

**Mechanistic characterisation of IRE1 α 's
activation cascade and how it is perturbed by
cancer-associated mutations**

Nicholas Cormac Hurst

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

The University of Leeds
School of Molecular and Cellular Biology

September 2019

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Acknowledgements

Firstly, I have to acknowledge my incredible supervisor Anastasia Zhuravleva, who provided me with the perfect amounts of direction, inspiration and scientific freedom to explore this project and make these four years truly enjoyable and fascinating, thank you.

I must also thank Łukasz Wieteska for his guidance and more importantly, friendship throughout the three years, as well as my new-found love for sailing, frozen vodka and gherkins.

Additionally, a huge thank you to present and past Zhuravleva, Breeze and NMR group members; in particular Sebastien Cardon for all of the cake and French wine and Sam Dawes for the foundation for much of the work on the luminal domain.

Thank you also to Arnout Kalverda, Iain Manfield, Nasir Kahn and all other facility managers that have aided me. Additionally, thank you to the White Rose BBSRC DTP for funding my studies.

Thank you also to Atenas for putting up with me, especially during writing this thesis, providing scientific insight and for her constant support and supply of amazing food.

Lastly, thank you to all my friends and my family; my Mum and Gama for always believing in me, my brother, Tim, for inspiring me and my Dad for telling me to ‘read around the subject’; well Dad, I’ve done quite enough of that for a while.

Abstract

The unfolded protein response is the cell's reaction to an increased burden on the endoplasmic reticulum's (ER) protein folding machinery. The most conserved sensor of ER stress is IRE1 α , which clusters in response to stress to initiate a cellular signal. IRE1 α activation is a complex, multi-step mechanism, triggered by IRE1 α 's luminal domain's (IRE1-LD) response to fluctuating ER stress levels. Currently, the mechanisms of IRE1-LD's activation and termination are only partially understood, with conflicting models proposed.

Using a set of biophysical approaches, IRE1-LD's complex conformational landscape is characterised in unstressed conditions, upon addition of unfolded protein mimics (representative of ER stress) and the molecular chaperone BiP. The outputs suggest that in the absence of stress IRE1-LD exists in a conformational equilibrium between monomers, homodimers and homooligomers. In a concentration-dependent manner the unfolded protein mimics shift this equilibrium towards increasingly large oligomers, indicative of a proportional response to ER stress levels. Interestingly, these substrate-induced oligomers adopt fibril-like structures, providing a plausible model of the protein's clustering *in vivo*. In turn, addition of the molecular chaperone, BiP to these oligomers results in their disassembly, revealing a novel IRE1 α -BiP interaction. Notably, this process requires BiP's chaperone activity and the presence of ATP, reminiscent of Hsp70-assisted clathrin uncoating. Following this, the effects of four IRE1-LD cancer-associated mutations on the activation cascade are characterised, revealing novel allosteric sites that enable tuning of IRE1-LD's conformational landscape and thus IRE1 α activation. Lastly, solution NMR is used to probe the conformation of IRE1-LD's functionally important intrinsically disordered regions, revealing an allosteric network between key functional sites in the protein: the substrate-binding cleft, dimerisation and oligomerisation interfaces and the C-terminal juxtamembrane linker. Therefore, the research here augments our understanding of IRE1 α 's response to stress and identifies allosteric sites in the protein that offer novel potential therapeutic targets.

Table of Contents

Acknowledgements.....	ii
Abstract.....	iii
Table of Contents.....	iv
List of Tables.....	xii
List of Figures.....	xii
List of Abbreviations.....	xvi
1. Introduction.....	1
1.1. The endoplasmic reticulum.....	1
1.1.1. Protein quality control network.....	1
1.1.2 The unfolded protein response.....	2
1.1.2. Stress sensors.....	3
1.1.2.1. ATF6.....	3
1.1.2.2. PERK.....	3
1.1.2.3. IRE1.....	4
1.2. IRE1 α 's signalling.....	5
1.2.1. IRE1 α 's upstream activation.....	5
1.2.2. IRE1 α 's adaptive response.....	5
1.2.3. IRE1 α 's apoptotic response.....	7
1.3. IRE1 α in health and disease.....	7
1.3.1. The adaptive response in disease.....	8
1.3.2. The apoptotic response in disease.....	8
1.3.3. IRE1 α in cancer.....	9
1.4. Luminal domain activation.....	12
1.4.1. Luminal domain architecture.....	12
1.4.2. BiP repression model of IRE1 α luminal domain activation.....	15
1.4.3. Unfolded protein binding model.....	18
1.4.4. Disulphide formation in IRE1 α activation.....	19

