzhubei, Sternberg and Makarov, 2013), whilst collagen triple helices are found at around -75˚ (ϕ) and 165˚ (ψ) (Berisio *et al.*, 2002; Bhattacharjee and Bansal, 2005). Therefore, caution should be exerted when predicting the secondary structure of the PRR region, in particular when distinguishing between polyproline-II and collagen triple helices. Using DANGLE predictions as a tool for revealing potential protein conformations supported the hypothesis of this study: the PRR region of UBQLN2 forms a distinct secondary structure and may be capable of forming a collagen-like triple helical structure. This prediction is strengthened by the chemical shift assignments for residues in the PRR region, where 2/3rd of the residues able to be assigned were located in the Ramachandran region of a collagen triple helix.

A second aspect of the hypothesis in this study is the conference of affinity and/or specificity of the UBA domain for ubiquitin binding. If the UBA domains are brought together in a trimeric structure, driven by the PRR region, then chemical shift changes were expected in the residues corresponding to the UBA domain (Raasi and Pickart, 2003). Comparison of the monomeric and trimeric protein revealed no interaction of the UBA domains when expressed as a trimer. This may be due to several possibilities. First, no true trimeric structure is formed downstream of foldon, meaning the UBA domains are loosely tethered to the trimeric foldon domains by flexible and disordered PRR and linker regions. Whilst there is no conclusive evidence relating to the structure of the PRR region in FPRR, there is evidence suggestive of a propensity to form a structure. Therefore, this conclusion cannot be drawn definitely. Secondly, an incorrect register of the PRR regions may be occurring if the linker between the foldon and PRR regions is not of the correct length and flexibility to drive assembly. Again, there is not enough evidence to say with any certainty whether this is the case. The construct encoding HFPRR has been designed in such a way that this linker can be altered, and further investigations using this variable linker may shed light on this option. The final possibility is that although a trimeric structure brings the UBA domains closer, they do not interact with one another to confer specificity or affinity. The UBA domain of Mud1 was determined to be monomeric, with no self-association (Trempe *et al.*, 2005), suggesting that interaction of UBA domains is not necessary

for ubiquitin binding. Furthermore, the UBA domain of mouse p62 forms a dimer. However, this dimeric structure is incompatible with ubiquitin binding. Thus, only monomeric UBA domains of p62 are capable of binding ubiquitin (Isogai *et al.*, 2011). Therefore, a lack of direct interaction between the UBA domains of UBQLN2 does not necessarily cause a lack of functionality in the domain, and further investigations are conducted later in this study.

To conclude, the work conducted in this chapter confirms a novel investigative method for generating an obligate trimer, using foldon at the N-, as opposed to C-, terminus of the protein of interest. This trimerization did not perturb the structure of the foldon or UBA domains and can be used in future investigations. The data collected suggests a structure may be formed in the PRR region, but final conclusions cannot be drawn regarding the structure of the PRR region with the experiments conducted in this study. Further experiments are required to assign proton resonances and side chain assignments and improved experimental design for revealing prolines is vital if the structure of the PRR region is to be determined. These experiments involve ^{13}C or ^{15}N direct detection based on CON experiments rather than ^{15}N-HSQC. Finally, there is no evidence of structural interaction between the UBA domains in the obligate trimer FPRR and further analysis to investigate the impact of an obligate trimer on the function of the UBA domain was conducted next in this study.

7 Investigating the binding affinity of the UBA domain in a monomeric and trimeric state

7.1 Introduction

Determining the binding affinity of protein domains is one of the key techniques employed to elucidate the biological function of ubiquitin-binding proteins. Establishing which domains preferentially bind which proteins, and the residues required to achieve this, is vital in deducing which regions are sufficient and necessary for protein function (Kim *et al.*, 2006). As such, information regarding the role of these proteins in disease can be extrapolated. Due to its similarity to UBQLN1 (Zhang, Raasi and Fushman, 2008), the UBA domain of UBQLN2 has had minimal investigation into the mechanism of its binding, although the residues required for binding have been identified (Dao *et al.*, 2018). Furthermore, the affinity for different chain lengths and linkage types has been determined (Nguyen, Puthenveetil and Vinogradova, 2017; Dao *et al.*, 2018) though not as thoroughly as UBQLN1 (Zhang, Raasi and Fushman, 2008). Notably, the binding affinity of the UBA domain of UBQLN2 has only been investigated under monomeric conditions, despite evidence from other ubiquitin-binding proteins that multiple UBA domains are required to regulate affinity and specificity (Chen *et al.*, 2001). This chapter aims to determine whether expression of the UBA domains as a trimer in the HFPRR protein impacts on binding affinity for ubiquitin and di-ubiquitin.

7.2 Optical analysis of binding interactions

Surface Plasmon Resonance (SPR) involves the immobilisation of the protein of interest (UBQLN2 peptides in this case) onto a Ni-NTA chip and flowing ligand across it. As binding occurs, the accumulation of mass on the surface of the chip changes the solvents refractive index. These changes can be monitored in real-time, allowing detailed information regarding binding kinetics to be extracted (Nelson *et al.*, 2001). Furthermore, SPR consumes much less sample but with a much higher throughput. It is therefore an excellent tool to determine the dissociation constant (K_d) and compliments NMR spectroscopy (Homola, 2003).

Constructs were designed and proteins expressed for SPR experiments. SPR classically uses a His_{12} tag to bind the protein of interest tightly to the chip. Three constructs were designed: His_{12}-PRR-UBA, His_{12}-UBA and His_{12}-UBA^{L619A}. The latter construct is a ubiquitin non-binding mutant to use as a control. The viability of these proteins was tested by monitoring how well each protein was immobilised on the Ni-NTA chip. All proteins produced the expected response units. However, the His_{12}-PRR-UBA protein was notably less stable on the chip.

A test series of 0-100 μM monoubiquitin (Ub) and K48-linked diubiquitin (K48-Ub_2) was flowed across each channel and the binding measured. Due to the instability on the chip, His_{12}-PRR-UBA did not show any binding whilst His_{12}-UBA produced K_d of 20.3±1.5 μM and 5.3 ±2.5 μM for Ub and K48-Ub_2 respectively. These values are comparable with those for the UBA domain of UBQLN1 (Zhang, Raasi and Fushman, 2008).

To overcome the problem of stability with the His_{12}-PRR-UBA protein, the obligate trimer HFPRR was trialled on the Ni-NTA chip. It was hoped no extra cloning would be required to insert a His_{12} tag as this protein already contained 3x His_6 tags in close proximity immediately N-terminal to the foldon-PRR-UBA protein. Therefore, this protein may already have the Ni-NTA avidity boost which is created using the His_{12} tag. However, HFPRR experienced issues binding to the Ni-NTA chip. Due to time constraints, further optimisation of this was not possible.

7.3 Spectroscopic analysis of binding interactions

Chemical shift perturbation (CSP) is a common experimental technique employed to investigate ligand binding to a protein. Using a ^{15}N-labelled protein and titrating in an unlabelled ligand, several insights can be gained such as location of the binding site, the affinity of the ligand, and potentially the structure of the complex.

Chemical exchange is a modulation in the chemical shift caused by microsecond to millisecond motions. This exchange can be caused by the protein's internal motions, such as a conformational change, or through interaction with another molecule, such as during complex formation. If chemical exchange between two states is slow in comparison to the NMR experiment, two separate peaks will be

observed in the spectrum (representing each state). If exchange is fast, a single peak at the average chemical shift of the two states will be observed in the spectrum. As such, only residues in fast exchange can be used to map the movement of peaks upon ligand binding. Therefore, CSP for K_d estimation is only possible for residues in fast exchange, which equates to a K_d weaker than 3 μM (Williamson, 2013).

Previous work by (Zhang, Raasi and Fushman, 2008) was used to calculate the predicted exchange regime of ligand binding. The UBA domain of the various UBQLNs is extremely well conserved, with only one residue difference (Ser→Asn at residue 590 in UBQLN2) between the UBA domains of UBQLN1 and UBQLN2. This residue difference was predicted not to impact on the binding of ubiquitin as the residue in question does not lie among the established ubiquitin-binding residues in UBQLN2 (Dao *et al.*, 2018). Monoubiquitin interacting with a single UBA domain of UBQLN1 has a K_d of 20 μM (Zhang, Raasi and Fushman, 2008). Assuming a similar K_d to the UBQLN1 UBA domain, the UBA domain of HFPRR is expected to be in fast exchange for large chemical shift change.

CSP analysis was therefore employed in this study to investigate the binding of ubiquitin to the HFPRR protein. 2D ^{15}N-HSQCs were acquired to monitor each stage of the titration. The chemical shift changes of selected residues at each titration point were measured and the movement of peaks mapped throughout the titration. Identifying the peaks with the most movement identifies the likely binding site for ubiquitin. By fitting to a hyperbolic titration curve (chemical shift vs concentration of ligand) the dissociation constant, K_d, of the complex can be determined.

It has previously been established that the UBA domain of UBQLN2 has a similar affinity to monoubiquitin as its counterpart domain in UBQLN1 (Dao *et al.*, 2018), though longer ubiquitin-chain lengths and types have not been investigated. As K48-linked ubiquitin chains are the primary proteasomal degradation signal (Thrower *et al.*, 2000), it is hypothesised UBQLN2 may have an increased affinity for K48-linked ubiquitin chains over monoubiquitin and other chain linkages. This set of experiments aimed to confirm whether the PRR-UBA portions of the protein conform to the current literature and assumptions, and to establish whether expression as an obligate trimer impacts on ubiquitin binding ability and affinity.

7.4 Experimental Design

7.4.1 Modelling saturation curves for ligand binding

To determine the range of concentrations required to experimentally sample the binding curve adequately, a simulation of the binding curve for the observed species was completed. The saturation of binding (Y) between the observed, non-varied, ^{15}N-labelled species (Macromolecule, M_{tot}) and the varied ligand (L_{tot}) was modelled using the relationship outlined in Green, 1965, implemented in Python as a Jupyter notebook hosted on a Google Colab page. The equation is outlined below, with the resulting simulated binding curve for ^{15}N-labelled Ub as the macromolecule and HFPRR as the ligand displayed as an example in Figure 7.1A.

$$Y = \frac{([M_{tot}] + [L_{tot}] + K_d) \pm \sqrt{([M_{tot}] + [L_{tot}] + K_d)^2 - 4[M_{tot}][L_{tot}]}}{2[M_{tot}]}$$

Successive addition of ligand material from a concentrated stock solution into the observed species will increase the total volume of the sample with each titration point. As a consequence, the concentration of the initial observed species will be reduced over the course of the experiment. To account for this, the volume (v_t) can be expressed in terms of ligand concentration, where L_s is the stock concentration of the ligand and v_0 is the initial volume.

$$v_t = v_0 \times \frac{[L_s]}{[L_s] - [L_{tot}]}$$

The observed macromolecule can therefore be expressed as below, where M_0 is the initial macromolecule concentration.

$$[M_{tot}] = [M_0] \times {}^{v_0}/_{vt} = [M_0]\left(1 - {}^{[L_{tot}]}/_{[L_s]}\right)$$

Finally, the response curve can be expressed as follows, with the simulated binding curve for this model displayed in Figure 7.1B.

$$Y = \frac{([M_0]\left(1 - {}^{[L_{tot}]}/_{[L_s]}\right) + [L_{tot}] + K_d) \pm \sqrt{\left([M_0](1 - {}^{[L_{tot}]}/_{[L_s]} + [L_{tot}] + K_d\right)^2 - 4[M_0]\left(1 - {}^{[L_{tot}]}/_{[L_s]}\right)[L_{tot}]}}{2[M_0]\left(1 - {}^{[L_{tot}]}/_{[L_s]}\right)}$$

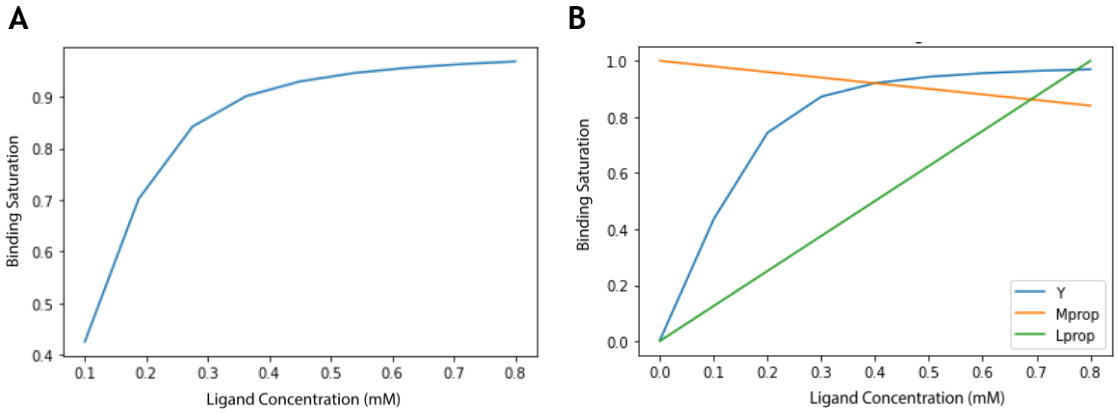


Figure 7.1. Simulated binding saturation. The saturation of the binding site with increasing ligand concentration was modelled to aid in experimental design. (A) An example of the saturation curve (blue) where the macromolecule concentration is maintained throughout the titration. (B) Simulated binding curve when macromolecule concentration (Mprop) is not maintained as ligand (Lprop) is added. 97 % of total protein is bound by 0.8 mM ligand concentration.

Simulation of the binding curve identified the range of ligand concentrations required to experimentally sample the binding curve as 0 mM -0.8 mM, when the macromolecule concentration is 0.2 mM. At 0.8 mM, 97 % of macromolecule will be bound by ligand and the experiment complete.

7.4.2 Experimental set up

In order to ascertain the ability of the HFPRR protein to bind ubiquitin, a series of titrations were set up. The first using ^{15}N-labelled HFPRR with increasing amounts of unlabelled mono-ubiquitin titrated in (^{15}N-HFPRR), and the second using ^{15}N-labelled mono-ubiquitin and titrating in unlabelled HFPRR (^{15}N-Ub). This would provide similar information relating to the binding of ubiquitin but from two differing perspectives: that of the HFPRR and from that of mono-ubiquitin. Finally, differential binding affinities to K48-linked chains were investigated using distally ^{15}N-labelled diubiquitin linked via K48 (^{15}N-K48-Ub$_2$).

During experimental set-up with ^{15}N-HFPRR, measurements at 9 points between 0 and 4 molar ratio were planned. However, only 5 points were measured between

0 and 2 molar ratio, as no further change was apparent after this point, so the data were judged sufficient. It became evident during analysis that this decision may have been premature, due to the large errors in K_d (Figure 7.3B(i)) and continued increase in the experimental saturation curves. Furthermore, the fitting equation used for calculating K_d in the analysis software cannot account for the slight dilution of ^{15}N-HFPRR protein when adding sequential volumes of ligand. This dilution and inability to process it mathematically may also contribute to the larger K_d error values.

To overcome this problem, subsequent experiments were conducted by exchanging samples. Two samples were prepared, both containing equal concentrations of the observed species but with two different concentrations of ligand: one with the lowest concentration and the other with the highest concentration of ligand to be explored. Exchange of volumes between these two samples generated the intermediate experimental titration points. Using this method, the problems with macromolecule dilution are avoided, as the concentration of the macromolecule being observed is kept constant. A range of molar ratios could then be explored, outlined in Table 7.1.

Table 7.1 Experimental design of titrations. Each titration experiment is outlined here with the molar ratios measured and the method by which the titration was achieved.

Experiment	^{15}N-labelled protein	Unlabelled ligand	Molar ratios measured	Method of titration
1	HFPRR	Monoubiquitin	0, 0.5, 1, 1.5, 2	Sequential addition of ligand
2	Monoubiquitin	HFPRR	0, 0.5, 1, 2, 3, 4, 4.5, 5	Exchange of sample
3	K48 linked-diubiquitin	HFPRR	0, 0.4, 0.8, 1.9, 2.3, 2.7	Exchange of sample

7.4.3 Quality Assurance

A series of 2D-^{15}N-HSQCs were recorded whilst titrating the ^{15}N-labelled macromolecule with the unlabelled ligand. Spectra were analysed using the CcpNMR analysis v2.4.2 software (Vranken *et al.*, 2005). The deposited assignment for monoubiquitin (BMRB 16228) was imported in, and the residue assignments transferred to the ^{15}N-Ub and ^{15}N-K48-Ub_2 spectra. Where HFPRR was the observed ^{15}N-labelled macromolecule, the UBA assignments completed in Chapter 6 were used.

Peak position changes in the spectra over the course of the titration indicates a change in the electronic environment of the corresponding nucleus as a result of the binding interaction. Therefore, residues with the strongest CSPs likely indicate regions of binding (Williamson, 2013), and these were initially identified and verified within the spectra visually (Figure 7.2A). This revealed numerous residues that could not be included in the analysis. This was due to incomplete data due to peaks becoming overlapped, for example where two residues' trajectories intersect (Figure 7.2C), or due to excessively broadened signals caused by intermediate exchange occurring (Figure 7.2B), resulting in the loss of peak signal at certain concentrations. Including residues undergoing intermediate exchange would also invalidate the fast exchange assumption in the fitting (see section 7.3) and as such these were excluded from analysis.

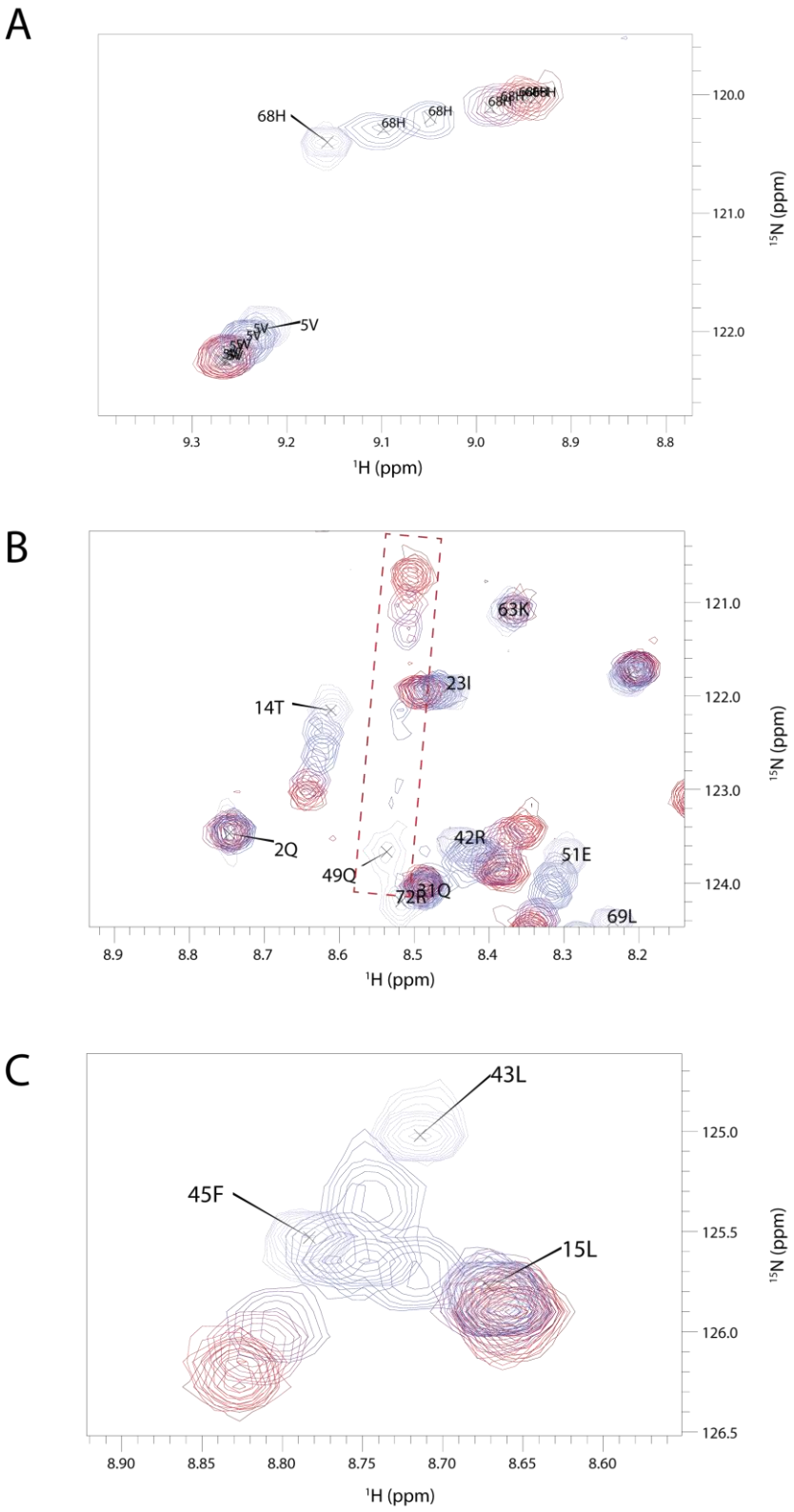


Figure 7.2 Visual considerations when mapping CSPs. Examples of cross-peak trajectories in the ^{15}N-HSQC of experiment 2. Cross peaks move from no ligand (HFPRR) (light blue) through to 0.5 mM ligand concentration (dark red), with ubiquitin concentration kept at a constant 0.1 mM. (A) Residue 68 has a considerable shift change with increasing ligand concentration whilst residue 5's change is less. (B) Residue 14 is another example of large chemical shift change whilst residue 49's peaks are broadened to the limit of detection at some intermediate concentrations (red box). (C) The cross-peak trajectories of residues 45F, 43L, and 15L intersect.

The binding affinity between HFPRR and ubiquitin was quantified by fitting the observed CSPs as a function of the protein and ligand concentrations to a single binding mode. As the data from all residues are reporting on the same event, the same binding model can be applied but with different overall magnitude of observed CSPs, which correlates to the sensitivity achieved experimentally.

Residues were ordered by chemical shift distance and the K_d error investigated. Using the ^{15}N-Ub experiment as an example, an explanation of the process of quality assurance is outlined as follows. Comparison of shift distance and K_d error revealed at a shift distance of <0.05 ppm, the errors associated with each calculate K_d value are much larger. This was also observed when inspecting the experimental vs mathematically fitted saturation graphs of each residue (Figure 7.3A), comparing the experimental shift plotted against the predicted shift calculated using the equation in section 2.2.4.10. Establishment of the cut off for significant K_d values was finally achieved by plotting the K_d of each residue as a bar chart, along with its corresponding error bars, in decreasing shift distance along the X axis (Figure 7.3B(i)). This highlighted a clear step in the reliability of the data between residues 54 and 5. This was the determined cut-off point for K_d values to be used in subsequent calculations for this experiment. These quality assurance techniques were employed in each titration experiment, whose results are each displayed in Figure 7.3B. It was not possible to determine a reasonable cut off range for reliable K_d values in the ^{15}N-HFPRR experiment, and as such the K_d values calculated were not used in subsequent analysis.