1.4.5.	Discrepancies in the models	20
1.5.	IRE1 α 's cytoplasmic domain	22
1.5.1.	Cytoplasmic domain architecture	22
1.5.2.	Monomer	29
1.5.3.	Face-to-face dimer	29
1.5.4.	Back-to-back dimer	30
1.5.5.	Oligomer	34
1.6.	Methods for the characterisation of protein interactions	36
1.6.1.	Techniques to monitor protein interactions and changes to protein conformation 36	
1.6.1.1.	Fluorescence polarisation	36
1.6.1.2.	Microscale thermophoresis	37
1.6.1.3.	Nuclear magnetic resonance	38
1.6.2.	Size exclusion chromatography	43
1.6.3.	Turbidity assays	44
1.7.	Project aims	45
2.	Materials and methods	47
2.1.	List of general reagents and kits	47
2.2.	<i>E. coli</i> strains	49
2.3.	Molecular biology methods	49
2.3.1.	Polymerase chain reaction	49
2.3.1.1.	PCR reaction for DNA insert	49
2.3.1.2.	PCR clean up reaction	50
2.3.2.	SspI Digest	50
2.3.3.	Agarose gel	50
2.3.3.1.	Running agarose gels	50
2.3.3.2.	Gel extraction	50
2.3.4.	Ligation independent cloning reaction	50

2.3.4.1.	T4 Polymerase Digestion	50
2.3.4.2.	LIC Annealing Reaction	50
2.3.5.	Transformation.....	51
2.3.6.	Plasmid extraction.....	51
2.3.7.	Sequencing.....	51
2.3.8.	Mutagenesis	51
2.4.	General bacterial methods.....	52
2.4.1.	Sterilisation	52
2.4.2.	LB media.....	53
2.4.3.	LB agar plates	53
2.4.4.	M9 Media.....	53
2.4.5.	Protein expression.....	53
2.4.5.1.	30 and 37°C expression.....	53
2.4.5.2.	20°C expression	54
2.4.5.3.	Labelled expression.....	54
2.4.5.4.	Different construct's expression conditions.....	54
2.4.6.	Preparation of expression culture pellet for purification.....	55
2.4.6.1.	Cytoplasmic domain	55
2.4.6.2.	Luminal domain and BiP	56
2.5.	General protein methods	56
2.5.1.	Bacterial overexpression solubility tests.....	56
2.5.2.	IMAC purification of proteins	56
2.5.2.1.	Cytoplasmic domain purification.....	57
2.5.2.2.	Luminal domain purification.....	57
2.5.2.3.	BiP purification	58
2.5.3.	A280 concentration.....	58
2.5.4.	Bradford assay.....	59
2.5.5.	SDS-PAGE gels	59

2.5.6.	Buffer exchange protocols.....	59
2.5.6.1.	Dialysis.....	59
2.5.6.2.	NAP5 columns.....	59
2.5.7.	Concentrating protein.....	60
2.6.	General analytical protein methods.....	60
2.6.1.	Mass spectrometry.....	60
2.6.1.1.	Sequence identification.....	60
2.6.1.2.	Molecular mass.....	60
2.6.2.	Size exclusion chromatography.....	61
2.6.2.1.	Methodology.....	61
2.6.2.2.	Calibrations.....	61
2.6.3.	Circular Dichroism.....	62
2.7.	Nuclear magnetic resonance.....	63
2.7.1.	Amide TROSY experiments.....	63
2.7.1.1.	Cytoplasmic domain denatured and refolded.....	63
2.7.1.2.	Luminal domain constructs.....	63
2.7.1.3.	Truncation mutant assignment.....	63
2.7.1.4.	Temperature correlation.....	64
2.7.1.5.	Luminal domain with peptide experiments.....	64
2.7.2.	Solid-state NMR.....	64
2.8.	Cytoplasmic domain methods.....	64
2.8.1.	Refolding techniques.....	64
2.8.1.1.	Dilution refolding (final protocol).....	64
2.8.1.2.	Column refolding.....	65
2.8.1.3.	Dialysis refolding.....	65
2.8.2.	Optim 1000.....	65
2.8.3.	Dynamic light scattering.....	65
2.8.4.	Kinase assay.....	66

2.8.5.	Autophosphorylation assay	66
2.8.6.	Endoribonuclease assays.....	67
2.8.6.1.	Time dependent assays.....	67
2.8.6.2.	Concentration dependent assays	67
2.8.7.	TEV cleavage reaction	67
2.9.	Luminal domain methods.....	67
2.9.1.	Differential scanning calorimetry	67
2.9.2.	Preparation of peptides.....	68
2.9.3.	FITC-labelling of the luminal domain	68
2.9.4.	Fluorescence polarisation.....	68
2.9.5.	Microscale thermophoresis	68
2.9.6.	Turbidity assays	69
2.9.6.1.	OD400.....	69
2.9.6.2.	Soluble concentration determination.....	69
2.9.7.	BiP chaperone activity assay.....	69
2.9.8.	Negative stain electron microscopy	69
3.	Expression and characterisation of hIRE1 α 's cytoplasmic domain in <i>E. coli</i>	71
3.1.	Subcloning of the cytoplasmic domain construct	72
3.1.1.	Ligation independent subcloning principle.....	73
3.1.2.	Optimising the LIC reaction for cytoplasmic domain constructs	74
3.2.	Optimisation <i>E. coli</i> recombinant overexpression of the cytoplasmic domain.....	75
3.2.1.	Strategy to optimise protein solubility and yield	75
3.2.2.	Assessment of <i>E. coli</i> expression conditions	77
3.2.2.1.	Temperature	77
3.2.2.2.	IPTG concentration.....	77
3.2.2.3.	Small molecule additives	78
3.2.2.4.	<i>E. coli</i> expression strains.....	78
3.2.2.5.	Solubility tags	79

3.2.3.	Selection of an optimal condition	80
3.3.	Purification of the cytoplasmic domain protein from the insoluble fraction	80
3.3.1.	Preparation and purification of the insoluble fraction	80
3.3.2.	Mass spectrometry to validate the size and sequence of the cytoplasmic domain 82	
3.4.	Refolding reaction optimisation	84
3.4.1.	Refolding techniques in arginine buffer	84
3.4.2.	Buffer optimisation for stability	85
3.4.3.	Optimisation of refolding technique and subsequent removal of arginine	86
3.4.4.	Validation of refolding reaction	88
3.5.	Measuring the activity of the refolded cytoplasmic domain.....	90
3.5.1.	Species present in activity assay buffers	90
3.5.2.	The refolded cytoplasmic domain has kinase activity	91
3.5.3.	Autophosphorylation of the protein.....	93
3.5.4.	Analysis of the protein's endoribonuclease activity	95
3.5.4.1.	The refolded cytoplasmic domain cleaves the XBP-1 mRNA mimic	96
3.5.4.2.	The effect of concentration on endoribonuclease activity	97
3.5.4.3.	Characterising the endoribonuclease activity of monomeric and oligomeric species 99	
3.6.	Conclusion.....	101
4.	Mechanistic understanding of the luminal domain's activation cascade.....	103
4.1.	Introduction	103
4.2.	IRE1 α luminal domain conformations in the absence of binding partners	106
4.3.	IRE1 α substrate-induced oligomerisation	109
4.3.1.	Interaction with peptide induces luminal domain oligomerisation.....	109
4.3.2.	Quantification of peptide-induced formation of insoluble oligomers.....	110
4.3.2.1.	Δ EspP promotes formation of insoluble species in a concentration-dependent manner 111	