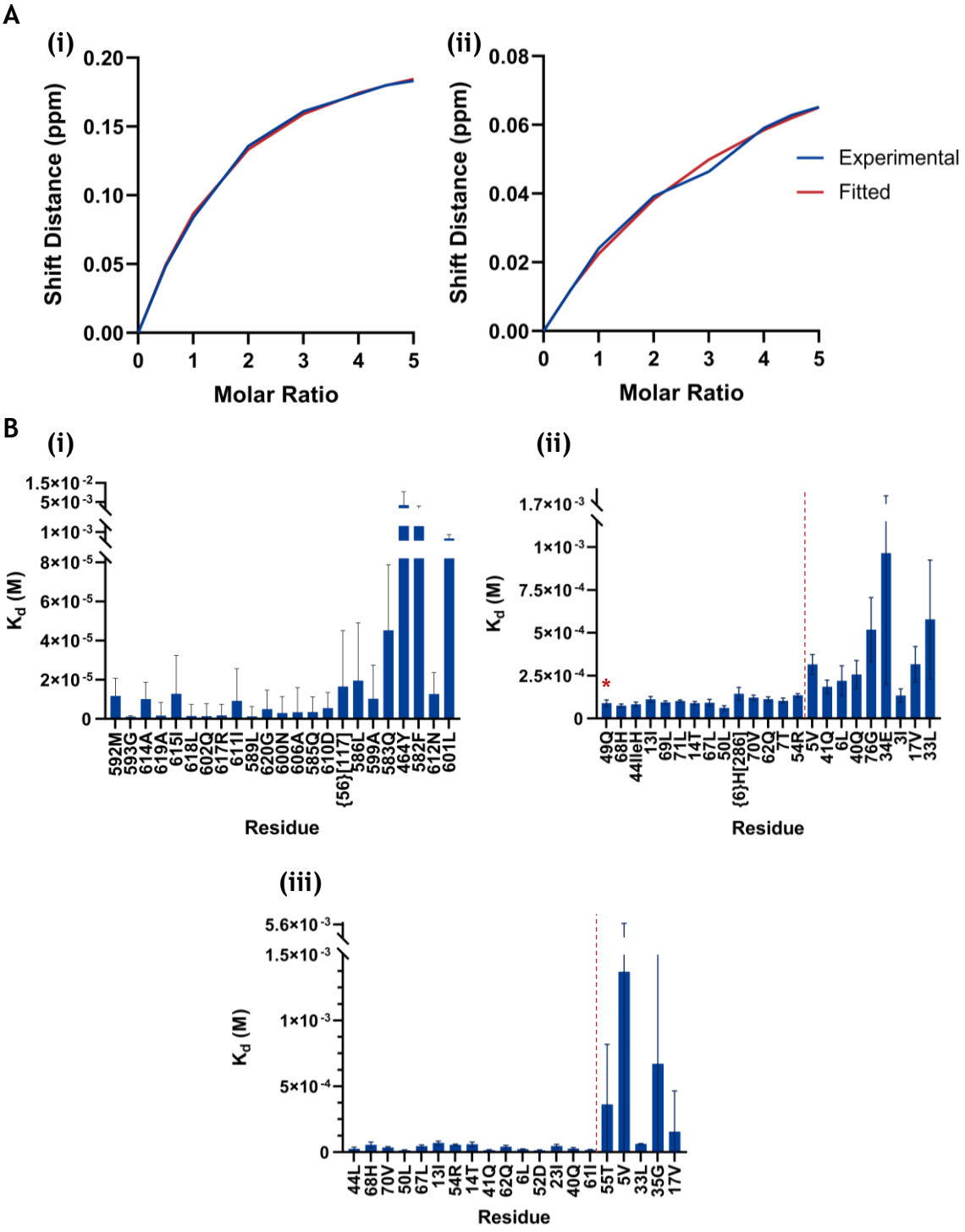


Figure 7.3 Numerically establishing the range of significant K_d values. (A) Examples of saturation curves for two residues, 71(i) and 5 (ii), where the fit is good (i) and less good (ii). (B) Bar charts illustrating the final cut off values for significant K_d values (left of the red dashed line) for experiment 2 (ii) and experiment 3 (iii). It was not possible to determine a reasonable cut-off range for significant K_d values for experiment 1 (i). Residue numbers were assigned from the BMRB deposited assignment for monoubiquitin (BMRB 16228) in experiments 2 (ii) and 3 (iii) or from UBA assignments completed in Chapter 3 (iii). Residues are ordered by increasing chemical shift distance along the X axis, where CSPs=$[(\Delta H)+(0.15*\Delta N)]^{1/2}$. It was not possible to assign a residue to peaks {6}H[286] and {56}[117] but their shift distance and K_d value and error were significant enough to include as it was determined that it was not necessary to assign a residue to this peak for the purpose of this chapter. The red asterix above residue 49Q indicates a discounted residue from this cut-off due to undergoing intermediate exchange.

7.5 Determination of the dissociation constant through spectroscopic analysis

7.5.1 Apparent Kd

Quantitative analysis of each of the titration experiments yielded an equilibrium dissociation constant (K_d) for each residue. The apparent K_d of each residue was plotted against its shift distance in Figure 7.4. The regions and residues of ubiquitin involved in binding to the UBA domain of UBQLN1 have been previously determined as the ß1 and ß2 loops and ß2 strand (containing residues 7T, 8L, I13, T14), ß3 and ß4 (residues 43L-51E), and ß5 (residues 67L-73L)(Beal *et al.*, 1996, 1998; Zhang, Raasi and Fushman, 2008). The residues with the strongest CSPs were investigated manually and found to correspond to these established ubiquitin binding residues.

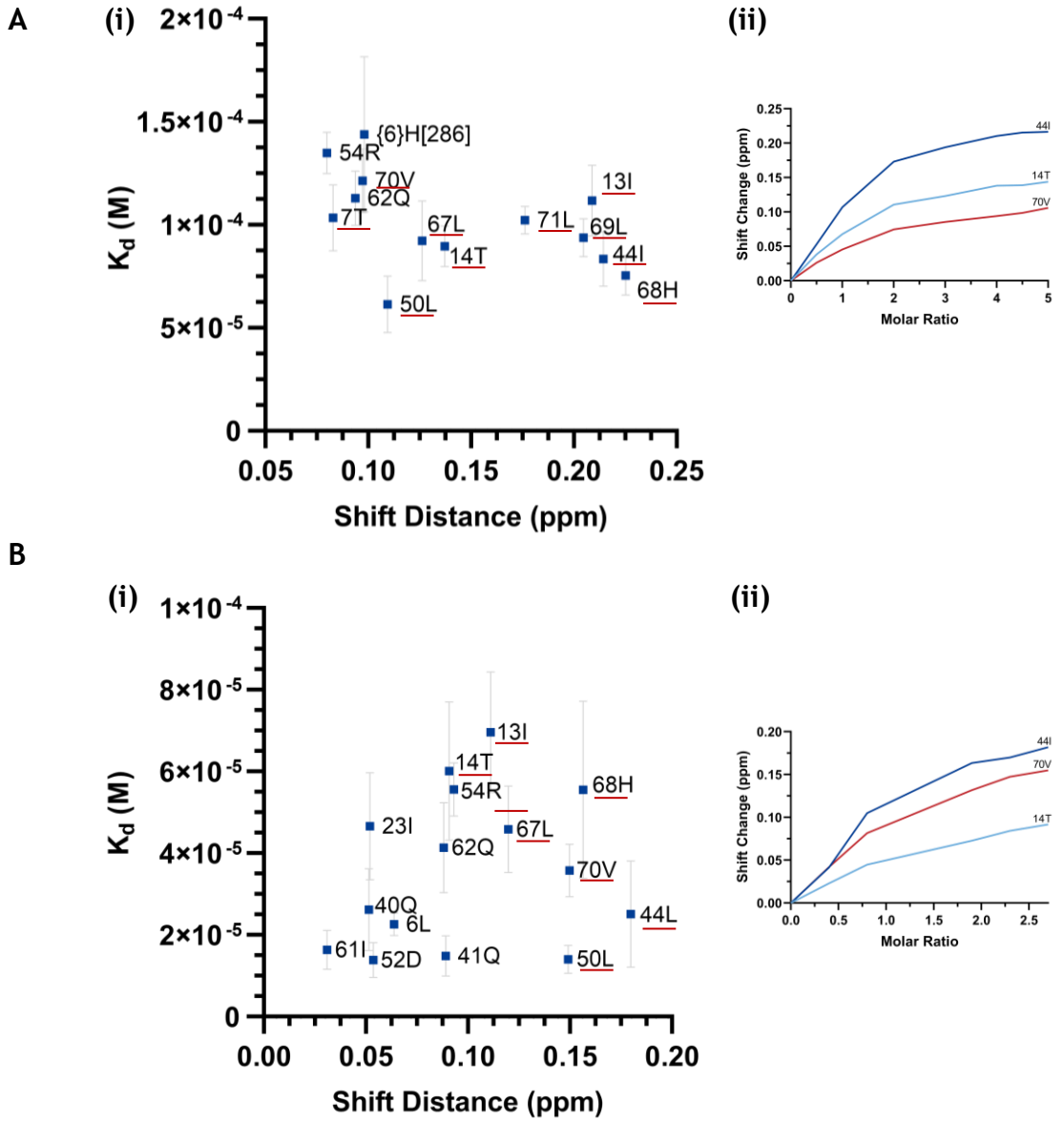


Figure 7.4 The apparent K_d values of ubiquitin residues involved during binding to HFPRR. Scatter plots (i) of the residues of ubiquitin likely to be involved when binding to HFPRR UBA domains, with measured shift change plotted against apparent K_d. Residues previously established to be involved in the binding of ubiquitin to UBA domain of UBQLN1 are underlined in red. (A) represents ^{15}N-Ub whilst (B) represents ^{15}N-K48-Ub$_2$. Example titration curves for a selection of perturbed residues (ii), where CSPs are plotted as a function of molar ratio of the titrant and ^{15}N-labelled protein under observation. Assignments for residues were extracted from the deposited spectra of monoubiquitin (BMRB 16228).

Previous studies have reported a K_d value for the UBA domain binding Ub as approximately 20 μM, determined by quantitative analysis of NMR titration experiments and SPR analysis (Zhang, Raasi and Fushman, 2008). Averaging the K_d values calculated in this study and their error values (Table 7.2) revealed a binding affinity for monoubiquitin approximately 5-fold higher than already published. An approximate 3-fold increase in affinity for K48-linked diubiquitin in comparison to monoubiquitin was determined in this study. However, this dissociation constant

was still 3-fold larger than the equivalent published data. Therefore, it is possible the existence of the PRR region and UBA domains as an obligate trimer perturbed rather than improved the binding of ubiquitin.

Table 7.2 Average K_d for different ubiquitin chains. The individual K_d values of residues with strong CSP upon binding were averaged, and their respective error values squared, averaged and then square-rooted. K_d for UBQLN1 was obtained from (Zhang, Raasi and Fushman, 2008).

Ubiquitin chain length and linkage	This Study Average K_d (μM) (UBQLN2)	Zhang et al 2008 Average K_d (μM) (UBQLN1)
Ub	113±17	20±5
K48-Ub_2 (distal Ub)	36±11	12±10

7.6 Chapter Discussion

Expression of the PRR-UBA region as an obligate trimer was originally hypothesised to increase binding affinity to ubiquitin, in a similar mechanism to multimeric expression of the UBA domains of UBQLN1 (Hjerpe *et al.*, 2009). However, there is evidence in other UBAs that multimeric expression in fact prevents ubiquitin binding (Isogai *et al.*, 2011). This chapter aimed to elucidate whether expression of the UBA domains as a trimer would reveal any regulatory impacts on ubiquitin binding similar to those already published.

Quantitative analysis of the CSPs in experiments ^{15}N-Ub and ^{15}N-K48-Ub_2 yielded a K_d value of 113± 17 μM and 36 ±11 μM respectively. Whilst there is a 3-fold difference between the binding of monoubiquitin and diubiquitin, both the values determined in this study were larger than those previously published (Zhang, Raasi and Fushman, 2008; Dao *et al.*, 2018). This difference may be due to incomplete saturation of the binding site at the final titration point of the experiments, as modelling the saturation curve predicted 97 % saturation (Figure 7.1) which does not appear to have been achieved for all residues (Figure 7.4). The balance between available ubiquitin for experimentation and the number of titration points measured would need improving in future studies. It is worth noting that the published comparisons used in this study were conducted on the UBA domain

of UBQLN1. Without an appropriate control for monomeric UBQLN2 UBA domain, no conclusive arguments can be made. However, SPR data collected in this study confirmed monomeric UBQLN2 UBA domains bind ubiquitin with a comparable affinity to UBQLN1 (20.3±1.5 μM). Therefore, the assumptions made in this study and published data previously, that one residue difference between UBQLN1 and UBQLN2 in sequence does not affect ubiquitin binding, are likely true. Therefore, the changes in ubiquitin binding observed in this study are likely due to the presence of the PRR region and/or multimerization.

Another possible cause for the difference in K_d may lie in how the parameters of significant CSPs were determined. In previous publications (Zhang, Raasi and Fushman, 2008), only a selection of K_d values for the established interacting residues were included in the quantitative analysis and final K_d average, despite other residues' established role in ubiquitin binding (Beal *et al.*, 1996, 1998; Hicke, Schubert and Hill, 2005; Hurley, Lee and Prag, 2006). The exclusion of these residues was not justified by the authors and no clear pattern is apparent when comparing their CSPs to those included. In this study, quality assurance and validation of significant CSPs was achieved numerically, with the aim of removing bias. Furthermore, it was not possible to identify a number of binding residues in the spectra collected in this study. The published analysis determines CSPs from analysis of TROSY experiments, whilst this study used 2D-^{15}N-HSQCs. TROSY experiments have better resolution, allowing identification of all residues involved in binding. The omission of these residues in this study, albeit unavoidable, may have contributed to the difference in calculated K_d values. Furthermore, TROSY experiments may produce a more precise peak position which could improve K_d calculations, especially in noisier data.

Experimental set-up may be another contributor to the difference between the K_d values calculated in this study and those previously established. First, in the published data (Zhang, Raasi and Fushman, 2008), increasing volumes of ligand were sequentially added to the macromolecule. However, it is unclear whether the dilution caused as a consequence of this sequential ligand addition was accounted for in the fitting model used. Secondly, the titration experiments were performed at high macromolecule concentrations which equate to 35x the measured K_d. In comparison, the macromolecule concentrations used in this study never exceeded 2x K_d. An optimum concentration of macromolecule has been

calculated as 0.5x K_d (Granot, 1983), though concentrations up to 10x K_d are acceptable (Markin and Spyracopoulos, 2012). The high macromolecule concentration used in the published data can result in a loss of accuracy in the K_d calculation (Granot, 1983; Markin and Spyracopoulos, 2012), which may cast doubt over the validity of using these values as a comparison. Future investigation would benefit from a control, where foldon has been mutated to prevent trimerization (Habazettl, Reiner and Kiefhaber, 2009). This would allow complete comparison of constructs without the need to refer to previously published data.

A final explanation of the difference between K_d values in this study and the published data could be a genuine difference in ubiquitin binding. In the case of p62, dimeric UBA interactions are not conducive to ubiquitin binding, and a shift from dimeric to monomeric UBA triggered by the addition of ubiquitin has been reported (Isogai *et al.*, 2011). A similar phenomenon occurs when investigating the oligomeric profile of UBQLN2 during liquid-liquid phase separation (LLPS). Addition of mono-, di- and tetraubiquitin to phase-separated UBQLN2 eliminated the oligomerization (Dao *et al.*, 2018), revealing another mutually exclusive state between ubiquitin binding and oligomerization. A mechanism of regulation similar to p62 has been ruled out for the UBA domains of UBQLN2, as no structural interactions between the UBA domains were detected in the experiments conducted in this study. However, the currently published data demonstrates oligomerization may disrupt ubiquitin binding (Isogai *et al.*, 2011) and conversely, ubiquitin binding can disrupt oligomerization of UBQLN2 (Dao *et al.*, 2018). Therefore, expressing the UBA domains as a constitutive trimer may interfere with this relationship. The control described above, where foldon is mutated to prevent trimerization, would again be useful in elucidating the relationship between oligomerization of UBQLN2 and ubiquitin binding.

The change in ubiquitin binding reflected in the data produced in this chapter supports the hypothesis that chain specificity may be conferred by multimerization of the UBA domains. To investigate this further, a variety of ubiquitin chain lengths and linkage types should be investigated. In particular, tetraubiquitin titrations may shed light on potential specificity of the UBA domains; diubiquitin may not cover the necessary conformational space to evoke a change in measured specificity. As a complete assignment could not be completed for the PRR region, it is not possible to say whether the change in

ubiquitin binding is due to the presence of the PRR region or due to oligomerization. Repeating these experiments with the Linker-UBA protein (described in Chapter 4) and monomeric HFPRR (as described above) would give a better indication to the cause of this ubiquitin binding perturbation. Future investigations would also benefit from ^{15}N- relaxation studies, comparing Linker-UBA to FPRR, to give an indication of whether the FPRR protein behaves significantly differently. For example, if the rod-like extension is formed by the PRR region when expressed as a trimer, this would be represented as a slower-tumbling species in NMR data. Should data from these experiments prove interesting, research could be extended to relaxation dispersion and Chemical Exchange Saturation Transfer (CEST) experimentation. Data from these techniques could inform on other states not directly observed. For example, if a conformation of interest is only adopted 5 % of the time, this would not be detected without relaxation data, which is capable of identifying signals from other states (Forsén and Huffman, 2004; Rangadurai, Shi and Al-Hashimi, 2020).

To conclude, this study determined a 5-fold and 3-fold decrease in mono- and diubiquitin binding respectively, when compared to previously published data. A slight increase in binding affinity towards diubiquitin was observed, though without further investigation with different chain lengths and linkages, this data is not striking enough to draw any conclusions. Nonetheless, the UBA domain of UBQLN2 is still capable of functionally binding ubiquitin which allows further avenues of investigation using the novel engineered protein HFPRR.

8 Discussion

8.1 Developing an iPSC-derived MN model for a UBQLN2 mutation

The original aim of this thesis was to generate a novel iPSC-derived motor neuron (MN) model of a UBQLN2 mutation to elucidate a function for UBQLN2 in the MNs and infer its role in the development of ALS. This was to be completed using CRISPR/Cas9 gene editing to produce a stably expressing UBQLN2^{P506T} or UBQLN2^{P497H} mutant. Multiple steps were taken to begin this process, such as validating the quality of the iPSC lines and the efficiency of differentiation into MNs, a process which had not previously been completed with this protocol. A viable line for engineering was chosen from quality control experiments, and endeavours made to begin the CRISPR/Cas9 process. All constructs had been designed and tested, with transfection and screening of the iPSCs beginning just before the Covid-19 pandemic.

Creating a physiologically relevant model of disease is the bold but controversial aim of many labs. Each field could argue its own case, justifying why their model is the best. For example, an animal model lacks the same genome as humans, but it has the benefit of a whole organism system, where cellular and tissue interactions can be monitored. In contrast, iPSC-derived cell models hold the advantage of containing the human genome, sometimes with patient-specific mutations which can be rescued with CRISPR/Cas9 to create the ultimate control. However, these too have significant drawbacks.

It is well established that neuronal connectivity drives development and maintenance of neurons (Kirkby *et al.*, 2013), resulting in the over simplified phrase “use it or lose it”. Cultivation of neurons *in vitro* lack this interplay and as a consequence will not mature to the comparable level as an *in vivo* human. A prime example of this is when comparing the electrophysiological profile of *in vitro* and *in vivo* MNs. Even after extended time in culture, human iPSC-derived MNs exhibit resting membrane potentials and action potential dynamics similar to embryonic stage rodent MNs (Devlin *et al.*, 2015). Functionally, this translates into artificially hyperexcitable neurons with limited ability to spontaneously generate

action potentials. In addition to this, iPSC-derived MNs demonstrate minimal ability to form synapses, a critical component of the neuron's function. Furthermore, spontaneous action potential activity is suggested to drive MN maturation and functionality in the spinal cord (Takazawa *et al.*, 2012). Although not fully understood, co-culture of iPSC-derived MNs with primary cortical mouse astrocytes causes a significant increase in the maturity of the MN, as well as improving network level activity (Johnson et al., 2007). Therefore, without the appropriate input neurons or supporting cells to promote maturity, iPSC-derived MNs will always face scrutiny in one major area: How can these cell models, which are comparable in some ways to the embryonic stage of development, be used to investigate a set of diseases which are broadly linked with ageing?

A valid question and, as with many problems in science, not an easy one to answer. However, iPSC-derived MNs can provide excellent potential for longevity studies, where cultures can be maintained for several months and the impacts of disease development scrutinised over time. This is especially relevant when considering on average, 55 % of MNs are lost in the patient at the time of diagnosis, with a range up to 90 % neuronal loss (Ravits *et al.*, 2007). Therefore, it is likely neuronal death has been occurring for a time prior to symptom development, with considerable compensation occurring in the body. The investigation into the time course of disease progression or identification of potential biomarkers (Mendez and Sattler, 2015; Feneberg *et al.*, 2018; Si *et al.*, 2021) could be revolutionary in the diagnosis and early treatment of neurodegenerative diseases. With the problems posed by accessibility of neuronal tissue for research and drug development, iPSC-derived cell models overcome several of the challenges previously faced in neurodegenerative research.

A substantial amount of time and thought went into the development of such a model occurred prior to the pandemic. This resulted in constructs which can be used in the future, should this avenue of investigation be picked up again. Furthermore, UBQLN2 expression in motor neurons was confirmed and the MNs were not sensitive to the stress assays. Whether the mutant lines would have been as robust, either to culture as iPSCs or when differentiated, remains uncertain. Nonetheless, the development of these skills and resources, combined with the establishment of preliminary assays, provided a good foundation which can be built upon.