4.3.2.2.	The effect of peptide binding affinity on luminal domain oligomerisation	114
4.3.2.3.	Luminal domain interface mutants affect peptide-induced oligomerisation	115
4.4.	BiP's interaction with oligomerised IRE1 α	120
4.5.	Conformational characterisation of the insoluble oligomers and BiP interaction	123
4.5.1.	Insoluble luminal domain species form fibrillar assemblies	124
4.5.2.	The fibrils consist of folded protein and BiP binding affects the luminal domain's disordered regions	126
4.6.	Conclusion	128
5.	Characterisation of the luminal domain's intrinsically disordered regions using solution NMR	133
5.1.	Introduction	133
5.2.	Initial characterisation of the luminal domain by solution NMR (repeated work of Sam Dawes)	136
5.3.	Partial assignments of IRE1 α 's luminal domain spectrum	137
5.4.	$^1\text{H}^{\text{N}}$ Chemical shift temperature coefficients suggests that no secondary structure is observed	139
5.5.	Different segments of the linker region experience differential dynamic processes	141
5.6.	Characterising different IRE1 α conformational states by NMR	144
5.6.1.	Monomeric protein has a differential linker region conformation to dimeric/oligomeric protein	145
5.6.2.	The dimer to oligomer transition involves changes in the protein's disordered regions	149
5.6.3.	There are no significant changes to luminal domain disordered regions upon peptide binding	151
5.7.	Conclusion	155
6.	Cancer-associated mutants affect the conformational landscape of IRE1 α	160
6.1.	Introduction	160
6.2.	Initial characterisation of the mutants' effects	163

6.2.1.	Three of the mutants can be overexpressed as soluble protein in <i>E. coli</i>	163
6.2.2.	The thermal stability of the expressed soluble cancer-associated mutants.....	164
6.2.3.	The cancer-associated mutants do not significantly perturb secondary structure 165	
6.3.	A414T mutation results in the formation of insoluble oligomers.....	166
6.3.1.	Determination of the linker region's effect on solubility	166
6.3.2.	Optimisation of expression does not promote soluble expression of the A414T construct 168	
6.3.3.	Perturbation of the dimer and oligomer interfaces results in soluble A414T expression.....	170
6.3.4.	The A414T mutation doesn't cause a significant change in secondary structure of the D123P construct	172
6.4.	How the cancer-associated mutations affect IRE1 α 's conformational landscape	173
6.4.1.	No perturbations to luminal domain activation are observed in the N244S mutant 173	
6.4.2.	The S296F mutant affects luminal domain oligomerisation and the stability of disulphide bonds.....	174
6.4.2.1.	The S296F mutant moderately stabilises oligomers.....	174
6.4.2.2.	The S296F mutant appears to protect formed disulphide bonds	176
6.4.3.	The A414T mutation promotes peptide-induced oligomerisation in the D123P construct 178	
6.4.4.	The V418M mutation does not perturb the luminal domain's conformation landscape 181	
6.5.	The cancer-associated mutations do not perturb BiP's interaction with oligomeric luminal domain.....	182
6.6.	Conclusion.....	185
7.	The effects of cancer-associated mutations on the luminal domain's intrinsically disordered regions	189
7.1.	Introduction	189

7.2.	Effects of the core domain cancer-associated mutations on luminal domain disordered regions	190
7.2.1.	The N244S mutant has no observable effects on the disordered regions of the protein	191
7.2.2.	The S296F mutation alters disordered region conformation.....	191
7.3.	How the linker region mutations affect the luminal domain's disordered regions ...	195
7.3.1.	The A414T mutation only causes local perturbations in the linker region	195
7.3.2.	The V418M mutation causes long-range perturbations of the linker region	196
7.4.	Conclusion	199
8.	Conclusions and future perspectives.....	204
9.	Appendices.....	212
10.	Bibliography	222

List of Tables

Table 1.1	10
Table 1.2	23
Table 2.1	47
Table 2.2	54
Table 5.1	141
Table 5.2	156
Table 6.1	161
Table 6.2	167
Table 9.1	218

List of Figures

Figure 1.1	2
Figure 1.2	4
Figure 1.3	6
Figure 1.4	13
Figure 1.5	14
Figure 1.6	16
Figure 1.7	17

Figure 1.8.....	18
Figure 1.9.....	19
Figure 1.10.....	21
Figure 1.11.....	22
Figure 1.12.....	26
Figure 1.13.....	28
Figure 1.14.....	30
Figure 1.15.....	32
Figure 1.16.....	34
Figure 1.17.....	37
Figure 1.18.....	38
Figure 1.19.....	40
Figure 1.20.....	42
Figure 1.21.....	44
Figure 1.22.....	45
Figure 2.1.....	58
Figure 2.2.....	61
Figure 2.3.....	62
Figure 2.4.....	66
Figure 3.1.....	72
Figure 3.2.....	73
Figure 3.3.....	75
Figure 3.4.....	76
Figure 3.5.....	77
Figure 3.6.....	78
Figure 3.7.....	79
Figure 3.8.....	80
Figure 3.9.....	81
Figure 3.10.....	83
Figure 3.11.....	85
Figure 3.12.....	86
Figure 3.13.....	87
Figure 3.14.....	88
Figure 3.15.....	89
Figure 3.16.....	90