8.2 UBQLN2 as a trimer

Multimerization has previously been observed in both full length and N-terminally truncated UBQLN2 (Hjerpe *et al.*, 2016; Alexander *et al.*, 2018; Dao *et al.*, 2018). The experimental work in this study originally set out to use a portion of the native protein, to first validate the hypothesis under investigation. The reductionist model was used to determine the propensity of the PRR region to form a multimer, and potentially elucidate the order of this multimerization. Whilst previous reports have identified the STI1-II region as governing one mode of oligomerization (LLPS)(Dao *et al.*, 2018), this study has identified another mode of oligomerization, likely driven by the PRR region.

The MW estimations generated during AUC analysis place the multimeric species approximately 2.5- and 5.5-fold larger than the monomeric species. AUC-SV relies on a series of assumptions regarding the shape of the macromolecule under investigation. Thus, it is difficult to conclude with true certainty the size of the species identified. If the hypothesis regarding the overall structure of the trimer forming a rod-like extension is correct, it is possible this shape would slow the sedimentation of the protein more than a globular protein 3x the size of the monomer. This would result in a larger predicted MW being reported. Indeed, the large frictional ratio calculated during SE experimentation does suggest an elongated protein is formed when existing as a trimer. As previous investigations into the multimerization of the full-length UBQLN2 revealed no species larger than trimeric (Hjerpe *et al.*, 2016), this theory of a rod-like extension is the best fit of the data.

The presence of numerous multimeric species in a single FPLC peak was intriguing. One possible explanation is the lack of a nucleation domain causing undirected, misfolding of any structure. This possibility was accounted for with protein engineering, adding the foldon domain N-terminally to direct trimerization. A second possibility is any structures formed by the PRR region may be less stable than other collagen like proteins. Due to the vast array of roles UBQLN2 has within the cell, the multimerization would likely need to be reversible to allow these varied functions to be fulfilled. The collagen-like sequence of UBQLN2 is composed of twelve $(Gly\text{-}Xaa\text{-}Yaa)_n$ repeats with three interruptions. Increasing numbers of

interruptions has been shown to have a destabilising effect on collagen-like proteins (Dölz, Engel and Kühn, 1988). However, given the numerous and varied roles UBQLN2 has within the cell, it would require more plasticity in when and where it is assembled and disassembled than a characteristic collagen-like protein. Regions of collagen with low 4-Hyp, and therefore lower stability, are capable of relaxing (Brodsky and Persikov, 2005); these regions can rapidly fold and unfold to retain flexibility in the collagen. Should a triple helix form from UBQLN2 monomers, this may be an example of why an interrupted sequence is favourable over a more rigid one. As a consequence, the structure may be more difficult to predict computationally or identify experimentally.

8.3 Determining the structure of trimeric UBQLN2

Several approaches were originally attempted to elucidate any structure formed in the FPRR protein. These included Small Angle X-ray Scattering (SAXS), X-ray crystallography, and NMR spectroscopy, with the latter producing the data reported in this study.

Despite encountering no problems with solubility during purification or NMR spectroscopy experiments, issues arose when attempting to employ SAXS and crystallography techniques. During sample preparation for SAXS, the protein crashed out. Similarly, numerous trays for x-ray crystallography yielded only large salt crystals, unless left for a very long period of time (~130 days), which was not reproducible given the time constraints of this study. Whilst SAXS could not provide as much detail on structure as NMR or x-ray crystallography, it could have informed on the overall shape of the molecule. This may have confirmed or refuted the working hypothesis that trimerization via the PRR region forms a rod-like extension from the UBA domains. X-ray crystallography analysis may have complemented NMR spectroscopy analysis or provided an alternate conclusion. UBQLN2 is largely cytoplasmic and soluble (N'Diaye *et al.*, 2009). X-ray crystallography would have elucidated the structure in a fixed, crystallized state, which may not have been representative of the soluble protein. It was for this reason both structural investigations were attempted in tandem. However, due to issues obtaining crystals, x-ray crystallography was not pursued further. Future

investigations may look to optimise these techniques further to gain new perspective on the propensity of the PRR region to form a triple helix.

No problems were encountered during sample preparation for NMR spectroscopy, and the structure and folding of the FPRR protein was investigated. The foldon and UBA domains were correctly folded, confirming trimeric expression of the PRR-UBA regions of UBQLN2 does not perturb their structure.

The majority of secondary structure calculation programmes predict the PRR region to be largely disordered. Data produced in Chapter 4 provides evidence that a polyproline-II helical secondary structure is formed by the PRR region. DANGLE predictions were executed on the entire FPRR protein, using experimental chemical shift data from this study when available (Cheung *et al.*, 2010). The majority of dihedral angle predictions for the PRR region fell within the range of those determined for a collagen triple helix. However, the dihedral angles of collagen triple helices and polyproline-II helices share significant portions of Ramachandran space (Bhattacharjee and Bansal, 2005). Therefore, distinguishing between these two structures with this information alone is not possible. To fully determine the structure of the PRR region, further experimental investigations are required to visualise the proline and repeated residues in the PRR region. This includes 3D carbonyl-carbon detected experiments (Felli and Pierattelli, 2014), as opposed to $^{1}H_{N}$ detected experiments. These experiments would reveal information on proline residues that would be otherwise unavailable, but new optimisation on a higher field magnet or different probe is required.

8.4 Hypothesising the functional role of a UBQLN2 trimer

The PRR region of UBQLN2 has often been disregarded as disordered, due to computational predictions (Hjerpe *et al.*, 2016; Dao *et al.*, 2018). Using *AlphaFold* as an example, this study has identified why dismissal of such regions based solely upon a predicted structure is a restrictive practice in structural biology. Furthermore, this study has gathered evidence through CD spectroscopy that the PRR region is structured as a polyproline-II helix, a common precursor for the collagen-like triple helix. AUC analysis suggested a multimeric form of the protein exists, distinct from that of an aggregate, which may exist as an elongated trimer.

Investigations using NMR spectroscopy identified dihedral angle predictions concurrent with a collagen-like triple helix in the obligate trimer protein FPRR, although the conclusiveness of this data could not be determined. Finally, a possible reduction in the ability to bind ubiquitin when expressed as a trimer was detected. However, suitable controls are required to fully elucidate the nature of this change in binding.

The length of the linker between the PRR region and UBA domains is relatively large at 44 residues. Linkers three to five residues in length are generally sufficient to decouple the motion of the domains, allowing them to tumble and function independently (Reddy Chichili, Kumar and Sivaraman, 2013). This larger linker length could suggest the effect of trimerization may not be observed directly in the function of the UBA domain, but somewhere closer in the protein's domain architecture, such as in the flexible STI1-II domain, only 24 residues away. The STI1-II domain was originally predicted to be disordered (Hjerpe *et al.*, 2016; Dao *et al.*, 2018), but recent *AlphaFold* analysis predicts these cargo-binding regions to have a helical structure, albeit with a mixed per-residue confidence in this determination. Nonetheless, trimerization may improve organisation of the cargo-binding sites, such as that with HSP70 (Kaye *et al.*, 2000; Hjerpe *et al.*, 2016). In ALS-mutants, HSP70 interaction is disrupted (Hjerpe *et al.*, 2016), potentially due to poor organization of these binding sites as a consequence of perturbed trimerization.

There is evidence of interaction between the STI1-II and PRR regions (Dao *et al.*, 2018). The STI1-II domain was removed in this study, but it is possible that trimerization protects UBQLN2 from oligomerization via LLPS, and thus prevents it from stress granule recruitment. An ALS-causing mutant may prevent this trimeric structure forming, and this protective feature lost, resulting in an increase of LLPS and puncta formed (Dao *et al.*, 2019). Furthermore, the STI1-II domain is a low complexity domain with prion-like characteristics (Lancaster *et al.*, 2014; Zheng, Yang and Castañeda, 2020). Proteins with similar prion-like sequences have a tendency to form pathological ß-sheet aggregates (Martin and Mittag, 2018). Trimerization via the PRR region may organise these STI1-II domains in a way which prevents this. As a consequence, ALS-causing mutations in the PRR region may impact on the protective organisation of the domains and lead to a dysfunction over time in a two-pronged problem: First, an increase in aggregation

overall within the cell. Secondly, a decrease in the available, functional UBQLN2 to clear aggregates of its own along with aggregates from other proteins.

If genuine, the reduction in ubiquitin binding in the context of a trimer could be causal or coincidental. Ubiquitin binding may be impacted as a coincidental consequence of the protection from LLPS provided during trimerization, hypothesised above. However, this reduction in binding may serve a purpose: to direct UBQLN2 towards a cellular role independent of the UPS. Although not as well characterised, UBQLN2 has numerous roles outside of the UPS, such as in the autophagy pathways (Xu et al., 2013; Xia et al., 2014; Chuang et al., 2016), as well as in ER-Golgi trafficking and DNA/RNA metabolism (Gilpin, Chang and Monteiro, 2015; Halloran et al., 2019). UBQLN2 stabilises Heterogeneous Nuclear Ribonucleoprotein A1 (hnRNPA1) (Gilpin, Chang and Monteiro, 2015) and plays a role in ER-Golgi trafficking (Halloran *et al.*, 2019) in two independent processes. However, ALS-causing mutations within the PRR region both reduce binding to hnRNPA1 and alter its subcellular localisation (Gilpin, Chang and Monteiro, 2015) as well as inhibit protein transport from the ER to the Golgi in neuronal cells, causing a disorganised and fragmented Golgi (Halloran *et al.*, 2019). The role of multimeric structures in these processes has not been investigated, though may provide unique insights into the functional role of UBQLN2 oligomerization within the cell.

In support of this hypothesis, the yeast homologue of UBQLN2, Dsk2, which lacks the PRR region found in UBQLN2, only exists as a monomer or dimer (Sasaki *et al.*, 2005). Dsk2 mainly directs clients to the proteasome for degradation under normal circumstances (Wilkinson *et al.*, 2001; Funakoshi *et al.*, 2002). However, modifications to enable Dsk2 to oligomerize result in direction of proteins to the autophagy, not UPS, pathway (Lu, den Brave and Jentsch, 2017). This work demonstrates the simple principle that oligomerization of UBL-UBA proteins alters the function of the protein within the cell. Given the multiple methods by which UBQLN2 may oligomerize, it is possible trimerization via the PRR region may direct UBQLN2 to the even broader range of pathways currently identified, such as DNA/RNA metabolism or ER-golgi trafficking (Gilpin, Chang and Monteiro, 2015; Halloran et al., 2019). In these pathways, the mechanism of binding is not ubiquitin dependent, so a coincidental decrease in ubiquitin binding as a consequence of multimerization would not cause a problem functionally. In

contrast, ubiquitin binding may be reduced in the trimer to allow function in the pathways outside of degradation, without being recruited back to the UPS through strong ubiquitin binding. Similarly, ubiquitin binding has been demonstrated to inhibit oligomerization via other domains. Therefore, a reduction in ubiquitin binding may be preferable to allow oligomerization, and the downstream functionalities created by this, to continue.

8.5 Future directions and concluding remarks

Whilst a lot of the above hypotheses are largely conjecture, they are worth considering to continue research effectively. For instance, should trimerization affect the role of UBQLN2 in the cell, *in vitro* cellular assays could be completed to identify binding partners. Affinity-labelling using biotinylation and mass spectrometry could identify transient binding partners of UBQLN2 and lead to identification of which pathways are involved under particular cellular conditions or identify potential triggers for trimerization. Furthermore, should oligomerization affect the localisation of UBQLN2, this could be detected and explored further using the stress assays and microscopy techniques developed in Chapter 3.

Questions surrounding the stability of the collagen triple helix have often been investigated. Post-translational hydroxylation of the proline (4-Hyp) in the Yaa position of the $(Gly\text{-}Xaa\text{-}Yaa)_n$ repeat has been identified as the best stabilisation technique for mammalian collagens (Shoulders and Raines, 2009). The lack of any prolines in this position in the PRR region rules this stabilisation technique out for UBQLN2 if comparing it to a mammalian collagen. However, there are reports of 4-Hyp occurring in the Xaa position of some invertebrate collagens, such as the previously mentioned vent worm (Mann *et al.*, 1996). The effect of this 4R-hydroxylation on stability of collagen-like peptides was investigated and found to be acceptable for triple helix formation, provided proline does not occupy the Yaa position (Bann and Bachinger, 2000; Mizuno, Hayashi and Bächinger, 2003; Mizuno *et al.*, 2004). None of the $(Gly\text{-}Xaa\text{-}Yaa)_n$ repeats in the PRR region contain a proline in the Yaa position. Furthermore, increased stability in the helix is provided when the Yaa position is accommodated by threonine, valine, and alanine (Kimura *et al.*, 1989; Bann and Bachinger, 2000; Mizuno, Hayashi and

Bächinger, 2003; Mizuno *et al.*, 2004). Of the 12 $(Gly\text{-}Xaa\text{-}Yaa)_n$ repeats in the PRR region, half fulfil this criteria, suggesting potential elements of stability may be conferred in the PRR region different to classical collagen stability.

Another mammalian prolyl hydroxylase exists, distinct from prolyl-4-hydroxylase, called Prolyl-3-hydroxylase (Tryggvason, Risteli and Kivirikko, 1976). Post-translational modification by this enzyme produces 3S, 2S-hydroxyproline (3-hydroxyproline; 3-Hyp). Originally thought to be destabilising (Jenkins *et al.*, 2003), further investigation has revealed a marginal improvement in collagen triple helix stability with 3-hydroxyproline (Mizuno *et al.*, 2008). Notably, the replacement of Pro with 3-Hyp when Pro occupies the Xaa position stabilises the collagen triple helix (Jenkins *et al.*, 2003; Mizuno *et al.*, 2008). As the prolines occurring in the PRR region of UBQLN2 all fall in this Xaa position, it is possible that post-translational modification via this residue provides an element of stability over the structures formed. Interestingly, deficiency in 3S-hydroxylation result in disease phenotype (Cabral *et al.*, 2007), providing further evidence that 3S-hydroxylation may play a larger role in protein stability and function than previously described. There are few 3-Hyp per collagen chain, whilst ~100 4-Hyp per chain exist, known to create stability in the helix. Despite its rarity, the collagen modification 3-Hyp is highly conserved across the animal kingdom (Weis *et al.*, 2010; Hudson, Weis and Eyre, 2011) and even found in the most primitive extant multicellular animal (Ehrlich *et al.*, 2010). It is unlikely such conservation would exist without a role in protein structure and function. However, this role may be more nuanced than simply providing stability. For example, 3-Hyp may play a role in intermolecular recognition and binding; there is evidence of self-association dependant on 3-hydroxylation of Pro986 in Col 1 α1 chain (Hudson *et al.*, 2012).

Together, the above information demonstrates how little is still understood about the role of post-translational modifications of collagen and collagen-like peptides. This is an important avenue to investigate further as both prolyl-3-hydroxylase and prolyl-4-hydroxylase are mammalian enzymes, not naturally expressed in bacteria. It is therefore possible that stable trimeric helices will not form in the PRR region when expressing recombinant proteins without these enzymes. Thus, determination of the structure may be more difficult without this potential stabilisation occurring.

Another avenue of investigation considered in this study, but unable to be executed, was the persistence of a trimeric structure when the nucleation event is removed. A mechanism for determining how stable any structure formed is without foldon would provide information which could be used to infer the biological role of UBQLN2 trimerization within the neurons. Primers were designed to insert a chemical cleavage site (sequence-specific nickel-assisted cleavage (SNAC)-tag)(Dang *et al.*, 2019) into the linker region between the foldon and PRR regions. A chemical cleavage site was chosen due to the potential obscuring of the cleavage site if a structure is formed. This structure could prevent proteases from accessing the cleavage site, resulting in inefficient cleavage. The SNAC-tag would lie within the nested restriction sites, thus retaining a level of manipulation over the linker region if this was desired in downstream investigations. Whilst the design strategy was verified *in silico*, time and resource constraints prevented any further progression of this construct. If a structure is determined within the PRR region of UBQLN2, future experiments may benefit from using this system to elucidate a physiological role for the trimeric protein.

As outlined earlier, high proline content is proposed to be one of the stabilisation mechanisms in play in bacterial collagen-like proteins. Intriguingly, most ALS-causing mutations alter prolines within the PRR region, but outside of the (Gly-Xaa-Yaa)$_n$ repeats. A substitution in these prolines may fatally destabilise the triple helix, whilst ALS-causing mutations may perturb the structure of the helix to impact on functionality of UBQLN2. Primers were designed to introduce a series of ALS-causing mutations (P497H, P506T, P525S) with a range of severity (Deng *et al.*, 2011; Dao *et al.*, 2019) and ubiquitin non-binding mutants (L619A, F594V) as controls for ubiquitin binding analyses. Site-directed mutagenesis was completed to introduce the ALS-causing mutation (P506T) into the PRR region of HFPRR. If more time was available, the structural and functional behaviour of this mutant would have been compared to the wild-type in order to infer a potential disease mechanism for the development of ALS. The differences would have been investigated using CD spectroscopy, AUC, and NMR spectroscopy.

To conclude, the work presented in this chapter successfully applied the objectives of generating a reductionist model of the PRR region and UBA domains of UBQLN2 to investigate structure and ubiquitin binding function. The technology developed during this study has enabled the identification of a previously

undescribed secondary structure for the PRR region. The knowledge obtained provides strong justification for further investigation into the structure of the PRR region and the proteins engineered provide an excellent springboard for this research. This thesis presents an important step in determining the structure of the PRR region, a region crucial in the development of ALS. By combining computational modelling with biological experimentation, new perspectives were gained on the structure and possible function of UBQLN2 in the development of ALS.

Appendices

Appendix I List of Chemical Shifts.

	H	N	C	C_a	C_b	
452 Gly			173.18	45.03		
453 Tyr	7.29	119.74	175.97	55.90	38.65	
454 Ile	8.26	118.37	173.78	59.05	38.71	
455 Pro			175.54	61.77	33.07	
456 Glu	7.73	118.64	175.65	55.95	30.88	
457 Ala	8.22	131.13	174.06	50.98	14.87	
459 Arg			174.71	54.63	28.71	
460 Asp	8.28	123.26	177.56	52.80	41.13	
461 Gly	9.56	111.45	174.44	45.67		
462 Gln	8.27	118.19	174.24	53.61	30.83	
463 Ala	7.75	120.81	177.51	50.45	20.71	
464 Tyr	9.44	121.57	175.09	59.17	43.75	
465 Val	9.52	111.85	173.62	59.26	33.30	
466 Arg	8.37	124.35	174.55	53.65	33.16	
467 Lys	8.91	127.92	175.46	55.88	37.27	
468 Asp	10.74	130.19	175.40	56.59	40.61	
469 Gly	9.21	104.40	172.71	45.65		
470 Glu	7.69	117.17	175.72	54.05	33.66	
471 Trp						N^{e1} 131.22, H^{e1} 10.17
471 Trp	8.89	122.45	175.82	55.93	29.96	
472 Val	9.73	122.63	175.16	59.69	35.27	
473 Leu	8.65	126.91	178.77	55.74	42.15	
474 Leu	8.45	126.05	178.37	57.58	41.66	
475 Ser	8.93	112.83	176.09	61.15	62.21	
476 Thr	7.11	116.54	174.95	64.91	68.86	
477 Phe	7.78	117.93	175.65	58.88	40.32	
478 Leu	7.17	118.00	177.24	54.82	42.20	
479 Gly	8.05	108.27		45.62		
491 Pro			177.11	63.16	32.06	
492 Val	8.11	120.10	176.29	62.10	32.75	
493 Gly	8.10	112.65	171.56	44.45		
503 Pro			176.79	62.97	32.07	
504 Ile	8.12	121.55	176.10	61.10	38.52	
505 Val	8.02	125.86		59.44	32.70	
521 Pro			176.63	63.07	32.10	
522 Ala	8.22	124.62	176.96	52.08	19.25	
523 Ala	8.06	124.94	174.99	50.19	18.32	
532 Gly						
543 Pro			176.19	63.40	32.05	
544 Ser	7.91	122.25	178.69	60.07	64.95	
546 Thr			174.00	61.56	69.93	
547 Thr	7.76	122.15	179.10	63.32	70.59	
569 Gly						
578 Pro			176.49	65.42	32.14	
579 Glu			177.12	59.17	28.99	
580 Val	7.28	118.44	178.50	64.41	32.47	
581 Arg	7.87	121.42	177.79	58.56	30.56	
582 Phe	7.93	114.64	175.49	54.72	38.17	
583 Gln	7.23	120.88	178.21	60.68	28.80	
584 Gln			178.86	59.45	28.24	
585 Gln			178.49	60.74	29.02	

		H	N	C	C_a	C_b		
586	Leu	8.68	120.11	180.29	58.43	41.40		
587	Glu	7.86	121.27	179.58	59.42	29.26		
588	Gln	8.11	121.72	178.88	59.23	28.35		
589	Leu	8.51	119.32	179.46	58.38	41.32		
590	Asn	7.87	118.80	178.87	57.30	39.50		
591	Ala	8.00	123.15	179.14	54.71	17.88		
592	Met	7.64	115.99	175.64	56.32	34.61		
592	Met							
593	Gly			173.49	45.08			
594	Phe	7.85	122.50	176.30	58.08	36.85		
595	Leu	8.13	123.24	177.30	55.79	42.24		
596	Asn			173.71	52.16	37.82		
597	Arg			177.42	60.31	30.57		
598	Glu	8.15	118.09	179.03	59.96	29.16		
599	Ala	7.51	123.35	180.81	54.45	17.47		
599	Ala							
600	Asn	8.50	120.67	177.22	55.67	38.05		
601	Leu	8.35	120.75	178.22	58.20	41.64		
602	Gln	7.86	116.77	179.14	59.00	28.05		
603	Ala	8.00	122.30	178.91	55.35	17.87		
604	Leu	7.95	118.30	179.01	57.14	41.79		
605	Ile	8.54	120.62	179.94	65.98		$C^{?}$ [280]	38.65
606	Ala	7.70	121.44	179.22	54.70	18.32		
607	Thr	7.58	105.52	176.25	60.62	69.64		
608	Gly	7.66	112.60	175.70	46.96			
609	Gly			172.25	45.50			
610	Asp		120.67	176.81	53.50	41.67		
610	Asp	6.82						
611	Ile	8.64	129.71	175.59	64.66	38.60		
612	Asn	7.85	120.42	178.10	56.84	37.83		
613	Ala	7.99	123.51	180.47	54.16	18.31		
614	Ala	8.27	123.34	178.65	55.52	17.09		
615	Ile	8.14	119.34	177.17	66.25	38.23		
616	Glu			179.69	59.75	29.46		
617	Arg			179.31	58.42	29.93		
618	Leu	8.24	122.54	179.85	57.49	42.66		
619	Leu	8.26	119.34	178.81	56.29	42.46		
620	Gly			174.33	45.89			
623	Pro			175.20	62.25	33.03		
624	Ser	7.87	125.38	178.44	59.99	64.97		

References

Adzhubei, A.A. *et al.* (1987) “Third type of secondary structure: Noncooperative mobile conformation. Protein Data Bank analysis,” *Biochemical and Biophysical Research Communications*, 146(3), pp. 934–938. doi:10.1016/0006-291X(87)90736-4.

Adzhubei, A.A. and Sternberg, M.J.E. (1993) “Left-handed Polyproline II Helices Commonly Occur in Globular Proteins,” *Journal of Molecular Biology*, 229(2), pp. 472–493. doi:10.1006/JMBI.1993.1047.

Adzhubei, A.A. and Sternberg, M.J.E. (1994) “Conservation of polyproline II helices in homologous proteins: implications for structure prediction by model building,” *Protein science : a publication of the Protein Society*, 3(12), pp. 2395–2410. doi:10.1002/PRO.5560031223.

Adzhubei, A.A., Sternberg, M.J.E. and Makarov, A.A. (2013) “Polyproline-II helix in proteins: Structure and function,” *Journal of Molecular Biology*, 425(12), pp. 2100–2132. doi:10.1016/J.JMB.2013.03.018.

Aitio, O. *et al.* (2010) “Recognition of tandem PxxP motifs as a unique Src homology 3-binding mode triggers pathogen-driven actin assembly,” *Proceedings of the National Academy of Sciences of the United States of America*, 107(50), pp. 21743–21748. doi:10.1073/PNAS.1010243107/SUPPL_FILE/PNAS.201010243SI.PDF.

Albert, S. *et al.* (2017) “Proteasomes tether to two distinct sites at the nuclear pore complex,” *Proceedings of the National Academy of Sciences of the United States of America*, 114(52), pp. 13726–13731. doi:10.1073/PNAS.1716305114/-/DCSUPPLEMENTAL.

Al-Chalabi, A. *et al.* (2012) “The genetics and neuropathology of amyotrophic lateral sclerosis,” *Acta neuropathologica*, 124(3), pp. 339–352. doi:10.1007/S00401-012-1022-4.

Alexander, E.J. *et al.* (2018) “Ubiquilin 2 modulates ALS/FTD-linked FUS–RNA complex dynamics and stress granule formation,” *Proceedings of the National Academy of Sciences of the United States of America*, 115(49), pp. E11485–E11494. doi:10.1073/PNAS.1811997115/-/DCSUPPLEMENTAL.

Al-Hakim, A. *et al.* (2010) “The ubiquitous role of ubiquitin in the DNA damage response,” *DNA repair*, 9(12), pp. 1229–1240. doi:10.1016/J.DNAREP.2010.09.011.

Al-Hakim, A.K. *et al.* (2008) “Control of AMPK-related kinases by USP9X and atypical Lys29/Lys33-linked polyubiquitin chains,” *Biochemical Journal*, 411(2), pp. 249–260. doi:10.1042/BJ20080067.

Anazawa, Y. *et al.* (2005) “PCOTH, a novel gene overexpressed in prostate cancers, promotes prostate cancer cell growth through phosphorylation of oncoprotein TAF-Ibeta/SET,” *Cancer research*, 65(11), pp. 4578–4586. doi:10.1158/0008-5472.CAN-04-4564.

Antonioli, M. *et al.* (2014) “AMBRA1 interplay with cullin E3 ubiquitin ligases regulates autophagy dynamics,” *Developmental cell*, 31(6), pp. 734–746. doi:10.1016/J.DEVCEL.2014.11.013.

Bachert, B.A. *et al.* (2015) “A Unique Set of the Burkholderia Collagen-Like Proteins Provides Insight into Pathogenesis, Genome Evolution and Niche Adaptation, and Infection Detection,” *PLoS ONE*, 10(9), p. e0137578. doi:10.1371/journal.pone.0137578.

BÄCHINGER, H.P. *et al.* (1980) “Folding mechanism of the triple helix in type-III collagen and type-III pN-collagen. Role of disulfide bridges and peptide bond isomerization,” *European journal of biochemistry*, 106(2), pp. 619–632. doi:10.1111/J.1432-1033.1980.TB04610.X.

Bann, J.G. and Bachinger, H.P. (2000) “Glycosylation/Hydroxylation-induced stabilization of the collagen triple helix. 4-trans-hydroxyproline in the Xaa position can stabilize the triple helix,” *The Journal of biological chemistry*, 275(32), pp. 24466–24469. doi:10.1074/JBC.M003336200.

Bard, J.A.M. *et al.* (2019) “The 26S Proteasome utilizes a kinetic gateway to prioritize substrate degradation,” *Cell*, 177(2), p. 286. doi:10.1016/J.CELL.2019.02.031.

Bax, A. and Clore, G.M. (2019) “Protein NMR: Boundless Opportunities,” *Journal of magnetic resonance (San Diego, Calif. : 1997)*, 306, p. 187. doi:10.1016/J.JMR.2019.07.037.

Baxt, L.A., Garza-Mayers, A.C. and Goldberg, M.B. (2013) “Bacterial subversion of host innate immune pathways,” *Science (New York, N.Y.)*, 340(6133), pp. 697–701. doi:10.1126/SCIENCE.1235771.

Beal, R. *et al.* (1996) “Surface hydrophobic residues of multiubiquitin chains essential for proteolytic targeting,” *Proceedings of the National Academy of Sciences of the United States of America*, 93(2), pp. 861–866. doi:10.1073/PNAS.93.2.861.

Beal, R.E. *et al.* (1998) “The Hydrophobic Effect Contributes to Polyubiquitin Chain Recognition†,” *Biochemistry*, 37(9), pp. 2925–2934. doi:10.1021/BI972514P.

Bella, J. *et al.* (1994) “Crystal and molecular structure of a collagen-like peptide at 1.9 A resolution,” *Science (New York, N.Y.)*, 266(5182), pp. 75–81. doi:10.1126/SCIENCE.7695699.

Benco, L. *et al.* (2001) “Upper Limit of the O−H···O Hydrogen Bond. Ab Initio Study of the Kaolinite Structure,” *Journal of Physical Chemistry B*, 105(44), pp. 10812–10817. doi:10.1021/JP0124802.

Berisio, R. *et al.* (2002) “Crystal structure of the collagen triple helix model [(Pro-Pro-Gly)10]3,” *Protein Science : A Publication of the Protein Society*, 11(2), p. 262. doi:10.1110/PS.32602.

Berisio, R. *et al.* (2006) “Polyproline Helices in Protein Structures: A Statistical Survey,” *Protein & Peptide Letters*, 13(8), pp. 847–854. doi:10.2174/092986606777841154.

Berisio, R. *et al.* (2009) “Role of side chains in collagen triple helix stabilization and partner recognition,” *Journal of Peptide Science*, 15(3), pp. 131–140. doi:10.1002/PSC.1082.

Berjanskii, M. v., Neal, S. and Wishart, D.S. (2006) “PREDITOR: a web server for predicting protein torsion angle restraints,” *Nucleic Acids Research*, 34(suppl_2), pp. W63–W69. doi:10.1093/NAR/GKL341.

Bermel, W. *et al.* (2006) “Novel 13C direct detection experiments, including extension to the third dimension, to perform the complete assignment of proteins,” *Journal of Magnetic Resonance*, 178(1), pp. 56–64. doi:10.1016/J.JMR.2005.08.011.

Bertolaet, B.L. *et al.* (2001a) “UBA domains mediate protein-protein interactions between two DNA damage-inducible proteins,” *Journal of Molecular Biology*, 313(5), pp. 955–963. doi:10.1006/JMBI.2001.5105.

Bertolaet, B.L. *et al.* (2001b) “UBA domains of DNA damage-inducible proteins interact with ubiquitin,” *Nature structural biology*, 8(5), pp. 417–422. doi:10.1038/87575.

Bertram, L. *et al.* (2005) “Family-based association between Alzheimer’s disease and variants in UBQLN1,” *The New England journal of medicine*, 352(9), pp. 884–894. doi:10.1056/NEJMOA042765.

Bhardwaj, A. *et al.* (2008) “Foldon-guided self-assembly of ultra-stable protein fibers,” *Protein Science : A Publication of the Protein Society*, 17(9), p. 1475. doi:10.1110/PS.036111.108.

Bhardwaj, A. *et al.* (2013) “Characterizing TDP-43 interaction with its RNA targets,” *Nucleic acids research*, 41(9), pp. 5062–5074. doi:10.1093/NAR/GKT189.

Bhattacharjee, A. and Bansal, M. (2005) “Collagen structure: The Madras triple helix and the current scenario,” *IUBMB Life*, 57(3), pp. 161–172. doi:10.1080/15216540500090710.

Bhowmick, M. and Fields, G.B. (2013) “Chapter 11 Stabilization of Collagen-Model, Triple-Helical Peptides for In Vitro and In Vivo Applications,” *Methods in Molecular Biology*, 1081, pp. 167–194. doi:10.1007/978-1-62703-652-8_11.

Biggins, S., Ivanovska, I. and Rose, M.D. (1996) “Yeast ubiquitin-like genes are involved in duplication of the microtubule organizing center.,” *Journal of Cell Biology*, 133(6), pp. 1331–1346. doi:10.1083/JCB.133.6.1331.

Birgisdottir, Å.B., Lamark, T. and Johansen, T. (2013) “The LIR motif - crucial for selective autophagy,” *Journal of cell science*, 126(Pt 15), pp. 3237–3247. doi:10.1242/JCS.126128.

Bjørkøy, G. *et al.* (2005) “p62/SQSTM1 forms protein aggregates degraded by autophagy and has a protective effect on huntingtin-induced cell death,” *Journal of Cell Biology*, 171(4), pp. 603–614. doi:10.1083/JCB.200507002/VIDEO-3.

Blokhuis, A.M. *et al.* (2013) “Protein aggregation in amyotrophic lateral sclerosis,” *Acta neuropathologica*, 125(6), pp. 777–794. doi:10.1007/S00401-013-1125-6.

Boisvert, F.M. *et al.* (2012) “A quantitative spatial proteomics analysis of proteome turnover in human cells,” *Molecular & cellular proteomics : MCP*, 11(3). doi:10.1074/MCP.M111.011429.

Boudko, S. *et al.* (2002) “Nucleation and propagation of the collagen triple helix in single-chain and trimerized peptides: Transition from third to first order kinetics,” *Journal of Molecular Biology*, 317(3), pp. 459–470. doi:10.1006/JMBI.2002.5439.

Boudko, S.P. *et al.* (2002) “Domain organization, folding and stability of bacteriophage T4 fibritin, a segmented coiled-coil protein,” *European journal of biochemistry*, 269(3), pp. 833–841. doi:10.1046/J.1432-1033.2002.02734.X.

Boudko, S.P. *et al.* (2008) “Crystal structure of human type III collagen Gly991-Gly1032 cystine knot-containing peptide shows both 7/2 and 10/3 triple helical symmetries,” *The Journal of biological chemistry*, 283(47), pp. 32580–32589. doi:10.1074/JBC.M805394200.

Boughton, A.J., Krueger, S. and Fushman, D. (2020) “Branching via K11 and K48 Bestows Ubiquitin Chains with a Unique Interdomain Interface and Enhanced Affinity for Proteasomal Subunit Rpn1,” *Structure (London, England : 1993)*, 28(1), pp. 29-43.e6. doi:10.1016/J.STR.2019.10.008.

Boutaud, A. *et al.* (2000) “Type IV collagen of the glomerular basement membrane. Evidence that the chain specificity of network assembly is encoded by the noncollagenous NC1 domains,” *The Journal of biological chemistry*, 275(39), pp. 30716–30724. doi:10.1074/JBC.M004569200.

Boydston, J.A. *et al.* (2005) “Orientation within the exosporium and structural stability of the collagen-like glycoprotein BclA of Bacillus anthracis,” *Journal of Bacteriology*, 187(15), pp. 5310–5317. doi:10.1128/JB.187.15.5310-5317.2005.

Brautigam, C.A. (2015) “Calculations and Publication-Quality Illustrations for Analytical Ultracentrifugation Data,” *Methods in enzymology*, 562, pp. 109–133. doi:10.1016/BS.MIE.2015.05.001.

Bremm, A., Freund, S.M.V. and Komander, D. (2010) “Lys11-linked ubiquitin chains adopt compact conformations and are preferentially hydrolyzed by the deubiquitinase Cezanne,” *Nature Structural & Molecular Biology 2010 17:8*, 17(8), pp. 939–947. doi:10.1038/nsmb.1873.

Bretscher, L.E. *et al.* (2001) “Conformational Stability of Collagen Relies on a Stereoelectronic Effect,” *Journal of the American Chemical Society*, 123(4), pp. 777–778. doi:10.1021/JA005542V.

Brodsky, B. and Persikov, A. v. (2005) “Molecular Structure of the Collagen Triple Helix,” *Advances in Protein Chemistry*, 70, pp. 301–339. doi:10.1016/S0065-3233(05)70009-7.

Bruijn, L.I., Miller, T.M. and Cleveland, D.W. (2004) “Unraveling the mechanisms involved in motor neuron degeneration in ALS,” *Annual review of neuroscience*, 27, pp. 723–749. doi:10.1146/ANNUREV.NEURO.27.070203.144244.

Cabral, W.A. *et al.* (2007) “Prolyl 3-hydroxylase 1 deficiency causes a recessive metabolic bone disorder resembling lethal/severe osteogenesis imperfecta,” *Nature genetics*, 39(3), pp. 359–365. doi:10.1038/NG1968.

Callis, J. (2014) “The Ubiquitination Machinery of the Ubiquitin System,” *The Arabidopsis Book / American Society of Plant Biologists*, 12, p. e0174. doi:10.1199/TAB.0174.

Cambridge, S.B. *et al.* (2011) “Systems-wide proteomic analysis in mammalian cells reveals conserved, functional protein turnover,” *Journal of Proteome Research*, 10(12), pp. 5275–5284. doi:10.1021/PR101183K/SUPPL_FILE/PR101183K_SI_002.XLS.

Caswell, C.C., Barczyk, M., *et al.* (2008) “Identification of the First Prokaryotic Collagen Sequence Motif That Mediates Binding to Human Collagen Receptors, Integrins α2β1 and α11β1,” *Journal of Biological Chemistry*, 283(52), pp. 36168–36175. doi:10.1074/JBC.M806865200.

Caswell, C.C., Han, R., *et al.* (2008) “The Scl1 protein of M6-type group A Streptococcus binds the human complement regulatory protein, factor H, and inhibits the alternative pathway of complement,” *Molecular Microbiology*, 67(3), pp. 584–596. doi:10.1111/J.1365-2958.2007.06067.X.

Chan, N.C. *et al.* (2011) “Broad activation of the ubiquitin-proteasome system by Parkin is critical for mitophagy,” *Human molecular genetics*, 20(9), pp. 1726–1737. doi:10.1093/HMG/DDR048.

Chang, L. and Monteiro, M.J. (2015) “Defective Proteasome Delivery of Polyubiquitinated Proteins by Ubiquilin-2 Proteins Containing ALS Mutations,” *PLoS ONE*, 10(6). doi:10.1371/JOURNAL.PONE.0130162.

Chaton, C.T. and Herr, A.B. (2015) “Elucidating complicated assembling systems in biology using size-and-shape analysis of sedimentation velocity data,” *Methods in enzymology*, 562, p. 187. doi:10.1016/BS.MIE.2015.04.004.

Chau, V. *et al.* (1989) “A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein,” *Science (New York, N.Y.)*, 243(4898), pp. 1576–1583. doi:10.1126/SCIENCE.2538923.

Chen, L. *et al.* (2001) “Ubiquitin-associated (UBA) domains in Rad23 bind ubiquitin and promote inhibition of multi-ubiquitin chain assembly,” *EMBO reports*, 2(10), pp. 933–938. doi:10.1093/EMBO-REPORTS/KVE203.

Chen, L. and Madura, K. (2002) “Rad23 promotes the targeting of proteolytic substrates to the proteasome,” *Molecular and cellular biology*, 22(13), pp. 4902–4913. doi:10.1128/MCB.22.13.4902-4913.2002.

Chen, R., Li, L. and Weng, Z. (2003) “ZDOCK: An initial-stage protein-docking algorithm,” *Proteins: Structure, Function, and Bioinformatics*, 52(1), pp. 80–87. doi:10.1002/PROT.10389.

Chen, T. *et al.* (2018) “Mutant UBQLN2 P497H in motor neurons leads to ALS-like phenotypes and defective autophagy in rats,” *Acta neuropathologica communications*, 6(1), p. 122. doi:10.1186/S40478-018-0627-9.

Chen, X. *et al.* (2019) “Structure of hRpn10 Bound to UBQLN2 UBL Illustrates Basis for Complementarity between Shuttle Factors and Substrates at the Proteasome,” *Journal of molecular biology*, 431(5), pp. 939–955. doi:10.1016/J.JMB.2019.01.021.

Cheung, M.S. *et al.* (2010) “DANGLE: A Bayesian inferential method for predicting protein backbone dihedral angles and secondary structure,” *Journal of Magnetic Resonance*, 202(2), pp. 223–233. doi:10.1016/J.JMR.2009.11.008.

Chiò, A. *et al.* (2013) “Global Epidemiology of Amyotrophic Lateral Sclerosis: a Systematic Review of the Published Literature,” *Neuroepidemiology*, 41(2), p. 118. doi:10.1159/000351153.

Chuang, K.H. *et al.* (2016) “Ubiquilin/Dsk2 promotes inclusion body formation and vacuole (lysosome)-mediated disposal of mutated huntingtin,” *Molecular Biology of the Cell*, 27(13), pp. 2025–2036. doi:10.1091/MBC.E16-01-0026/ASSET/IMAGES/LARGE/MBC-27-2025-G008.JPEG.

Ciechanover, A. *et al.* (1980) “ATP-dependent conjugation of reticulocyte proteins with the polypeptide required for protein degradation.,” *Proceedings of the National Academy of Sciences of the United States of America*, 77(3), p. 1365. doi:10.1073/PNAS.77.3.1365.

Clague, M.J. *et al.* (2013) “Deubiquitylases from genes to organism,” *Physiological Reviews*, 93(3), pp. 1289–1315. doi:10.1152/PHYSREV.00002.2013/SUPPL_FILE/TABLES2.PDF.

Clevers, H. and Nusse, R. (2012) “Wnt/ß-Catenin Signaling and Disease,” *Cell*, 149(6), pp. 1192–1205. doi:10.1016/J.CELL.2012.05.012.

Clubb, R.T., Thanabal, V. and Wagner, G. (1992) “A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue 1HN, 15N, and 13C′ chemical shifts in 15N□13C-labelled proteins,” *Journal of Magnetic Resonance (1969)*, 97(1), pp. 213–217. doi:10.1016/0022-2364(92)90252-3.

Cohen, L.D. *et al.* (2013) “Metabolic Turnover of Synaptic Proteins: Kinetics, Interdependencies and Implications for Synaptic Maintenance,” *PLOS ONE*, 8(5), p. e63191. doi:10.1371/JOURNAL.PONE.0063191.

Cohen-Kaplan, V. *et al.* (2016) “p62- and ubiquitin-dependent stress-induced autophagy of the mammalian 26S proteasome,” *Proceedings of the National Academy of Sciences of the United States of America*, 113(47), pp. E7490–E7499. doi:10.1073/PNAS.1615455113/SUPPL_FILE/PNAS.1615455113.SD01.XLSX.

Cole, J.L. *et al.* (2008) “Analytical Ultracentrifugation: Sedimentation Velocity and Sedimentation Equilibrium,” *Methods in Cell Biology*. NIH Public Access, pp. 143–179. doi:10.1016/S0091-679X(07)84006-4.

Collins, G.A. and Goldberg, A.L. (2017) “The Logic of the 26S Proteasome,” *Cell*, 169(5), p. 792. doi:10.1016/J.CELL.2017.04.023.

Collins, G.A. and Goldberg, A.L. (2020) “Proteins containing ubiquitin-like (Ubl) domains not only bind to 26S proteasomes but also induce their activation,” *Proceedings of the National Academy of Sciences of the United States of America*,

117(9), pp. 4664–4674. doi:10.1073/PNAS.1915534117/SUPPL_FILE/PNAS.1915534117.SAPP.PDF.

Compton, L.A. and Johnson, W.C. (1986) “Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication,” *Analytical biochemistry*, 155(1), pp. 155–167. doi:10.1016/0003-2697(86)90241-1.

Conklin, D. *et al.* (2000) “Molecular cloning, chromosome mapping and characterization of UBQLN3 a testis-specific gene that contains an ubiquitin-like domain,” *Gene*, 249(1–2), pp. 91–98. doi:10.1016/S0378-1119(00)00122-0.

Conley, M.P. and Wood, W.B. (1975) “Bacteriophage T4 whiskers: a rudimentary environment-sensing device,” *Proceedings of the National Academy of Sciences of the United States of America*, 72(9), pp. 3701–3705. doi:10.1073/PNAS.72.9.3701.

Cook, W.J. *et al.* (1992) “Structure of a diubiquitin conjugate and a model for interaction with ubiquitin conjugating enzyme (E2),” *The Journal of biological chemistry*, 267(23), pp. 16467–16471. doi:10.2210/PDB1AAR/PDB.

Coombs, D.H. and Eiserling, F.A. (1977) “Studies on the structure, protein composition and assembly of the neck of bacteriophage T4,” *Journal of Molecular Biology*, 116(3), pp. 375–405. doi:10.1016/0022-2836(77)90076-6.

Cornilescu, G., Delaglio, F. and Bax, A. (1999) “Protein backbone angle restraints from searching a database for chemical shift and sequence homology,” *Journal of Biomolecular NMR 1999 13:3*, 13(3), pp. 289–302. doi:10.1023/A:1008392405740.

Coux, O., Tanaka, K. and Goldberg, A.L. (1996) “Structure and functions of the 20S and 26S proteasomes,” *Annual review of biochemistry*, 65, pp. 801–847. doi:10.1146/ANNUREV.BI.65.070196.004101.

Cram, D.J. (1988) “The design of molecular hosts, guests, and their complexes,” *Journal of inclusion phenomena 1988 6:4*, 6(4), pp. 397–413. doi:10.1007/BF00658982.

Cubellis, M. v. *et al.* (2005) “Properties of polyproline II, a secondary structure element implicated in protein-protein interactions,” *Proteins*, 58(4), pp. 880–892. doi:10.1002/PROT.20327.

Dang, B. *et al.* (2019) “SNAC-tag for sequence-specific chemical protein cleavage,” *Nature Methods 2019 16:4*, 16(4), pp. 319–322. doi:10.1038/s41592-019-0357-3.