Figure 3.17	91
Figure 3.18	92
Figure 3.19	93
Figure 3.20	95
Figure 3.21	96
Figure 3.22	97
Figure 3.23	98
Figure 3.24	99
Figure 3.25	100
Figure 3.26	101
Figure 4.1	105
Figure 4.2	107
Figure 4.3	108
Figure 4.4	109
Figure 4.5	110
Figure 4.6	112
Figure 4.7	113
Figure 4.8	115
Figure 4.9	116
Figure 4.10	118
Figure 4.11	119
Figure 4.12	120
Figure 4.13	121
Figure 4.14	122
Figure 4.15	123
Figure 4.16	125
Figure 4.17	126
Figure 4.18	128
Figure 4.19	131
Figure 5.1	134
Figure 5.2	135
Figure 5.3	137
Figure 5.4	138
Figure 5.5	140
Figure 5.6	142

Figure 5.7.....	144
Figure 5.8.....	147
Figure 5.9.....	148
Figure 5.10.....	150
Figure 5.11.....	151
Figure 5.12.....	152
Figure 5.13.....	154
Figure 5.14.....	158
Figure 6.1.....	162
Figure 6.2.....	164
Figure 6.3.....	165
Figure 6.4.....	166
Figure 6.5.....	167
Figure 6.6.....	168
Figure 6.7.....	169
Figure 6.8.....	171
Figure 6.9.....	172
Figure 6.10.....	174
Figure 6.11.....	175
Figure 6.12.....	177
Figure 6.13.....	180
Figure 6.14.....	181
Figure 6.15.....	183
Figure 6.16.....	184
Figure 6.17.....	186
Figure 7.1.....	190
Figure 7.2.....	191
Figure 7.3.....	192
Figure 7.4.....	194
Figure 7.5.....	196
Figure 7.6.....	197
Figure 7.7.....	199
Figure 7.8.....	200
Figure 7.9.....	202
Figure 8.1.....	205

Figure 8.2	208
Figure 8.3	210
Figure 9.1	212
Figure 9.2	213
Figure 9.3	213
Figure 9.4	214
Figure 9.5	215
Figure 9.6	218
Figure 9.7	220
Figure 9.8	221

List of Abbreviations

A260 - Absorbance at 260nm
A280 - Absorbance at 280nm
ADP - Adenosine diphosphate
APP - Alzheimer's precursor protein
ASK1 - Apoptosis signal-regulating kinase
ATF6 - Activating transcription factor 6
ATP - Adenosine triphosphate
BCM - Barycentric mean
BEST - Band-selective Excitation Short-Transient
BHQ - Black hole quencher
BiP - Immunoglobulin binding protein
BSA - Bovine serum albumin
CD - Circular dichroism
C/EBP - CCAAT/enhancer-binding proteins
CHOP - CCAAT-enhancer-binding protein homologous protein
CSTC - Chemical shift temperature coefficient
Da - Daltons
dCTP - Deoxycytosine triphosphate
dGTP - Deoxyguanine triphosphate
DLS - Dynamic light scattering
DMSO - Dimethyl sulfoxide
DNA - Deoxyribonucleic acid
DTT - Dithiothreitol