Dantuma, N.P. and Hoppe, T. (2012) “Growing sphere of influence: Cdc48/p97 orchestrates ubiquitin-dependent extraction from chromatin,” *Trends in cell biology*, 22(9), pp. 483–491. doi:10.1016/J.TCB.2012.06.003.

Dao, T.P. *et al.* (2018) “Ubiquitin Modulates Liquid-Liquid Phase Separation of UBQLN2 via Disruption of Multivalent Interactions,” *Molecular Cell*, 69(6), pp. 965-978.e6. doi:10.1016/j.molcel.2018.02.004.

Dao, T.P. *et al.* (2019) “ALS-Linked Mutations Affect UBQLN2 Oligomerization and Phase Separation in a Position- and Amino Acid-Dependent Manner,” *Structure (London, England : 1993)*, 27(6), pp. 937-951.e5. doi:10.1016/J.STR.2019.03.012.

DeJesus-Hernandez, M. *et al.* (2011) “Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS,” *Neuron*, 72(2), pp. 245–256. doi:10.1016/J.NEURON.2011.09.011.

Deng, H.X. *et al.* (2011) “Mutations in UBQLN2 cause dominant X-linked juvenile and adult-onset ALS and ALS/dementia,” *Nature*, 477(7363), pp. 211–215. doi:10.1038/nature10353.

Devlin, A.C. *et al.* (2015) “Human iPSC-derived motoneurons harbouring TARDBP or C9ORF72 ALS mutations are dysfunctional despite maintaining viability,” *Nature communications*, 6. doi:10.1038/NCOMMS6999.

Dikic, I. (2017) “Proteasomal and Autophagic Degradation Systems,” *Annual review of biochemistry*, 86, pp. 193–224. doi:10.1146/ANNUREV-BIOCHEM-061516-044908.

Dillen, L. *et al.* (2013) “Explorative genetic study of UBQLN2 and PFN1 in an extended Flanders-Belgian cohort of frontotemporal lobar degeneration patients,” *Neurobiology of aging*, 34(6), pp. 1711.e1-1711.e5. doi:10.1016/J.NEUROBIOLAGING.2012.12.007.

DÖLZ, R., ENGEL, J. and KÜHN, K. (1988) “Folding of collagen IV,” *European journal of biochemistry*, 178(2), pp. 357–366. doi:10.1111/J.1432-1033.1988.TB14458.X.

Dörrbaum, A.R. *et al.* (2018) “Local and global influences on protein turnover in neurons and glia,” *eLife*, 7. doi:10.7554/ELIFE.34202.

Douglas, P.M. and Dillin, A. (2010) “Protein homeostasis and aging in neurodegeneration,” *The Journal of cell biology*, 190(5), pp. 719–729. doi:10.1083/JCB.201005144.

Drake, A.F., Siligardi, G. and Gibbons, W.A. (1988) “Reassessment of the electronic circular dichroism criteria for random coil conformations of poly(L-lysine) and the implications for protein folding and denaturation studies,” *Biophysical chemistry*, 31(1–2), pp. 143–146. doi:10.1016/0301-4622(88)80019-X.

van Drie, J.H. (2011) “Protein folding, protein homeostasis, and cancer,” *Chinese Journal of Cancer*, 30(2), p. 124. doi:10.5732/CJC.010.10162.

Dukor, R.K. and Keiderling, T.A. (1991) “Reassessment of the random coil conformation: vibrational CD study of proline oligopeptides and related polypeptides,” *Biopolymers*, 31(14), pp. 1747–1761. doi:10.1002/BIP.360311409.

Duncan, C. *et al.* (2011) “Lcl of Legionella pneumophila is an immunogenic GAG binding adhesin that promotes interactions with lung epithelial cells and plays a crucial role in biofilm formation,” *Infection and Immunity*, 79(6), pp. 2168–2181. doi:10.1128/IAI.01304-10/ASSET/DB77D021-A910-46C5-8520-E6CB9BF2A3DA/ASSETS/GRAPHIC/ZII9990991610009.JPEG.

Dyson, H.J. and Wright, P.E. (2001) “Nuclear magnetic resonance methods for elucidation of structure and dynamics in disordered states,” *Methods in enzymology*, 339, pp. 258–270. doi:10.1016/S0076-6879(01)39317-5.

Dyson, H.J. and Wright, P.E. (2005) “Intrinsically unstructured proteins and their functions,” *Nature reviews. Molecular cell biology*, 6(3), pp. 197–208. doi:10.1038/NRM1589.

Ehlers, M.D. (2003) “Activity level controls postsynaptic composition and signaling via the ubiquitin-proteasome system,” *Nature neuroscience*, 6(3), pp. 231–242. doi:10.1038/NN1013.

Ehrlich, H. *et al.* (2010) “Mineralization of the metre-long biosilica structures of glass sponges is templated on hydroxylated collagen,” *Nature Chemistry 2010 2:12*, 2(12), pp. 1084–1088. doi:10.1038/nchem.899.

Eletr, Z.M. and Wilkinson, K.D. (2014) “Regulation of Proteolysis by Human Deubiquitinating Enzymes,” *Biochimica et biophysica acta*, 1843(1), pp. 114–128. doi:10.1016/J.BBAMCR.2013.06.027.

Elia, A.E.H. *et al.* (2015) “Quantitative Proteomic Atlas of Ubiquitination and Acetylation in the DNA Damage Response,” *Molecular Cell*, 59(5), pp. 867–881. doi:10.1016/J.MOLCEL.2015.05.006/ATTACHMENT/BFFA0EB8-AC9A-4B22-92B5-D0121EBD1ADF/MMC7.XLS.

Engel, J. and Bächinger, H.P. (2000) “Cooperative equilibrium transitions coupled with a slow annealing step explain the sharpness and hysteresis of collagen folding,” *Matrix Biology*, 19(3), pp. 235–244. doi:10.1016/S0945-053X(00)00071-8.

Engel, J. and Prockop, D.J. (1991) “The Zipper-Like Folding of Collagen Triple Helices and the Effects of Mutations that Disrupt the Zipper,” *Annu. Rev. Biophys. Biophys. Chem*, 20, pp. 137–152. doi:10.1146/ANNUREV.BB.20.060191.001033.

Evans, R. *et al.* (2021) “Protein complex prediction with AlphaFold-Multimer,” *bioRxiv*, p. 2021.10.04.463034. doi:10.1101/2021.10.04.463034.

Fahed, A.C. *et al.* (2014) “UBQLN2 Mutation Causing Heterogeneous X-linked Dominant Neurodegeneration,” *Annals of neurology*, 75(5), p. 793. doi:10.1002/ANA.24164.

Fallas, J.A. *et al.* (2012) “Structural Insights into Charge Pair Interactions in Triple Helical Collagen-like Proteins,” *The Journal of Biological Chemistry*, 287(11), p. 8039. doi:10.1074/JBC.M111.296574.

Feigin, V.L. *et al.* (2019) “Global, regional, and national burden of neurological disorders, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016,” *The Lancet Neurology*, 18(5), pp. 459–480. doi:10.1016/S1474-4422(18)30499-X/ATTACHMENT/2830E1E6-5A7C-42F2-AC83-0BD8437532C0/MMC1.PDF.

Felli, I.C. and Pierattelli, R. (2014) “Novel methods based on 13C detection to study intrinsically disordered proteins,” *Journal of Magnetic Resonance*, 241(1), pp. 115–125. doi:10.1016/J.JMR.2013.10.020.

Feneberg, E. *et al.* (2018) “Towards a TDP-43-Based Biomarker for ALS and FTLD,” *Molecular neurobiology*, 55(10), pp. 7789–7801. doi:10.1007/S12035-018-0947-6.

Fernandez-Vizarra, E. and Zeviani, M. (2021) “Mitochondrial disorders of the OXPHOS system,” *FEBS letters*, 595(8), pp. 1062–1106. doi:10.1002/1873-3468.13995.

Filimonenko, M. *et al.* (2010) “The Selective Macroautophagic Degradation of Aggregated Proteins Requires the PI3P-Binding Protein Alfy,” *Molecular Cell*, 38(2), pp. 265–279. doi:10.1016/J.MOLCEL.2010.04.007.

Fletcher, J.M. *et al.* (2012) “A basis set of de novo coiled-Coil peptide oligomers for rational protein design and synthetic biology,” *ACS Synthetic Biology*, 1(6), pp. 240–250. doi:10.1021/SB300028Q/SUPPL_FILE/SB300028Q_SI_001.PDF.

Forsén, S. and Huffman, R.A. (2004) “Study of Moderately Rapid Chemical Exchange Reactions by Means of Nuclear Magnetic Double Resonance,” *The Journal of Chemical Physics*, 39(11), p. 2892. doi:10.1063/1.1734121.

Frank, S. *et al.* (2001) “Stabilization of short collagen-like triple helices by protein engineering,” *Journal of Molecular Biology*, 308(5), pp. 1081–1089. doi:10.1006/JMBI.2001.4644.

Frank, S. *et al.* (2003) “Collagen Triple Helix Formation Can Be Nucleated at Either End *,” *Journal of Biological Chemistry*, 278(10), pp. 7747–7750. doi:10.1074/JBC.C200698200.

Fujimoto, K. *et al.* (2016) “Regulation of intestinal homeostasis by the ulcerative colitis-associated gene RNF186,” *Mucosal Immunology 2017 10:2*, 10(2), pp. 446–459. doi:10.1038/mi.2016.58.

Funakoshi, M. *et al.* (2002) “Budding yeast Dsk2p is a polyubiquitin-binding protein that can interact with the proteasome,” *Proceedings of the National Academy of Sciences of the United States of America*, 99(2), pp. 745–750. doi:10.1073/PNAS.012585199.

Gao, Y. *et al.* (2010) “The Scl1 of M41-type group A Streptococcus binds the high-density lipoprotein,” *FEMS Microbiology Letters*, 309(1), pp. 55–61. doi:10.1111/J.1574-6968.2010.02013.X.

Gatti, M. *et al.* (2015) “RNF168 Promotes Noncanonical K27 Ubiquitination to Signal DNA Damage,” *Cell Reports*, 10(2), pp. 226–238. doi:10.1016/J.CELREP.2014.12.021.

Gellera, C. *et al.* (2013) “Ubiquilin 2 mutations in Italian patients with amyotrophic lateral sclerosis and frontotemporal dementia,” *Journal of Neurology, Neurosurgery & Psychiatry*, 84(2), pp. 183–187. doi:10.1136/JNNP-2012-303433.

Gerlach, B. *et al.* (2011) “Linear ubiquitination prevents inflammation and regulates immune signalling,” *Nature 2011 471:7340*, 471(7340), pp. 591–596. doi:10.1038/nature09816.

Gilpin, K.M., Chang, L. and Monteiro, M.J. (2015a) “ALS-linked mutations in ubiquilin-2 or hnRNPA1 reduce interaction between ubiquilin-2 and hnRNPA1,” *Human molecular genetics*, 24(9), pp. 2565–2577. doi:10.1093/HMG/DDV020.

Gilpin, K.M., Chang, L. and Monteiro, M.J. (2015b) “ALS-linked mutations in ubiquilin-2 or hnRNPA1 reduce interaction between ubiquilin-2 and hnRNPA1,” *Human molecular genetics*, 24(9), pp. 2565–2577. doi:10.1093/HMG/DDV020.

Gkazi, S.A. *et al.* (2019) “Striking phenotypic variation in a family with the P506S UBQLN2 mutation including amyotrophic lateral sclerosis, spastic paraplegia, and frontotemporal dementia,” *Neurobiology of aging*, 73, pp. 229.e5-229.e9. doi:10.1016/J.NEUROBIOLAGING.2018.08.015.

Glickman, M.H. *et al.* (1998) “The regulatory particle of the Saccharomyces cerevisiae proteasome,” *Molecular and cellular biology*, 18(6), pp. 3149–3162. doi:10.1128/MCB.18.6.3149.

Glickman, M.H. and Ciechanover, A. (2002) “The Ubiquitin-Proteasome Proteolytic Pathway: Destruction for the Sake of Construction,” *Physiological Reviews*, 82(2), pp. 373–428. doi:10.1152/physrev.00027.2001.

Granot, J. (1983) “Determination of dissociation constants of 1:1 complexes from NMR data. Optimization of the experimental setup by statistical analysis of simulated experiments,” *Journal of Magnetic Resonance (1969)*, 55(2), pp. 216–224. doi:10.1016/0022-2364(83)90234-2.

Green, N.M. (1965) “A Spectrophotometric Assay for Avidin and Biotin Based on Binding of Dyes by Avidin,” *Biochemical Journal*, 94(3), pp. 23C-24C. doi:10.1042/BJ0940023C.

Grzesiek, S. and Bax, A. (1992) “Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein,” *Journal of Magnetic Resonance (1969)*, 96(2), pp. 432–440. doi:10.1016/0022-2364(92)90099-S.

Guerler, A., Govindarajoo, B. and Zhang, Y. (2013) “Mapping Monomeric Threading to Protein–Protein Structure Prediction,” *Journal of chemical information and modeling*, 53(3), p. 717. doi:10.1021/CI300579R.

Güthe, S. *et al.* (2004) “Very fast folding and association of a trimerization domain from bacteriophage T4 fibritin,” *Journal of molecular biology*, 337(4), pp. 905–915. doi:10.1016/J.JMB.2004.02.020.

Habazettl, J., Reiner, A. and Kiefhaber, T. (2009) “NMR structure of a monomeric intermediate on the evolutionarily optimized assembly pathway of a small trimerization domain,” *Journal of molecular biology*, 389(1), pp. 103–114. doi:10.1016/J.JMB.2009.03.073.

Halloran, M. *et al.* (2019) “Amyotrophic lateral sclerosis-linked UBQLN2 mutants inhibit endoplasmic reticulum to Golgi transport, leading to Golgi fragmentation and ER stress,” 1, p. 3. doi:10.1007/s00018-019-03394-w.

Han, R., Zwiefka, A., *et al.* (2006) “Assessment of prokaryotic collagen-like sequences derived from streptococcal Scl1 and Scl2 proteins as a source of recombinant GXY polymers,” *Applied Microbiology and Biotechnology*, 72(1), pp. 109–115. doi:10.1007/S00253-006-0387-5/FIGURES/4.

Han, R., Caswell, C.C., *et al.* (2006) “Binding of the low-density lipoprotein by streptococcal collagen-like protein Scl1 of Streptococcus pyogenes,” *Molecular Microbiology*, 61(2), pp. 351–367. doi:10.1111/J.1365-2958.2006.05237.X.

Hara, T. *et al.* (2006) “Suppression of basal autophagy in neural cells causes neurodegenerative disease in mice,” *Nature*, 441(7095), pp. 885–889. doi:10.1038/NATURE04724.

Harbury, P.B. *et al.* (1993) “A Switch Between Two-, Three-, and Four-stranded Coiled Coils in GCN4 Leucine Zipper Mutants,” *Science*, 262(5138), pp. 1401–1407. doi:10.1126/SCIENCE.8248779.

Harman, C.A. and Monteiro, M.J. (2019) “The specificity of ubiquitin binding to ubiquilin-1 is regulated by sequences besides its UBA domain,” *Biochimica et Biophysica Acta - General Subjects*, 1863(10), pp. 1568–1574. doi:10.1016/j.bbagen.2019.06.002.

He, J. *et al.* (2012) “The structure of the 26S proteasome subunit Rpn2 reveals its PC repeat domain as a closed toroid of two concentric α-helical rings,” *Structure*, 20(3), pp. 513–521. doi:10.1016/J.STR.2011.12.015/ATTACHMENT/6B44B44D-EAB3-42FE-B462-1C7C04190698/MMC1.PDF.

Heo, J.M. *et al.* (2015) “The PINK1-PARKIN Mitochondrial Ubiquitylation Pathway Drives a Program of OPTN/NDP52 Recruitment and TBK1 Activation to Promote Mitophagy,” *Molecular cell*, 60(1), pp. 7–20. doi:10.1016/J.MOLCEL.2015.08.016.

Hershko, A. *et al.* (1983) “Components of ubiquitin-protein ligase system. Resolution, affinity purification, and role in protein breakdown.,” *The Journal of biological chemistry*, 258(13), pp. 8206–8214.

Hershko, A. and Ciechanover, A. (1998) “The ubiquitin system,” *Annual review of biochemistry*, 67, pp. 425–479. doi:10.1146/ANNUREV.BIOCHEM.67.1.425.

Hicke, L., Schubert, H.L. and Hill, C.P. (2005) “Ubiquitin-binding domains,” *Nature reviews. Molecular cell biology*, 6(8), pp. 610–621. doi:10.1038/NRM1701.

Hicks, J.M. and Hsu, V.L. (2004) “The extended left-handed helix: a simple nucleic acid-binding motif,” *Proteins*, 55(2), pp. 330–338. doi:10.1002/PROT.10630.

Hirsilä, M. *et al.* (2003) “Characterization of the human prolyl 4-hydroxylases that modify the hypoxia-inducible factor,” *The Journal of biological chemistry*, 278(33), pp. 30772–30780. doi:10.1074/JBC.M304982200.

Hjerpe, R. *et al.* (2009) “Efficient protection and isolation of ubiquitylated proteins using tandem ubiquitin-binding entities,” *EMBO reports*, 10(11), pp. 1250–1258. doi:10.1038/EMBOR.2009.192.

Hjerpe, R. *et al.* (2016) “UBQLN2 Mediates Autophagy-Independent Protein Aggregate Clearance by the Proteasome,” *Cell* [Preprint]. doi:10.1016/j.cell.2016.07.001.

Hoeller, D. *et al.* (2006) “Regulation of ubiquitin-binding proteins by monoubiquitination,” *Nature cell biology*, 8(2), pp. 163–169. doi:10.1038/NCB1354.

Homola, J. (2003) “Present and future of surface plasmon resonance biosensors,” *Analytical and bioanalytical chemistry*, 377(3), pp. 528–539. doi:10.1007/S00216-003-2101-0.

Hospenthal, M.K., Freund, S.M.V. and Komander, D. (2013) “Assembly, analysis and architecture of atypical ubiquitin chains,” *Nature Structural & Molecular Biology 2013 20:5*, 20(5), pp. 555–565. doi:10.1038/nsmb.2547.

Hsu, S.T.D., Bertoncini, C.W. and Dobson, C.M. (2009) “Use of protonless NMR spectroscopy to alleviate the loss of information resulting from exchange-broadening,” *Journal of the American Chemical Society*, 131(21), pp. 7222–7223. doi:10.1021/JA902307Q/SUPPL_FILE/JA902307Q_SI_001.PDF.

Huang, H. *et al.* (2010) “K33-linked polyubiquitination of T cell receptor-ζ regulates proteolysis-independent T cell signaling,” *Immunity*, 33(1), pp. 60–70. doi:10.1016/J.IMMUNI.2010.07.002/ATTACHMENT/5BCFA5C2-1671-471B-BB1F-29B803011D3E/MMC1.PDF.

Huang, X., Shen, S. and Fan, D. (2017) “No evidence for pathogenic role of UBQLN2 mutations in sporadic amyotrophic lateral sclerosis in the mainland Chinese population,” *PLoS ONE*, 12(1). doi:10.1371/journal.pone.0170943.

Hudson, D.M. *et al.* (2012) “Peptidyl 3-hydroxyproline binding properties of type i collagen suggest a function in fibril supramolecular assembly,” *Biochemistry*, 51(12), pp. 2417–2424. doi:10.1021/BI2019139/ASSET/IMAGES/MEDIUM/BI-2011-019139_0006.GIF.

Hudson, D.M., Weis, M.A. and Eyre, D.R. (2011) “Insights on the Evolution of Prolyl 3-Hydroxylation Sites from Comparative Analysis of Chicken and Xenopus Fibrillar Collagens,” *PLOS ONE*, 6(5), p. e19336. doi:10.1371/JOURNAL.PONE.0019336.

Humtsoe, J.O. *et al.* (2005) “A Streptococcal Collagen-like Protein Interacts with the α2β1 Integrin and Induces Intracellular Signaling,” *Journal of Biological Chemistry*, 280(14), pp. 13848–13857. doi:10.1074/JBC.M410605200.

Hurley, J.H., Lee, S. and Prag, G. (2006) “Ubiquitin-binding domains,” *The Biochemical journal*, 399(3), pp. 361–372. doi:10.1042/BJ20061138.

Husnjak, K. *et al.* (2008) “Proteasome subunit Rpn13 is a novel ubiquitin receptor,” *Nature*, 453(7194), pp. 481–488. doi:10.1038/NATURE06926.

Husnjak, K. and Dikic, I. (2012) “Ubiquitin-Binding Proteins: Decoders of Ubiquitin-Mediated Cellular Functions,” *http://dx.doi.org/10.1146/annurev-biochem-051810-094654*, 81, pp. 291–322. doi:10.1146/ANNUREV-BIOCHEM-051810-094654.

Ichimura, Y. *et al.* (2013) “Phosphorylation of p62 activates the Keap1-Nrf2 pathway during selective autophagy,” *Molecular cell*, 51(5), pp. 618–631. doi:10.1016/J.MOLCEL.2013.08.003.

Ikeya, T. *et al.* (2009) “Automated NMR structure determination of stereo-array isotope labeled ubiquitin from minimal sets of spectra using the SAIL-FLYA system,” *Journal of biomolecular NMR*, 44(4), pp. 261–272. doi:10.1007/S10858-009-9339-6.

Improta, R., Benzi, C. and Barone, V. (2001) “Understanding the role of stereoelectronic effects in determining collagen stability. 1. A quantum mechanical study of proline, hydroxyproline, and fluoroproline dipeptide analogues in aqueous solution,” *Journal of the American Chemical Society*, 123(50), pp. 12568–12577. doi:10.1021/JA010599I/SUPPL_FILE/JA010599I_S.PDF.

Isogai, S. *et al.* (2011) “Crystal Structure of the Ubiquitin-associated (UBA) Domain of p62 and Its Interaction with Ubiquitin,” *The Journal of Biological Chemistry*, 286(36), p. 31864. doi:10.1074/JBC.M111.259630.

Iwata, J.I. *et al.* (2006) “Excess peroxisomes are degraded by autophagic machinery in mammals,” *The Journal of biological chemistry*, 281(7), pp. 4035–4041. doi:10.1074/JBC.M512283200.

Jachimowicz, R.D. and Reinhardt, H.C. (2019) “UBQLN4 promotes non-homologous end joining by repressing DNA end-resection,” *Molecular & cellular oncology*, 6(2). doi:10.1080/23723556.2019.1575692.

Jacobson, A.D. *et al.* (2009) “The lysine 48 and lysine 63 ubiquitin conjugates are processed differently by the 26 s proteasome,” *The Journal of biological chemistry*, 284(51), pp. 35485–35494. doi:10.1074/JBC.M109.052928.

Jantrapirom, S., Piccolo, L. lo and Yamaguchi, M. (2019) “Non-Proteasomal UbL-UbA Family of Proteins in Neurodegeneration,” *International journal of molecular sciences*, 20(8). doi:10.3390/IJMS20081893.

Jenkins, C.L. *et al.* (2003) “Effect of 3-hydroxyproline residues on collagen stability,” *Journal of the American Chemical Society*, 125(21), pp. 6422–6427. doi:10.1021/JA034015J/SUPPL_FILE/JA034015JSI20030106_014839.CIF.

Jha, A.K. *et al.* (2005) “Helix, sheet, and polyproline II frequencies and strong nearest neighbor effects in a restricted coil library,” *Biochemistry*, 44(28), pp. 9691–9702. doi:10.1021/BI0474822/SUPPL_FILE/BI0474822SI20050415_122743.PDF.

Jinek, M. *et al.* (2012) “A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity,” *Science*, 337(6096), pp. 816–821. doi:10.1126/SCIENCE.1225829/SUPPL_FILE/JINEK.SM.PDF.

Johnson, M Austin *et al.* (2007) “Development/Plasticity/Repair Functional Neural Development from Human Embryonic Stem Cells: Accelerated Synaptic Activity via Astrocyte Coculture.” doi:10.1523/JNEUROSCI.4562-06.2007.

Johnson, M. Austin *et al.* (2007a) “Functional neural development from human embryonic stem cells: accelerated synaptic activity via astrocyte coculture,” *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 27(12), pp. 3069–3077. doi:10.1523/JNEUROSCI.4562-06.2007.

Johnson, M. Austin *et al.* (2007b) “Functional neural development from human embryonic stem cells: accelerated synaptic activity via astrocyte coculture,” *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 27(12), pp. 3069–3077. doi:10.1523/JNEUROSCI.4562-06.2007.

Jumper, J. *et al.* (2021) “Highly accurate protein structure prediction with AlphaFold,” *Nature 2021 596:7873*, 596(7873), pp. 583–589. doi:10.1038/s41586-021-03819-2.

Kabashi, E. *et al.* (2008) “TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis,” *Nature genetics*, 40(5), pp. 572–574. doi:10.1038/NG.132.

Kageyama, S. *et al.* (2014) “Proteasome dysfunction activates autophagy and the Keap1-Nrf2 pathway,” *The Journal of biological chemistry*, 289(36), pp. 24944–24955. doi:10.1074/JBC.M114.580357.

Kaiser, S.E. *et al.* (2011) “Protein standard absolute quantification (PSAQ) method for the measurement of cellular ubiquitin pools,” *Nature Methods 2011 8:8*, 8(8), pp. 691–696. doi:10.1038/nmeth.1649.

Kajava, A. v. (2002) “What Curves α-Solenoids?: EVIDENCE FOR AN α-HELICAL TOROID STRUCTURE OF Rpn1 AND Rpn2 PROTEINS OF THE 26 S PROTEASOME *,” *Journal of Biological Chemistry*, 277(51), pp. 49791–49798. doi:10.1074/JBC.M204982200.

Kane, L.A. *et al.* (2014) “PINK1 phosphorylates ubiquitin to activate Parkin E3 ubiquitin ligase activity,” *The Journal of Cell Biology*, 205(2), p. 143. doi:10.1083/JCB.201402104.

Kang, H. and Schuman, E.M. (1996) “A requirement for local protein synthesis in neurotrophin-induced hippocampal synaptic plasticity,” *Science (New York, N.Y.)*, 273(5280), pp. 1402–1406. doi:10.1126/SCIENCE.273.5280.1402.

Kang, Y., Chen, X., *et al.* (2007) “Defining how ubiquitin receptors hHR23a and S5a bind polyubiquitin,” *Journal of molecular biology*, 369(1), pp. 168–176. doi:10.1016/J.JMB.2007.03.008.

Kang, Y., Zhang, N., *et al.* (2007) “Ubiquitin receptor proteins hHR23a and hPLIC2 interact,” *Journal of molecular biology*, 365(4), p. 1093. doi:10.1016/J.JMB.2006.10.056.

Kaplun, L. *et al.* (2005) “The DNA damage-inducible UbL-UbA protein Ddi1 participates in Mec1-mediated degradation of Ho endonuclease,” *Molecular and cellular biology*, 25(13), pp. 5355–5362. doi:10.1128/MCB.25.13.5355-5362.2005.

Karlström, Å. *et al.* (2004) “Identification of a novel collagen-like protein, SclC, in Streptococcus equi using signal sequence phage display,” *Veterinary Microbiology*, 104(3–4), pp. 179–188. doi:10.1016/J.VETMIC.2004.09.014.

Karlström, Å., Jacobsson, K. and Guss, B. (2006) “SclC is a member of a novel family of collagen-like proteins in Streptococcus equi subspecies equi that are recognised by antibodies against SclC,” *Veterinary Microbiology*, 114(1–2), pp. 72–81. doi:10.1016/J.VETMIC.2005.10.036.

Karplus, P.A. and Schulz, G.E. (1985) “Prediction of chain flexibility in proteins,” *Naturwissenschaften 1985 72:4*, 72(4), pp. 212–213. doi:10.1007/BF01195768.

Kaur, P.J. *et al.* (2015) “The Self-assembly of a Mini-fibril with Axial Periodicity from a Designed Collagen-mimetic Triple Helix *,” *Journal of Biological Chemistry*, 290(14), pp. 9251–9261. doi:10.1074/JBC.M113.542241.

Kay, B.K., Williamson, M.P. and Sudol, M. (2000) “The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains,” *FASEB*, (14), pp. 224–231. Available at: http://www.pharmacia.mod- (Accessed: March 14, 2022).

Kay, L.E., Xu, G.Y. and Yamazaki, T. (1994) “Enhanced-Sensitivity Triple-Resonance Spectroscopy with Minimal H2O Saturation,” *Journal of Magnetic Resonance, Series A*, 109(1), pp. 129–133. doi:10.1006/JMRA.1994.1145.

Kaye, F.J. *et al.* (2000) “A family of ubiquitin-like proteins binds the ATPase domain of Hsp70-like Stch,” *FEBS letters*, 467(2–3), pp. 348–355. doi:10.1016/S0014-5793(00)01135-2.

Kaye, F.J. and Shows, T.B. (2000) “Assignment of ubiquilin2 (UBQLN2) to human chromosome xp11. 23-->p11.1 by GeneBridge radiation hybrids,” *Cytogenetics and cell genetics*, 89(1–2), pp. 116–117. doi:10.1159/000015588.

Kazlauskaite, A. *et al.* (2014) “Parkin is activated by PINK1-dependent phosphorylation of ubiquitin at Ser65,” *Biochemical Journal*, 460(1), pp. 127–141. doi:10.1042/BJ20140334.

Kelly, M.A. *et al.* (2001) “Host−Guest Study of Left-Handed Polyproline II Helix Formation†,” *Biochemistry*, 40(48), pp. 14376–14383. doi:10.1021/BI011043A.

Khoshnoodi, J. *et al.* (2006) “Molecular Recognition in the Assembly of Collagens: Terminal Noncollagenous Domains Are Key Recognition Modules in the Formation of Triple Helical Protomers,” *Journal of Biological Chemistry*, 281(50), pp. 38117–38121. doi:10.1074/JBC.R600025200.

Kiel, J.A.K.W. (2010) “Autophagy in unicellular eukaryotes,” *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1541), p. 819. doi:10.1098/RSTB.2009.0237.

Kim, H.J., Lee, S. and Jung, J.U. (2010) “When autophagy meets viruses: a double-edged sword with functions in defense and offense,” *Seminars in immunopathology*, 32(4), pp. 323–341. doi:10.1007/S00281-010-0226-8.

Kim, H.T. and Goldberg, A.L. (2018) “UBL domain of Usp14 and other proteins stimulates proteasome activities and protein degradation in cells,” *Proceedings of the National Academy of Sciences of the United States of America*, 115(50), pp. E11642–E11650. doi:10.1073/PNAS.1808731115.

Kim, M. *et al.* (2006) “Surface plasmon resonance imaging analysis of protein-protein interactions using on-chip-expressed capture protein,” *Analytical biochemistry*, 351(2), pp. 298–304. doi:10.1016/J.AB.2006.01.042.

Kim, W. *et al.* (2011) “Systematic and quantitative assessment of the ubiquitin-modified proteome,” *Molecular cell*, 44(2), pp. 325–340. doi:10.1016/J.MOLCEL.2011.08.025.

Kimura, T. *et al.* (1989) “The Human ar2(XI) Collagen (COLI 1A2) Chain: MOLECULAR CLONING OF cDNA AND GENOMIC DNA REVEALS CHARACTERISTICS OF A FIBRILLAR COLLAGEN WITH DIFFERENCES IN GENOMIC ORGANIZATION*,” *THE JOURNAL OF BIOLOGICAL CHEMISTRY*, 264(23), pp. 13910–13916. doi:10.1016/S0021-9258(18)80086-2.

Kirkby, L.A. *et al.* (2013) “A role for correlated spontaneous activity in the assembly of neural circuits,” *Neuron*, 80(5), p. 1129. doi:10.1016/J.NEURON.2013.10.030.

Kirkin, V. *et al.* (2009) “A role for NBR1 in autophagosomal degradation of ubiquitinated substrates,” *Molecular cell*, 33(4), pp. 505–516. doi:10.1016/J.MOLCEL.2009.01.020.

Kishore, U. *et al.* (1998) “Functional characterization of a recombinant form of the C-terminal, globular head region of the B-chain of human serum complement protein, C1q.,” *Biochemical Journal*, 333(Pt 1), p. 27. doi:10.1042/BJ3330027.

Kisselev, A.F. *et al.* (1999) “The sizes of peptides generated from protein by mammalian 26 and 20 S proteasomes. Implications for understanding the degradative mechanism and antigen presentation,” *The Journal of biological chemistry*, 274(6), pp. 3363–3371. doi:10.1074/JBC.274.6.3363.

Kleijnen, M.F. *et al.* (2000) “The hPLIC proteins may provide a link between the ubiquitination machinery and the proteasome,” *Molecular cell*, 6(2), pp. 409–419. doi:10.1016/S1097-2765(00)00040-X.

Komander, D., Clague, M.J. and Urbé, S. (2009) “Breaking the chains: structure and function of the deubiquitinases,” *Nature reviews. Molecular cell biology*, 10(8), pp. 550–563. doi:10.1038/NRM2731.

Komatsu, M. *et al.* (2006) “Loss of autophagy in the central nervous system causes neurodegeneration in mice,” *Nature*, 441(7095), pp. 880–884. doi:10.1038/NATURE04723.

Korac, J. *et al.* (2013) “Ubiquitin-independent function of optineurin in autophagic clearance of protein aggregates,” *Journal of Cell Science*, 126(2), pp. 580–592. doi:10.1242/JCS.114926/263245/AM/UBIQUITIN-INDEPENDENT-FUNCTION-OF-OPTINEURIN-IN.

Koyano, F. *et al.* (2014) “Ubiquitin is phosphorylated by PINK1 to activate parkin,” *Nature 2014 510:7503*, 510(7503), pp. 162–166. doi:10.1038/nature13392.

Kozakov, D. *et al.* (2017) “The ClusPro web server for protein-protein docking,” *Nature protocols*, 12(2), p. 255. doi:10.1038/NPROT.2016.169.

Kramer, R.Z. *et al.* (1999) “Sequence dependent conformational variations of collagen triple-helical structure,” *Nature Structural Biology 1999 6:5*, 6(5), pp. 454–457. doi:10.1038/8259.

Kramer, R.Z. *et al.* (2000) “Staggered molecular packing in crystals of a collagen-like peptide with a single charged pair,” *Journal of Molecular Biology*, 301(5), pp. 1191–1205. doi:10.1006/JMBI.2000.4017.

Kristariyanto, Y.A., Choi, S.Y., *et al.* (2015) "Assembly and structure of Lys33-linked polyubiquitin reveals distinct conformations," *Biochemical Journal*, 467(2), pp. 345–352. doi:10.1042/BJ20141502.

Kristariyanto, Y.A., Abdul Rehman, S.A., *et al.* (2015) "K29-Selective Ubiquitin Binding Domain Reveals Structural Basis of Specificity and Heterotypic Nature of K29 Polyubiquitin," *Molecular Cell*, 58(1), pp. 83–94. doi:10.1016/J.MOLCEL.2015.01.041.

Kuma, A. and Mizushima, N. (2010) "Physiological role of autophagy as an intracellular recycling system: with an emphasis on nutrient metabolism," *Seminars in cell & developmental biology*, 21(7), pp. 683–690. doi:10.1016/J.SEMCDB.2010.03.002.

Kwon, Y.T. and Ciechanover, A. (2017) "The Ubiquitin Code in the Ubiquitin-Proteasome System and Autophagy," *Trends in Biochemical Sciences*, 42(11), pp. 873–886. doi:10.1016/J.TIBS.2017.09.002.

Lamark, T. and Johansen, T. (2012) "Aggrephagy: Selective disposal of protein aggregates by macroautophagy," *International Journal of Cell Biology* [Preprint]. doi:10.1155/2012/736905.

Lamb, C.A., Yoshimori, T. and Tooze, S.A. (2013) "The autophagosome: origins unknown, biogenesis complex," *Nature Reviews Molecular Cell Biology 2013 14:12*, 14(12), pp. 759–774. doi:10.1038/nrm3696.

Lancaster, A.K. *et al.* (2014) "PLAAC: a web and command-line application to identify proteins with prion-like amino acid composition," *Bioinformatics*, 30(17), p. 2501. doi:10.1093/BIOINFORMATICS/BTU310.

Le, N.T.T. *et al.* (2016) "Motor neuron disease, TDP-43 pathology, and memory deficits in mice expressing ALS-FTD-linked UBQLN2 mutations," *Proceedings of the National Academy of Sciences of the United States of America*, 113(47), pp. E7580–E7589. doi:10.1073/PNAS.1608432113.

Lee, M.J. *et al.* (2011) "Trimming of Ubiquitin Chains by Proteasome-associated Deubiquitinating Enzymes," *Molecular & Cellular Proteomics : MCP*, 10(5). doi:10.1074/MCP.R110.003871.

Letarov, A. v. *et al.* (1999) "The Carboxy-Terminal Domain Initiates Trimerization of Bacteriophage T4 Fibritin," *Biochemistry (Moscow)*, 64(7), pp. 817–823.

Lilienbaum, A. (2013) "Relationship between the proteasomal system and autophagy," *International Journal of Biochemistry and Molecular Biology*, 4(1), p. 1. Available at: /pmc/articles/PMC3627065/ (Accessed: May 11, 2022).

Lim, J. *et al.* (2015) "Proteotoxic Stress Induces Phosphorylation of p62/SQSTM1 by ULK1 to Regulate Selective Autophagic Clearance of Protein Aggregates," *PLOS Genetics*, 11(2), p. e1004987. doi:10.1371/JOURNAL.PGEN.1004987.

Lim, P.J. *et al.* (2009) "Ubiquilin and p97/VCP bind erasin, forming a complex involved in ERAD," *The Journal of cell biology*, 187(2), pp. 201–217. doi:10.1083/JCB.200903024.

Liu, C.C. *et al.* (2016) “Cul3-KLHL20 Ubiquitin Ligase Governs the Turnover of ULK1 and VPS34 Complexes to Control Autophagy Termination,” *Molecular cell*, 61(1), pp. 84–97. doi:10.1016/J.MOLCEL.2015.11.001.

Lopes, J.L.S. *et al.* (2014) “Distinct circular dichroism spectroscopic signatures of polyproline II and unordered secondary structures: Applications in secondary structure analyses,” *Protein Science : A Publication of the Protein Society*, 23(12), p. 1765. doi:10.1002/PRO.2558.

Lopez, J. *et al.* (2016) “Studying Intrinsically Disordered Proteins under True In Vivo Conditions by Combined Cross-Polarization and Carbonyl-Detection NMR Spectroscopy,” *Angewandte Chemie International Edition*, 55(26), pp. 7418–7422. doi:10.1002/ANIE.201601850.

Löwe, J. *et al.* (1995) “Crystal structure of the 20S proteasome from the archaeon T. acidophilum at 3.4 A resolution,” *Science (New York, N.Y.)*, 268(5210), pp. 533–539. doi:10.1126/SCIENCE.7725097.

Lu, K., den Brave, F. and Jentsch, S. (2017) “Receptor oligomerization guides pathway choice between proteasomal and autophagic degradation,” *Nature Cell Biology 2017 19:6*, 19(6), pp. 732–739. doi:10.1038/ncb3531.

Lu, K., Psakhye, I. and Jentsch, S. (2014) “Autophagic clearance of polyQ proteins mediated by ubiquitin-Atg8 adaptors of the conserved CUET protein family,” *Cell*, 158(3), pp. 549–563. doi:10.1016/J.CELL.2014.05.048.

Lukomski, S. *et al.* (2000) “Identification and characterization of the scl gene encoding a group A Streptococcus extracellular protein virulence factor with similarity to human collagen,” *Infection and Immunity*, 68(12), pp. 6542–6553. doi:10.1128/IAI.68.12.6542-6553.2000.

Lutz, C. (2018) “Mouse models of ALS: Past, present and future,” *Brain Research*, 1693, pp. 1–10. doi:10.1016/J.BRAINRES.2018.03.024.

Makarov, A.A. *et al.* (1975) “A Conformational study of beta-melanocyte-stimulation hormone,” *Biochemical and biophysical research communications*, 67(4), pp. 1378–1383. doi:10.1016/0006-291X(75)90179-5.

Makarov, A.A. *et al.* (1984) “Secondary structure of polypeptide hormones of the anterior lobe of the pituitary gland,” *Biopolymers*, 23(1), pp. 5–22. doi:10.1002/bip.360230103.

Makarov, A.A. *et al.* (1992) “Natural polypeptides in left-handed helical conformation. A circular dichroism study of the linker histones’ C-terminal fragments and beta-endorphin,” *FEBS letters*, 306(1), pp. 63–65. doi:10.1016/0014-5793(92)80838-8.

Mann, K. *et al.* (1996) “Glycosylated threonine but not 4-hydroxyproline dominates the triple helix stabilizing positions in the sequence of a hydrothermal vent worm cuticle collagen,” *Journal of molecular biology*, 261(2), pp. 255–266. doi:10.1006/JMBI.1996.0457.

Marín, I. (2014) “The ubiquilin gene family: evolutionary patterns and functional insights,” *BMC evolutionary biology*, 14(1). doi:10.1186/1471-2148-14-63.

Markin, C.J. and Spyracopoulos, L. (2012) “Increased precision for analysis of protein-ligand dissociation constants determined from chemical shift titrations,” *Journal of Biomolecular NMR*, 53(2), pp. 125–138. doi:10.1007/S10858-012-9630-9/FIGURES/12.

Marshall, R.S. *et al.* (2015) “Autophagic Degradation of the 26S Proteasome Is Mediated by the Dual ATG8/Ubiquitin Receptor RPN10 in Arabidopsis,” *Molecular cell*, 58(6), pp. 1053–1066. doi:10.1016/J.MOLCEL.2015.04.023.

Marshall, R.S., McLoughlin, F. and Vierstra, R.D. (2016) “Autophagic Turnover of Inactive 26S Proteasomes in Yeast Is Directed by the Ubiquitin Receptor Cue5 and the Hsp42 Chaperone,” *Cell reports*, 16(6), pp. 1717–1732. doi:10.1016/J.CELREP.2016.07.015.

Marshall, R.S. and Vierstra, R.D. (2015) “Eat or be eaten: The autophagic plight of inactive 26S proteasomes,” *Autophagy*, 11(10), pp. 1927–1928. doi:10.1080/15548627.2015.1078961.

Marshall, R.S. and Vierstra, R.D. (2019) “Dynamic regulation of the 26S proteasome: From synthesis to degradation,” *Frontiers in Molecular Biosciences*, 6(JUN), p. 40. doi:10.3389/FMOLB.2019.00040/BIBTEX.

Martin, E.W. and Mittag, T. (2018) “Relationship of Sequence and Phase Separation in Protein Low-Complexity Regions,” *Biochemistry*, 57(17), pp. 2478–2487. doi:10.1021/ACS.BIOCHEM.8B00008/ASSET/IMAGES/MEDIUM/BI-2018-00008N_0006.GIF.

Martinez-Fonts, K. *et al.* (2020) “The proteasome 19S cap and its ubiquitin receptors provide a versatile recognition platform for substrates,” *Nature Communications 2020 11:1*, 11(1), pp. 1–16. doi:10.1038/s41467-019-13906-8.

Mathieson, T. *et al.* (2018) “Systematic analysis of protein turnover in primary cells,” *Nature Communications 2018 9:1*, 9(1), pp. 1–10. doi:10.1038/s41467-018-03106-1.

Matsumoto, G. *et al.* (2011) “Serine 403 phosphorylation of p62/SQSTM1 regulates selective autophagic clearance of ubiquitinated proteins,” *Molecular cell*, 44(2), pp. 279–289. doi:10.1016/J.MOLCEL.2011.07.039.

Matsumoto, M.L. *et al.* (2010) “K11-Linked Polyubiquitination in Cell Cycle Control Revealed by a K11 Linkage-Specific Antibody,” *Molecular Cell*, 39(3), pp. 477–484. doi:10.1016/J.MOLCEL.2010.07.001.

Maury, Y. *et al.* (2014) “Combinatorial analysis of developmental cues efficiently converts human pluripotent stem cells into multiple neuronal subtypes,” *Nature Biotechnology*, 33. doi:10.1038/nbt.3049.

McEwan, D.G. and Dikic, I. (2011) “The Three Musketeers of Autophagy: phosphorylation, ubiquitylation and acetylation,” *Trends in cell biology*, 21(4), pp. 195–201. doi:10.1016/J.TCB.2010.12.006.

McKinnon, C. and Tabrizi, S.J. (2014) “The ubiquitin-proteasome system in neurodegeneration,” *Antioxidants & redox signaling*, 21(17), pp. 2302–2321. doi:10.1089/ARS.2013.5802.

Meier, S. *et al.* (2004) “Foldon, the natural trimerization domain of T4 fibritin, dissociates into a monomeric A-state form containing a stable beta-hairpin: atomic details of trimer dissociation and local beta-hairpin stability from residual dipolar couplings,” *Journal of molecular biology*, 344(4), pp. 1051–1069. doi:10.1016/J.JMB.2004.09.079.

Melendez, M.E. *et al.* (2018) “Construction and characterization of a new TRAIL soluble form, active at picomolar concentrations,” *Oncotarget*, 9(43), p. 27233. doi:10.18632/ONCOTARGET.25519.

Mendez, E.F. and Sattler, R. (2015) “Biomarker development for C9orf72 repeat expansion in ALS,” *Brain research*, 1607, pp. 26–35. doi:10.1016/J.BRAINRES.2014.09.041.

Meyer, H., Bug, M. and Bremer, S. (2012) “Emerging functions of the VCP/p97 AAA-ATPase in the ubiquitin system,” *Nature Cell Biology 2012 14:2*, 14(2), pp. 117–123. doi:10.1038/ncb2407.

Meyer, H.J. and Rape, M. (2014) “Enhanced protein degradation by branched ubiquitin chains,” *Cell*, 157(4), p. 910. doi:10.1016/J.CELL.2014.03.037.

Michel, M.A. *et al.* (2015) “Assembly and Specific Recognition of K29- and K33-Linked Polyubiquitin,” *Molecular Cell*, 58(1), pp. 95–109. doi:10.1016/J.MOLCEL.2015.01.042.

Millecamps, S. *et al.* (2012) “Mutations in UBQLN2 are rare in French amyotrophic lateral sclerosis,” *Neurobiology of Aging*, 33(4), pp. 839.e1-839.e3. doi:10.1016/J.NEUROBIOLAGING.2011.11.010.

Mizuno, K. *et al.* (2004) “Hydroxylation-induced Stabilization of the Collagen Triple Helix: ACETYL-(GLYCYL-4(R)-HYDROXYPROLYL-4(R)-HYDROXYPROLYL)10-NH2 FORMS A HIGHLY STABLE TRIPLE HELIX *,” *Journal of Biological Chemistry*, 279(36), pp. 38072–38078. doi:10.1074/JBC.M402953200.

Mizuno, K. *et al.* (2008) “Effect of the -Gly-3(S)-hydroxyprolyl-4(R)-hydroxyprolyl- tripeptide unit on the stability of collagen model peptides,” *The FEBS journal*, 275(23), pp. 5830–5840. doi:10.1111/J.1742-4658.2008.06704.X.

Mizuno, K., Hayashi, T. and Bächinger, H.P. (2003) “Hydroxylation-induced Stabilization of the Collagen Triple Helix: FURTHER CHARACTERIZATION OF PEPTIDES WITH 4(R)-HYDROXYPROLINE IN THE Xaa POSITION *,” *Journal of Biological Chemistry*, 278(34), pp. 32373–32379. doi:10.1074/JBC.M304741200.

Mizushima, N. and Komatsu, M. (2011) “Autophagy: renovation of cells and tissues,” *Cell*, 147(4), pp. 728–741. doi:10.1016/J.CELL.2011.10.026.

Mohs, A. *et al.* (2007) “Mechanism of Stabilization of a Bacterial Collagen Triple Helix in the Absence of Hydroxyproline,” *Journal of Biological Chemistry*, 282(41), pp. 29757–29765. doi:10.1074/JBC.M703991200.

Mori, S. *et al.* (1995) “Improved sensitivity of HSQC spectra of exchanging protons at short interscan delays using a new fast HSQC (FHSQC) detection scheme that avoids water saturation,” *Journal of magnetic resonance. Series B*, 108(1), pp. 94–98. doi:10.1006/JMRB.1995.1109.

Morris, J.R. and Solomon, E. (2004) “BRCA1 : BARD1 induces the formation of conjugated ubiquitin structures, dependent on K6 of ubiquitin, in cells during DNA replication and repair,” *Human Molecular Genetics*, 13(8), pp. 807–817. doi:10.1093/HMG/DDH095.

Muhandiram, D.R. and Kay, L.E. (1994) “Gradient-Enhanced Triple-Resonance Three-Dimensional NMR Experiments with Improved Sensitivity,” *Journal of Magnetic Resonance, Series B*, 103(3), pp. 203–216. doi:10.1006/JMRB.1994.1032.

Mukhopadhyay, D. and Riezman, H. (2007) “Proteasome-independent functions of ubiquitin in endocytosis and signaling,” *Science (New York, N.Y.)*, 315(5809), pp. 201–205. doi:10.1126/SCIENCE.1127085.

Murray, A.W. (2004) “Recycling the cell cycle: cyclins revisited,” *Cell*, 116(2), pp. 221–234. doi:10.1016/S0092-8674(03)01080-8.

Nahar, A. *et al.* (2019) “Two alternative mechanisms regulate the onset of chaperone-mediated assembly of the proteasomal ATPases,” *The Journal of Biological Chemistry*, 294(16), p. 6562. doi:10.1074/JBC.RA118.006298.

Nandi, D. *et al.* (2006) “The ubiquitin-proteasome system,” *Journal of biosciences*, 31(1), pp. 137–155. doi:10.1007/BF02705243.

Narendra, D. *et al.* (2008) “Parkin is recruited selectively to impaired mitochondria and promotes their autophagy,” *Journal of Cell Biology*, 183(5), pp. 795–803. doi:10.1083/JCB.200809125/VIDEO-1.

N’Diaye, E.N. *et al.* (2009) “PLIC proteins or ubiquilins regulate autophagy-dependent cell survival during nutrient starvation,” *EMBO reports*, 10(2), pp. 173–179. doi:10.1038/EMBOR.2008.238.

Neal, S. *et al.* (2006) “Accurate prediction of protein torsion angles using chemical shifts and sequence homology,” *Magnetic Resonance in Chemistry*, 44(S1), pp. S158–S167. doi:10.1002/MRC.1832.

Nelson, B.P. *et al.* (2001) “Surface plasmon resonance imaging measurements of DNA and RNA hybridization adsorption onto DNA microarrays,” *Analytical chemistry*, 73(1), pp. 1–7. doi:10.1021/AC0010431.

Nemec, A.A. *et al.* (2017) “Autophagic clearance of proteasomes in yeast requires the conserved sorting nexin Snx4,” *The Journal of biological chemistry*, 292(52), pp. 21466–21480. doi:10.1074/JBC.M117.817999.

Nguyen, K., Puthenveetil, R. and Vinogradova, O. (2017) “Investigation of the adaptor protein PLIC-2 in multiple pathways,” *Biochemistry and Biophysics Reports*, 9, pp. 341–348. doi:10.1016/J.BBREP.2017.01.013.

Nillegoda, N.B. *et al.* (2015) “Crucial HSP70 co-chaperone complex unlocks metazoan protein disaggregation,” *Nature*, 524(7564), pp. 247–251. doi:10.1038/NATURE14884.

Nishikawa, H. *et al.* (2004) “Mass Spectrometric and Mutational Analyses Reveal Lys-6-linked Polyubiquitin Chains Catalyzed by BRCA1-BARD1 Ubiquitin Ligase *,”

Journal of Biological Chemistry, 279(6), pp. 3916–3924. doi:10.1074/JBC.M308540200.

Nowicka, U. *et al.* (2015) "DNA-damage-inducible 1 protein (Ddi1) contains an uncharacteristic ubiquitin-like domain that binds ubiquitin," *Structure (London, England : 1993)*, 23(3), pp. 542–557. doi:10.1016/J.STR.2015.01.010.

Ogden, C.A. *et al.* (2001) "C1q and mannose binding lectin engagement of cell surface calreticulin and CD91 initiates macropinocytosis and uptake of apoptotic cells," *The Journal of experimental medicine*, 194(6), pp. 781–795. doi:10.1084/JEM.194.6.781.

Ohtake, F. *et al.* (2015) "Ubiquitin acetylation inhibits polyubiquitin chain elongation," *EMBO Reports*, 16(2), p. 192. doi:10.15252/EMBR.201439152.

Ordureau, A. *et al.* (2014a) "Quantitative proteomics reveal a feedforward mechanism for mitochondrial PARKIN translocation and ubiquitin chain synthesis," *Molecular Cell*, 56(3), pp. 360–375. doi:10.1016/J.MOLCEL.2014.09.007/ATTACHMENT/DBE9C9E7-C36A-4670-B04D-B54B07C9E8D9/MMC4.MP4.

Ordureau, A. *et al.* (2014b) "Quantitative proteomics reveal a feedforward mechanism for mitochondrial PARKIN translocation and ubiquitin chain synthesis," *Molecular Cell*, 56(3), pp. 360–375. doi:10.1016/J.MOLCEL.2014.09.007/ATTACHMENT/DBE9C9E7-C36A-4670-B04D-B54B07C9E8D9/MMC4.MP4.

Pankiv, S. *et al.* (2007) "p62/SQSTM1 binds directly to Atg8/LC3 to facilitate degradation of ubiquitinated protein aggregates by autophagy," *The Journal of biological chemistry*, 282(33), pp. 24131–24145. doi:10.1074/JBC.M702824200.

Papanikolopoulou, K. *et al.* (2004) "Formation of highly stable chimeric trimers by fusion of an adenovirus fiber shaft fragment with the foldon domain of bacteriophage t4 fibritin," *The Journal of biological chemistry*, 279(10), pp. 8991–8998. doi:10.1074/JBC.M311791200.

Papanikolopoulou, K., van Raaij, M.J. and Mitraki, A. (2008) "Creation of hybrid nanorods from sequences of natural trimeric fibrous proteins using the fibritin trimerization motif," *Methods in Molecular Biology*, 474, pp. 15–33. doi:10.1007/978-1-59745-480-3_2.

Parzych, K.R. and Klionsky, D.J. (2014) "An Overview of Autophagy: Morphology, Mechanism, and Regulation," *Antioxidants & Redox Signaling*, 20(3), p. 460. doi:10.1089/ARS.2013.5371.

Paterson, G.K. *et al.* (2008) "PclA, a pneumococcal collagen-like protein with selected strain distribution, contributes to adherence and invasion of host cells," *FEMS Microbiology Letters*, 285(2), pp. 170–176. doi:10.1111/J.1574-6968.2008.01217.X.

Peng, J. *et al.* (2003) "A proteomics approach to understanding protein ubiquitination," *Nature Biotechnology 2003 21:8*, 21(8), pp. 921–926. doi:10.1038/nbt849.

Pérez Berrocal, D.A. *et al.* (2020) “Hybrid Chains: A Collaboration of Ubiquitin and Ubiquitin-Like Modifiers Introducing Cross-Functionality to the Ubiquitin Code,” *Frontiers in Chemistry*, 7, p. 931. doi:10.3389/FCHEM.2019.00931/BIBTEX.

Perozzi, G. and Prakash, S. (1986) “RAD7 gene of Saccharomyces cerevisiae: transcripts, nucleotide sequence analysis, and functional relationship between the RAD7 and RAD23 gene products.,” *Molecular and Cellular Biology*, 6(5), p. 1497. doi:10.1128/MCB.6.5.1497.

Perret, S. *et al.* (2001) “Unhydroxylated Triple Helical Collagen I Produced in Transgenic Plants Provides New Clues on the Role of Hydroxyproline in Collagen Folding and Fibril Formation,” *Journal of Biological Chemistry*, 276(47), pp. 43693–43698. doi:10.1074/JBC.M105507200.

Persikov, A. v., Ramshaw, J.A.M. and Brodsky, B. (2005) “Prediction of Collagen Stability from Amino Acid Sequence,” *Journal of Biological Chemistry*, 280(19), pp. 19343–19349. doi:10.1074/JBC.M501657200.

Peters, L.Z. *et al.* (2015) “The Protein Quality Control Machinery Regulates Its Misassembled Proteasome Subunits,” *PLOS Genetics*, 11(4), p. e1005178. doi:10.1371/JOURNAL.PGEN.1005178.

Picher-Martel, V. *et al.* (2015) “Ubiquilin-2 drives NF-κB activity and cytosolic TDP-43 aggregation in neuronal cells,” *Molecular Brain*, 8(1), pp. 1–14. doi:10.1186/S13041-015-0162-6/FIGURES/5.

Pickart, C.M. (2001) “Mechanisms underlying ubiquitination,” *Annual review of biochemistry*, 70, pp. 503–533. doi:10.1146/ANNUREV.BIOCHEM.70.1.503.

Pizarro-Guajardo, M. *et al.* (2014) “Characterization of the collagen-like exosporium protein, BclA1, of Clostridium difficile spores,” *Anaerobe*, 25, pp. 18–30. doi:10.1016/J.ANAEROBE.2013.11.003.

Price, J.C. *et al.* (2010) “Analysis of proteome dynamics in the mouse brain,” *Proceedings of the National Academy of Sciences of the United States of America*, 107(32), pp. 14508–14513. doi:10.1073/PNAS.1006551107/SUPPL_FILE/ST02.DOC.

Prockop, D.J. (2004) “Collagens,” *Encyclopedia of Biological Chemistry*, pp. 482–487. doi:10.1016/B0-12-443710-9/00131-9.

Proctor, W.G. and Yu, F.C. (1950) “The Dependence of a Nuclear Magnetic Resonance Frequency upon Chemical Compound,” *Physical Review*, 77(5), p. 717. doi:10.1103/PhysRev.77.717.

Quiroz, F.G. *et al.* (2019) “Intrinsically disordered proteins access a range of hysteretic phase separation behaviors,” *Science Advances*, 5(10), pp. 5177–5195. doi:10.1126/SCIADV.AAX5177/SUPPL_FILE/AAX5177_SM.PDF.

Raasi, S. *et al.* (2004) “Binding of Polyubiquitin Chains to Ubiquitin-associated (UBA) Domains of HHR23A,” *Journal of Molecular Biology*, 341(5), pp. 1367–1379. doi:10.1016/J.JMB.2004.06.057.

Raasi, S. and Pickart, Cecile M (2003) “Rad23 Ubiquitin-associated Domains (UBA) Inhibit 26 S Proteasome-catalyzed Proteolysis by Sequestering Lysine 48-linked Polyubiquitin Chains*.” doi:10.1074/jbc.M212841200.

Raasi, S. and Pickart, Cecile M. (2003) “Rad23 Ubiquitin-associated Domains (UBA) Inhibit 26 S Proteasome-catalyzed Proteolysis by Sequestering Lysine 48-linked Polyubiquitin Chains *,” *Journal of Biological Chemistry*, 278(11), pp. 8951–8959. doi:10.1074/JBC.M212841200.

Ramachandran, G.N. and Kartha, G. (1954) “Structure of collagen,” *Nature*, 174(4423), pp. 269–270. doi:10.1038/174269C0.

Rangadurai, A., Shi, H. and Al-Hashimi, H.M. (2020) “Extending the Sensitivity of CEST NMR Spectroscopy to Micro-to-Millisecond Dynamics in Nucleic Acids Using High-Power Radio-Frequency Fields,” *Angewandte Chemie - International Edition*, 59(28), pp. 11262–11266. doi:10.1002/ANIE.202000493.

Rao, H. *et al.* (2001) “Degradation of a cohesin subunit by the N-end rule pathway is essential for chromosome stability,” *Nature*, 410(6831), pp. 955–959. doi:10.1038/35073627.

Rasmussen, M., Eden, A. and Bjorck, L. (2000) “SclA, a novel collagen-like surface protein of Streptococcus pyogenes,” *Infection and Immunity*, 68(11), pp. 6370–6377. doi:10.1128/IAI.68.11.6370-6377.2000.

Rasmussen, M., Jacobsson, M. and Björck, L. (2003) “Genome-based Identification and Analysis of Collagen-related Structural Motifs in Bacterial and Viral Proteins,” *Journal of Biological Chemistry*, 278(34), pp. 32313–32316. doi:10.1074/JBC.M304709200.

Rath, A., Davidson, A.R. and Deber, C.M. (2005) “The structure of ‘unstructured’ regions in peptides and proteins: role of the polyproline II helix in protein folding and recognition,” *Biopolymers*, 80(2–3), pp. 179–185. doi:10.1002/BIP.20227.

Ravits, J. *et al.* (2007) “Implications of ALS focality: rostral-caudal distribution of lower motor neuron loss postmortem,” *Neurology*, 68(19), pp. 1576–1582. doi:10.1212/01.WNL.0000261045.57095.56.

Reddy Chichili, V.P., Kumar, V. and Sivaraman, J. (2013) “Linkers in the structural biology of protein–protein interactions,” *Protein Science : A Publication of the Protein Society*, 22(2), p. 153. doi:10.1002/PRO.2206.

Reidick, C. *et al.* (2014) “Regulation of the Tumor-Suppressor Function of the Class III Phosphatidylinositol 3-Kinase Complex by Ubiquitin and SUMO,” *Cancers*, 7(1), pp. 01–29. doi:10.3390/CANCERS7010001.

Reimer, U. *et al.* (1998) “Side-chain effects on peptidyl-prolyl cis/trans isomerisation,” *Journal of Molecular Biology*, 279(2), pp. 449–460. doi:10.1006/JMBI.1998.1770.

Renaud, L. *et al.* (2019) “Key role of UBQLN2 in pathogenesis of amyotrophic lateral sclerosis and frontotemporal dementia,” *Acta Neuropathologica Communications*, 7(1), pp. 1–11. doi:10.1186/S40478-019-0758-7/TABLES/2.

Reuter, M. *et al*. (2010) “Binding of the Human Complement Regulators CFHR1 and Factor H by Streptococcal Collagen-like Protein 1 (Scl1) via Their Conserved C Termini Allows Control of the Complement Cascade at Multiple Levels,” *Journal of Biological Chemistry*, 285(49), pp. 38473–38485. doi:10.1074/JBC.M110.143727.

Ricard-Blum, S. (2011) “The Collagen Family,” *Cold Spring Harbor Perspectives in Biology*, 3(1), p. a004978. doi:10.1101/CSHPERSPECT.A004978.

Roberts, S., Dzuricky, M. and Chilkoti, A. (2015) “Elastin-like Polypeptides as Models of Intrinsically Disordered Proteins,” *FEBS letters*, 589(19 0 0), p. 2477. doi:10.1016/J.FEBSLET.2015.08.029.

Rosen, D.R. *et al*. (1993) “Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis,” *Nature 1993 362:6415*, 362(6415), pp. 59–62. doi:10.1038/362059a0.

Rosenberg, T. *et al*. (2014) “The roles of protein expression in synaptic plasticity and memory consolidation,” *Frontiers in Molecular Neuroscience*, 7(November), p. 86. doi:10.3389/FNMOL.2014.00086/BIBTEX.

Rosenbloom, J., Harsch, M. and Jimenez, S. (1973) “Hydroxyproline content determines the denaturation temperature of chick tendon collagen,” *Archives of biochemistry and biophysics*, 158(2), pp. 478–484. doi:10.1016/0003-9861(73)90539-0.

Rothenberg, C. *et al*. (2010) “Ubiquilin functions in autophagy and is degraded by chaperone-mediated autophagy,” *Human molecular genetics*, 19(16), pp. 3219–3232. doi:10.1093/HMG/DDQ231.

Rucker, A.L. and Creamer, T.P. (2002) “Polyproline II helical structure in protein unfolded states: Lysine peptides revisited,” *Protein Science*, 11(4), pp. 980–985. doi:10.1110/PS.4550102.

Sakakibara, S. *et al*. (1973) “Synthesis of (Pro-Hyp-Gly) n of defined molecular weights. Evidence for the stabilization of collagen triple helix by hydroxypyroline,” *Biochimica et biophysica acta*, 303(1), pp. 198–202. doi:10.1016/0005-2795(73)90164-5.

Sandoe, J. and Eggan, K. (2013) “Opportunities and challenges of pluripotent stem cell neurodegenerative disease models,” *Nature neuroscience*, 16(7), pp. 780–789. doi:10.1038/NN.3425.

Sasaki, T. *et al*. (2005) “Budding yeast Dsk2 protein forms a homodimer via its C-terminal UBA domain,” *Biochemical and biophysical research communications*, 336(2), pp. 530–535. doi:10.1016/J.BBRC.2005.08.126.

Schanzenbächer, C.T. *et al*. (2016) “Nascent Proteome Remodeling following Homeostatic Scaling at Hippocampal Synapses,” *Neuron*, 92(2), pp. 358–371. doi:10.1016/J.NEURON.2016.09.058.

Schiavina, M. *et al*. (2019) “Taking Simultaneous Snapshots of Intrinsically Disordered Proteins in Action,” *Biophysical Journal*, 117(1), pp. 46–55.

doi:10.1016/J.BPJ.2019.05.017/ATTACHMENT/84D741D1-252E-42C6-B5B9-75E4E04DD6CF/MMC1.PDF.

Schleucher, J., Sattler, M. and Griesinger, C. (1993) “Coherence Selection by Gradients without Signal Attenuation: Application to the Three-Dimensional HNCO Experiment,” *Angewandte Chemie International Edition in English*, 32(10), pp. 1489–1491. doi:10.1002/ANIE.199314891.

Schuck, P. (2000) “Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling,” *Biophysical journal*, 78(3), pp. 1606–1619. doi:10.1016/S0006-3495(00)76713-0.

Schumacher, M.A., Mizuno, K. and Bächinger, H.P. (2006) “The crystal structure of a collagen-like polypeptide with 3(S)-hydroxyproline residues in the Xaa position forms a standard 7/2 collagen triple helix,” *The Journal of biological chemistry*, 281(37), pp. 27566–27574. doi:10.1074/JBC.M602797200.

Schweitzer-Stenner, R. (2011) “Conformational propensities and residual structures in unfolded peptides and proteins,” *Molecular BioSystems*, 8(1), pp. 122–133. doi:10.1039/C1MB05225J.

Seibenhener, M.L. *et al.* (2004a) “Sequestosome 1/p62 Is a Polyubiquitin Chain Binding Protein Involved in Ubiquitin Proteasome Degradation,” *Molecular and Cellular Biology*, 24(18), p. 8055. doi:10.1128/MCB.24.18.8055-8068.2004.

Seibenhener, M.L. *et al.* (2004b) “Sequestosome 1/p62 Is a Polyubiquitin Chain Binding Protein Involved in Ubiquitin Proteasome Degradation,” *Molecular and Cellular Biology*, 24(18), p. 8055. doi:10.1128/MCB.24.18.8055-8068.2004.

Selvaraj, B.T., Livesey, M.R. and Chandran, S. (2017) “Modeling the C9ORF72 repeat expansion mutation using human induced pluripotent stem cells,” *Brain Pathology*, 27(4), pp. 518–524. doi:10.1111/bpa.12520.

Semple, C.A.M. *et al.* (2003) “The Comparative Proteomics of Ubiquitination in Mouse,” *Genome Research*, 13(6b), p. 1389. doi:10.1101/GR.980303.

Şentürk, M. *et al.* (2019) “Ubiquilins regulate autophagic flux through mTOR signalling and lysosomal acidification,” *Nature Cell Biology*, 21(3), pp. 384–396. doi:10.1038/s41556-019-0281-x.

Şentürk, M., Mao, D. and Bellen, H.J. (2019) “Loss of proteins associated with amyotrophic lateral sclerosis affects lysosomal acidification via different routes,” *Autophagy*, 15(8), pp. 1467–1469. doi:10.1080/15548627.2019.1609863.

Sharkey, L.M. *et al.* (2018) “Mutant UBQLN2 promotes toxicity by modulating intrinsic self-assembly,” *Proceedings of the National Academy of Sciences of the United States of America*, 115(44), pp. E10495–E10504. doi:10.1073/pnas.1810522115.

Shen, X. *et al.* (2011) “Processing of Optineurin in Neuronal Cells,” *The Journal of Biological Chemistry*, 286(5), p. 3618. doi:10.1074/JBC.M110.175810.

Shi, Yuan *et al.* (2016) “Rpn1 provides adjacent receptor sites for substrate binding and deubiquitination by the proteasome,” *Science (New York, N.Y.)*, 351(6275). doi:10.1126/SCIENCE.AAD9421.

Shoulders, M.D. and Raines, R.T. (2009) “COLLAGEN STRUCTURE AND STABILITY,” *Annual review of biochemistry*, 78, p. 929. doi:10.1146/ANNUREV.BIOCHEM.77.032207.120833.

Si, Y. *et al.* (2021) “FGF23, a novel muscle biomarker detected in the early stages of ALS,” *Scientific reports*, 11(1). doi:10.1038/S41598-021-91496-6.

Siligardi, G. and Drake, A.F. (1995a) “The importance of extended conformations and, in particular, the PII conformation for the molecular recognition of peptides,” *Biopolymers*, 37(4), pp. 281–292. doi:10.1002/BIP.360370406.

Siligardi, G. and Drake, A.F. (1995b) “The importance of extended conformations and, in particular, the PII conformation for the molecular recognition of peptides,” *Biopolymers*, 37(4), pp. 281–292. doi:10.1002/BIP.360370406.

Snellman, A. *et al.* (2000) “A short sequence in the N-terminal region is required for the trimerization of type XIII collagen and is conserved in other collagenous transmembrane proteins,” *The EMBO journal*, 19(19), pp. 5051–5059. doi:10.1093/EMBOJ/19.19.5051.

Spence, J. *et al.* (2000) “Cell cycle-regulated modification of the ribosome by a variant multiubiquitin chain,” *Cell*, 102(1), pp. 67–76. doi:10.1016/S0092-8674(00)00011-8.

Sreerama, N., Venyaminov, S.Y.U. and Woody, R.W. (1999) “Estimation of the number of alpha-helical and beta-strand segments in proteins using circular dichroism spectroscopy.,” *Protein Science : A Publication of the Protein Society*, 8(2), p. 370. doi:10.1110/PS.8.2.370.

Sreerama, N. and Woody, R.W. (2000) “Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set,” *Analytical biochemistry*, 287(2), pp. 252–260. doi:10.1006/ABIO.2000.4880.

Stapley, B.J. and Creamer, T.P. (1999) “A survey of left-handed polyproline II helices.,” *Protein Science : A Publication of the Protein Society*, 8(3), p. 587. doi:10.1110/PS.8.3.587.

van Stokkum, I.H.M. *et al.* (1990) “Estimation of protein secondary structure and error analysis from circular dichroism spectra,” *Analytical biochemistry*, 191(1), pp. 110–118. doi:10.1016/0003-2697(90)90396-Q.

Stolz, A. *et al.* (2011) “Cdc48: a power machine in protein degradation,” *Trends in biochemical sciences*, 36(10), pp. 515–523. doi:10.1016/J.TIBS.2011.06.001.

Sutton, M.A. and Schuman, E.M. (2006) “Dendritic protein synthesis, synaptic plasticity, and memory,” *Cell*, 127(1), pp. 49–58. doi:10.1016/J.CELL.2006.09.014.

Sylvestre, P., Couture-Tosi, E. and Mock, M. (2003) “Polymorphism in the collagen-like region of the Bacillus anthracis BclA protein leads to variation in exosporium filament length,” *Journal of Bacteriology*, 185(5), pp. 1555–1563. doi:10.1128/JB.185.5.1555-1563.2003.

Takahashi, K. *et al.* (2007) “Induction of pluripotent stem cells from adult human fibroblasts by defined factors,” *Cell*, 131(5), pp. 861–872. doi:10.1016/J.CELL.2007.11.019.

Takazawa, T. *et al.* (2012) “Maturation of spinal motor neurons derived from human embryonic stem cells,” *PloS one*, 7(7). doi:10.1371/JOURNAL.PONE.0040154.

Tang, X. *et al.* (2021) “UBQLN4 is activated by C/EBPβ and exerts oncogenic effects on colorectal cancer via the Wnt/β-catenin signaling pathway,” *Cell Death Discovery 2021 7:1*, 7(1), pp. 1–13. doi:10.1038/s41420-021-00795-4.

Tao, Y. *et al.* (1997) “Structure of bacteriophage T4 fibritin: a segmented coiled coil and the role of the C-terminal domain,” *Structure (London, England : 1993)*, 5(6), pp. 789–798. doi:10.1016/S0969-2126(97)00233-5.

Tasab, M., Jenkinson, L. and Bulleid, N.J. (2002) “Sequence-specific Recognition of Collagen Triple Helices by the Collagen-specific Molecular Chaperone HSP47,” *Journal of Biological Chemistry*, 277(38), pp. 35007–35012. doi:10.1074/JBC.M202782200.

Taylor, J.P., Brown, R.H. and Cleveland, D.W. (2016) “Decoding ALS: From Genes to Mechanism,” *Nature*, 539(7628), p. 197. doi:10.1038/NATURE20413.

Terzaghi, B.E., Terzaghi, E. and Coombs, D. (1979) “The role of the collar/whisker complex in bacteriophage T4D tail fiber attachment,” *Journal of molecular biology*, 127(1), pp. 1–14. doi:10.1016/0022-2836(79)90454-6.

Teyssou, E. *et al.* (2017) “Novel UBQLN2 mutations linked to amyotrophic lateral sclerosis and atypical hereditary spastic paraplegia phenotype through defective HSP70-mediated proteolysis,” *Neurobiology of Aging*, 58, pp. 239.e11-239.e20. doi:10.1016/J.NEUROBIOLAGING.2017.06.018.

Theillet, F.-X. *et al.* (2013) “The alphabet of intrinsic disorder,” *http://dx.doi.org/10.4161/idp.24360*, 1(1), p. e24360. doi:10.4161/IDP.24360.

Thrower, J.S. *et al.* (2000) “Recognition of the polyubiquitin proteolytic signal,” *The EMBO journal*, 19(1), pp. 94–102. doi:10.1093/EMBOJ/19.1.94.

Tifany, M.L. and Krimm, S. (1972) “Effect of temperature on the circular dichroism spectra of polypeptides in the extended state,” *Biopolymers*, 11(11), pp. 2309–2316. doi:10.1002/BIP.1972.360111109.

Tiffany, M.L. and Krimm, S. (1968) “New chain conformations of poly(glutamic acid) and polylysine,” *Biopolymers*, 6(9), pp. 1379–1382. doi:10.1002/BIP.1968.360060911.

Tiffany, M.L. and Krimm, S. (1969) “Circular dichroism of the ‘random’ polypeptide chain,” *Biopolymers*, 8(3), pp. 347–359. doi:10.1002/BIP.1969.360080306.

Traub, W. and Piez, K.A. (1971) “The chemistry and structure of collagen,” *Advances in protein chemistry*, 25(C), pp. 243–352. doi:10.1016/S0065-3233(08)60281-8.

Trempe, J.F. *et al.* (2005) “Mechanism of Lys48-linked polyubiquitin chain recognition by the Mud1 UBA domain,” *The EMBO journal*, 24(18), pp. 3178–3189. doi:10.1038/SJ.EMBOJ.7600797.

Tryggvason, K., Risteli, J. and Kivirikko, K.I. (1976) “Separation of prolyl 3-hydroxylase and 4-hydroxylase activities and the 4-hydroxyproline requirement for synthesis of 3-hydroxyproline,” *Biochemical and biophysical research communications*, 76(2), pp. 275–281. doi:10.1016/0006-291X(77)90722-7.

Tsuchiya, H. *et al.* (2017) “In Vivo Ubiquitin Linkage-type Analysis Reveals that the Cdc48-Rad23/Dsk2 Axis Contributes to K48-Linked Chain Specificity of the Proteasome,” *Molecular cell*, 66(4), pp. 488-502.e7. doi:10.1016/J.MOLCEL.2017.04.024.

Tsuchiya, H., Endo, A. and Saeki, Y. (2020) “Multi-step ubiquitin decoding mechanism for proteasomal degradation,” *Pharmaceuticals*. MDPI AG, pp. 1–14. doi:10.3390/ph13060128.

Tsuji, I., Iwamoto, K. and Shintani, Y. (2014) “Effective Expression and Purification of Bioactive Recombinant Soluble LIGHT,” *Methods in Molecular Biology*, 1155, pp. 201–213. doi:10.1007/978-1-4939-0669-7_17.

Vance, C. *et al.* (2009) “Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6,” *Science (New York, N.Y.)*, 323(5918), pp. 1208–1211. doi:10.1126/SCIENCE.1165942.

Vandersmissen, L. *et al.* (2010) “A Legionella pneumophila collagen-like protein encoded by a gene with a variable number of tandem repeats is involved in the adherence and invasion of host cells,” *FEMS Microbiology Letters*, 306(2), pp. 168–176. doi:10.1111/J.1574-6968.2010.01951.X.

Varshavsky, A. (2005) “Regulated protein degradation,” *Trends in biochemical sciences*, 30(6), pp. 283–286. doi:10.1016/J.TIBS.2005.04.005.

Vaz, I.M. *et al.* (2021) “Chromosomal aberrations after induced pluripotent stem cells reprogramming,” *Genetics and Molecular Biology*, 44(3). doi:10.1590/1678-4685-GMB-2020-0147.

Vengoechea, J. *et al.* (2013) “Clinical variability and female penetrance in X-linked familial FTD/ALS caused by a P506S mutation in UBQLN2,” *Amyotrophic lateral sclerosis & frontotemporal degeneration*, 14(7–8), pp. 615–619. doi:10.3109/21678421.2013.824001.

Verma, R. *et al.* (2004) “Multiubiquitin Chain Receptors Define a Layer of Substrate Selectivity in the Ubiquitin-Proteasome System,” *Cell*, 118(1), pp. 99–110. doi:10.1016/J.CELL.2004.06.014.

Vitagliano, L. *et al.* (2001) “Structural Bases of Collagen Stabilization Induced by Proline Hydroxylation,” 58, pp. 459–464. doi:10.1002/1097-0282.

Vlach, J., Hennecke, S. and Amati, B. (1997) “Phosphorylation-dependent degradation of the cyclin-dependent kinase inhibitor p27,” *The EMBO journal*, 16(17), pp. 5334–5344. doi:10.1093/EMBOJ/16.17.5334.

Vogel, P., Dux, E. and Wiessner, C. (1997) *Effect of Heat Shock on Neuronal Cultures: Importance of Protein Synthesis and HSP72 Induction for Induced Tolerance and Survival*, *Metabolic Brain Disease*. Available at: https://link.springer.com/content/pdf/10.1007%2FBF02674613.pdf (Accessed: May 23, 2019).

Voges, D., Zwickl, P. and Baumeister, W. (1999) “The 26S proteasome: a molecular machine designed for controlled proteolysis,” *Annual review of biochemistry*, 68, pp. 1015–1068. doi:10.1146/ANNUREV.BIOCHEM.68.1.1015.

Vranken, W.F. *et al.* (2005) “The CCPN data model for NMR spectroscopy: Development of a software pipeline,” *Proteins: Structure, Function, and Bioinformatics*, 59(4), pp. 687–696. doi:10.1002/PROT.20449.

Wallace, BA, & Janes, R. (2010) “Modern Techniques for Circular Dichroism and Synchrotron Radiation Circular Dichroism Spectroscopy,” *IOS Press*, pp. 7–8.

Walters, K.J. *et al.* (2002) “Structural studies of the interaction between ubiquitin family proteins and proteasome subunit S5a,” *Biochemistry*, 41(6), pp. 1767–1777. doi:10.1021/BI011892Y.

Wang, Q. *et al.* (2014) “The E3 Ubiquitin ligase AMFR and INSIG1 bridge the activation of TBK1 kinase by modifying the adaptor STING,” *Immunity*, 41(6), pp. 919–933. doi:10.1016/J.IMMUNI.2014.11.011/ATTACHMENT/BAD0839F-4C51-4211-AF91-7ECF98AD02A2/MMC1.PDF.

Watson, A. *et al.* (2019) “Surfactant Proteins A and D: Trimerized Innate Immunity Proteins with an Affinity for Viral Fusion Proteins,” *Journal of Innate Immunity*, 11(1), pp. 13–28. doi:10.1159/000492974.

Waugh, D.S. (2011) “An overview of enzymatic reagents for the removal of affinity tags,” *Protein Expression and Purification*, 80(2), p. 283. doi:10.1016/J.PEP.2011.08.005.

Weis, M.A. *et al.* (2010) “Location of 3-hydroxyproline residues in collagen types I, II, III, and V/XI implies a role in fibril supramolecular assembly,” *The Journal of biological chemistry*, 285(4), pp. 2580–2590. doi:10.1074/JBC.M109.068726.

Weissman, A.M. (2001) “Themes and variations on ubiquitylation,” *Nature Reviews Molecular Cell Biology 2001 2:3*, 2(3), pp. 169–178. doi:10.1038/35056563.

Whitmore, L. and Wallace, B.A. (2004) “DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data,” *Nucleic acids research*, 32(Web Server issue). doi:10.1093/NAR/GKH371.

Wilkinson, C.R.M. *et al.* (2001a) "Proteins containing the UBA domain are able to bind to multi-ubiquitin chains," *Nature cell biology*, 3(10), pp. 939–943. doi:10.1038/NCB1001-939.

Wilkinson, C.R.M. *et al.* (2001b) "Proteins containing the UBA domain are able to bind to multi-ubiquitin chains," *Nature cell biology*, 3(10), pp. 939–943. doi:10.1038/NCB1001-939.

Williams, K.L. *et al.* (2012) "UBQLN2/ubiquilin 2 mutation and pathology in familial amyotrophic lateral sclerosis," *Neurobiology of aging*, 33(10), pp. 2527.e3-2527.e10. doi:10.1016/J.NEUROBIOLAGING.2012.05.008.

Williamson, M.P. (1994) "The structure and function of proline-rich regions in proteins," *Biochemical Journal*, 297(2), pp. 249–260. doi:10.1042/BJ2970249.

Williamson, M.P. (2013) "Using chemical shift perturbation to characterise ligand binding," *Progress in Nuclear Magnetic Resonance Spectroscopy*, 73, pp. 1–16. doi:10.1016/J.PNMRS.2013.02.001.

Wing, S.S. (2003) "Deubiquitinating enzymes--the importance of driving in reverse along the ubiquitin-proteasome pathway," *The international journal of biochemistry & cell biology*, 35(5), pp. 590–605. doi:10.1016/S1357-2725(02)00392-8.

Wittekind, M. *et al.* (1993) "HNCACB, a High-Sensitivity 3D NMR Experiment to Correlate Amide-Proton and Nitrogen Resonances with the Alpha- and Beta-Carbon Resonances in Proteins," *JMRB*, 101(2), pp. 201–205. doi:10.1006/JMRB.1993.1033.

Woody, R.W. (1992) "Circular dichroism and conformation of unordered poly-peptides.," *Adv. Biophys. Chem.*, 2, pp. 37–79.

Wu, A.L. *et al.* (1999) "Ubiquitin-related proteins regulate interaction of vimentin intermediate filaments with the plasma membrane," *Molecular cell*, 4(4), pp. 619–625. doi:10.1016/S1097-2765(00)80212-9.

Wu, J. *et al.* (2018) "Tunicamycin specifically aggravates ER stress and overcomes chemoresistance in multidrug-resistant gastric cancer cells by inhibiting N-glycosylation," *Journal of Experimental and Clinical Cancer Research*, 37(1), pp. 1–12. doi:10.1186/S13046-018-0935-8/FIGURES/6.

Wu, J.J. *et al.* (2020) "ALS/FTD mutations in UBQLN2 impede autophagy by reducing autophagosome acidification through loss of function," *Proceedings of the National Academy of Sciences*, p. 201917371. doi:10.1073/pnas.1917371117.

Xia, Y. *et al.* (2014) "Pathogenic mutation of UBQLN2 impairs its interaction with UBXD8 and disrupts endoplasmic reticulum-associated protein degradation," *Journal of neurochemistry*, 129(1), pp. 99–106. doi:10.1111/JNC.12606.

Xu, C. *et al.* (2010) "Expanding the family of collagen proteins: Recombinant bacterial collagens of varying composition form triple-helices of similar stability," *Biomacromolecules*, 11(2), pp. 348–356. doi:10.1021/BM900894B/SUPPL_FILE/BM900894B_SI_001.PDF.

Xu, P. *et al.* (2009) “Quantitative proteomics reveals the function of unconventional ubiquitin chains in proteasomal degradation,” *Cell*, 137(1), pp. 133–145. doi:10.1016/J.CELL.2009.01.041.

Xu, Y. *et al.* (2002) “Streptococcal Scl1 and Scl2 Proteins Form Collagen-like Triple Helices,” *Journal of Biological Chemistry*, 277(30), pp. 27312–27318. doi:10.1074/JBC.M201163200.

Xu, Y. *et al.* (2003) “NMR and CD spectroscopy show that imino acid restriction of the unfolded state leads to efficient folding,” *Biochemistry*, 42(29), pp. 8696–8703. doi:10.1021/BI034006N.

Yamauchi, M. and Sricholpech, M. (2012) “Lysine post-translational modifications of collagen.” doi:10.1042/bse0520113.

Yoshii, S.R. *et al.* (2011) “Parkin mediates proteasome-dependent protein degradation and rupture of the outer mitochondrial membrane,” *The Journal of biological chemistry*, 286(22), pp. 19630–19640. doi:10.1074/JBC.M110.209338.

Yoshizumi, A. *et al.* (2009) “Self-association of streptococcus pyogenes collagen-like constructs into higher order structures,” *Protein Science*, 18(6), pp. 1241–1251. doi:10.1002/PRO.134.

Yoshizumi, A. *et al.* (2011) “Designed coiled coils promote folding of a recombinant bacterial collagen,” *Journal of Biological Chemistry*, 286(20), pp. 17512–17520. doi:10.1074/jbc.M110.217364.

Yu, J. *et al.* (2007) “Induced pluripotent stem cell lines derived from human somatic cells,” *Science (New York, N.Y.)*, 318(5858), pp. 1917–1920. doi:10.1126/SCIENCE.1151526.

Yu, Z. *et al.* (2010) “Noncollagenous region of the streptococcal collagen-like protein is a trimerization domain that supports refolding of adjacent homologous and heterologous collagenous domains,” *Protein Science*, 19(4), pp. 775–785. doi:10.1002/PRO.356.

Yu, Z. *et al.* (2014) “Bacterial collagen-like proteins that form triple-helical structures,” *Journal of Structural Biology*, 186(3), pp. 451–461. doi:10.1016/J.JSB.2014.01.003.

Yu, Z., Brodsky, B. and Inouye, M. (2011) “Dissecting a Bacterial Collagen Domain from Streptococcus pyogenes: SEQUENCE AND LENGTH-DEPENDENT VARIATIONS IN TRIPLE HELIX STABILITY AND FOLDING,” *Journal of Biological Chemistry*, 286(21), pp. 18960–18968. doi:10.1074/JBC.M110.217422.

Yuan, S. *et al.* (2015) “A Testis-specific Gene, Ubqlnl, Is Dispensable for Mouse Embryonic Development and Spermatogenesis,” *Molecular reproduction and development*, 82(6), p. 408. doi:10.1002/MRD.22504.

Yuan, W.C. *et al.* (2014) “K33-Linked Polyubiquitination of Coronin 7 by Cul3-KLHL20 Ubiquitin E3 Ligase Regulates Protein Trafficking,” *Molecular Cell*, 54(4), pp. 586–600. doi:10.1016/J.MOLCEL.2014.03.035/ATTACHMENT/BB844080-9B5B-40BC-A87D-EDA8FBB2D3C7/MMC3.MP4.

Yun Lee, D., Arnott, D. and Brown, E.J. (2013) “Ubiquilin4 is an adaptor protein that recruits Ubiquilin1 to the autophagy machinery,” *EMBO Reports*, 14(4), pp. 373–381. doi:10.1038/EMBOR.2013.22.

Zachari, M. and Ganley, I.G. (2017) “The mammalian ULK1 complex and autophagy initiation,” *Essays in Biochemistry*, 61(6), p. 585. doi:10.1042/EBC20170021.

Zhang, D., Raasi, S. and Fushman, D. (2008) “Affinity Makes the Difference: Nonselective Interaction of the UBA Domain of Ubiquilin-1 with Monomeric Ubiquitin and Polyubiquitin Chains,” *Journal of Molecular Biology*, 377(1), pp. 162–180. doi:10.1016/J.JMB.2007.12.029.

Zhao, H. *et al.* (2013) “Current Methods in Sedimentation Velocity and Sedimentation Equilibrium Analytical Ultracentrifugation,” *Current protocols in protein science / editorial board, John E. Coligan ... [et al.]*, 0 20(SUPPL.71). doi:10.1002/0471140864.PS2012S71.

Zhao, X. *et al.* (2015) “Collagen-Like Proteins (ClpA, ClpB, ClpC, and ClpD) Are Required for Biofilm Formation and Adhesion to Plant Roots by Bacillus amyloliquefaciens FZB42,” *PLOS ONE*, 10(2), p. e0117414. doi:10.1371/JOURNAL.PONE.0117414.

Zheng, T., Yang, Y. and Castañeda, C.A. (2020) “Structure, dynamics and functions of UBQLNs: at the crossroads of protein quality control machinery,” *The Biochemical journal*, 477(18), p. 3471. doi:10.1042/BCJ20190497.

Zinman, L. and Cudkowicz, M. (2011) “Emerging targets and treatments in amyotrophic lateral sclerosis,” *The Lancet. Neurology*, 10(5), pp. 481–490. doi:10.1016/S1474-4422(11)70024-2.