DSC - Differential scanning calorimetry
E. coli - Escherichia coli
EDTA - Ethylenediaminetetraacetic acid
eIF2 α - Eukaryotic translation initiation factor 2 α
EM - Electron microscopy
ER - Endoplasmic reticulum
ERdj4 - Endoplasmic reticulum-localised DnaJ 4
ERSE - Endoplasmic reticulum stress response element
ESI - Electron spray ionisation
FITC - Fluorescein isothiocyanate
FRET - Fluorescence resonance energy transfer
GDP - Guanosine diphosphate
GLS - Golgi localisation signal
GST - Glutathione S-transferase
GTP - Guanosine triphosphate
HEPES - 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HLE - Helix-loop element
HMK - HEPES, MgCl₂ and KCl buffer (Section 2.5.6.1)
Hsp - Heat shock protein
HSQC- Heteronuclear single quantum coherence
IMAC - Immobilised metal affinity chromatography
INEPT - Insensitive nuclei enhanced by polarisation transfer
IPTG - Isopropyl β -D-1-thiogalactopyranoside
IRE1 - Inositol requiring enzyme 1
IRS1 - Insulin receptor substrate 1
ITC - Isothermal titration calorimetry
JNK - c-Jun N-terminal kinase
KEN - Kinase-extension nuclease
LB - Lysogeny broth
LD - Luminal domain
LIC- Ligation independent cloning
MAVS - Mitochondrial antiviral-signalling protein
MBP - Maltose-binding protein
MHC - Major histocompatibility complex
miR - Micro messenger ribonucleic acid

mRNA - Messenger ribonucleic acid
MS - Mass spectrometry
MST - Microscale thermophoresis
MW - Molecular weight
NMR - Nuclear magnetic resonance
NOD - Nucleotide-binding oligomerisation domain
OD400 - Optical density/ absorbance at 400 nm
OD600 - Optical density/absorbance at 600 nm
PCR - Polymerase chain reaction
PDB - Protein data bank
PDI - Protein disulphide isomerase
PERK - Protein kinase R-like endoplasmic reticulum kinase
PONDR - Prediction of natural disordered regions
RIDD - Regulated IRE1-dependent decay
RMSD - Root mean square deviation
SDS-PAGE - Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEC - Size exclusion chromatography
S/N - Signal to noise
SP1/2 - Site-1/2 protease
SOFAST - band-Selective Optimized Flip-Angle Short-Transient
TAE - Tris acetate EDTA
TCEP - Tris (2-carboxyethyl) phosphine
TEM - Transmission electron microscopy
TEV - Tobacco etch virus
TNFR - Tumour necrosis factor receptor
TRAF2 - Tumour necrosis factor receptor associated factor 2
TRIS - Tris(hydroxymethyl)aminomethane
TROSY - Transverse relaxation optimized spectroscopy
UV - Ultraviolet
XBP-1 - X-box binding protein 1

1. Introduction

1.1. The endoplasmic reticulum

The cell is an exceptionally complex machine, in this machine the majority of operations critical for function are carried out by proteins. Each specific protein has one or many roles to carry out, for which it has evolved a polypeptide sequence that will allow it to fold into a structure specific for its function. 39% of the proteome is predicted to consist of membrane bound and secreted proteins, these often have critical cellular functions borne out by the fact that of all proteins targeted by marketed drugs, ~83% are membrane or secreted (Uhlen et al., 2015). Newly translated, unfolded membrane and secretory proteins are processed and achieve their native fold in a cellular organelle; the endoplasmic reticulum (ER).

1.1.1. Protein quality control network

Due to the requirement of secretory and membrane proteins, the concentration of proteins in the endoplasmic reticulum can reach 100mg/mL, this environment can severely complicate the process of an unfolded polypeptide chain from achieving its native fold (Stevens and Argon, 1999). The endoplasmic reticulum therefore possesses a protein quality control network which maintains the correct processing of the majority of newly translated proteins (Figure 1.1). This network, among other machinery, includes molecular chaperones. Molecular chaperones are proteins that interrogate the newly synthesised proteins. Correctly folded proteins are chaperoned to the next step of processing and misfolded proteins are retained to be potentially repaired but otherwise degraded (Adams et al., 2019).

In some cases, the protein quality control machinery can be overwhelmed by unfolded and misfolding proteins, this is called endoplasmic reticulum stress. This stress can be triggered by multiple factors including physiological, such as during antibody (Reimold et al., 2001) and hormone production (Sharma et al., 2015) and also pathological factors such as hypoxia, nutrient deprivation, inflammation or infection (Yoshida, 2007). Here, the increased amount of unfolded proteins in the endoplasmic reticulum mean that the molecular chaperones binding to them are saturated and the protein quality control network is overwhelmed. In this case the cell requires another response to adapt to the stress and return the endoplasmic reticulum to an unstressed equilibrium by increasing the capacity of the protein quality control machinery; this is known as the unfolded protein response (Figure 1.1).

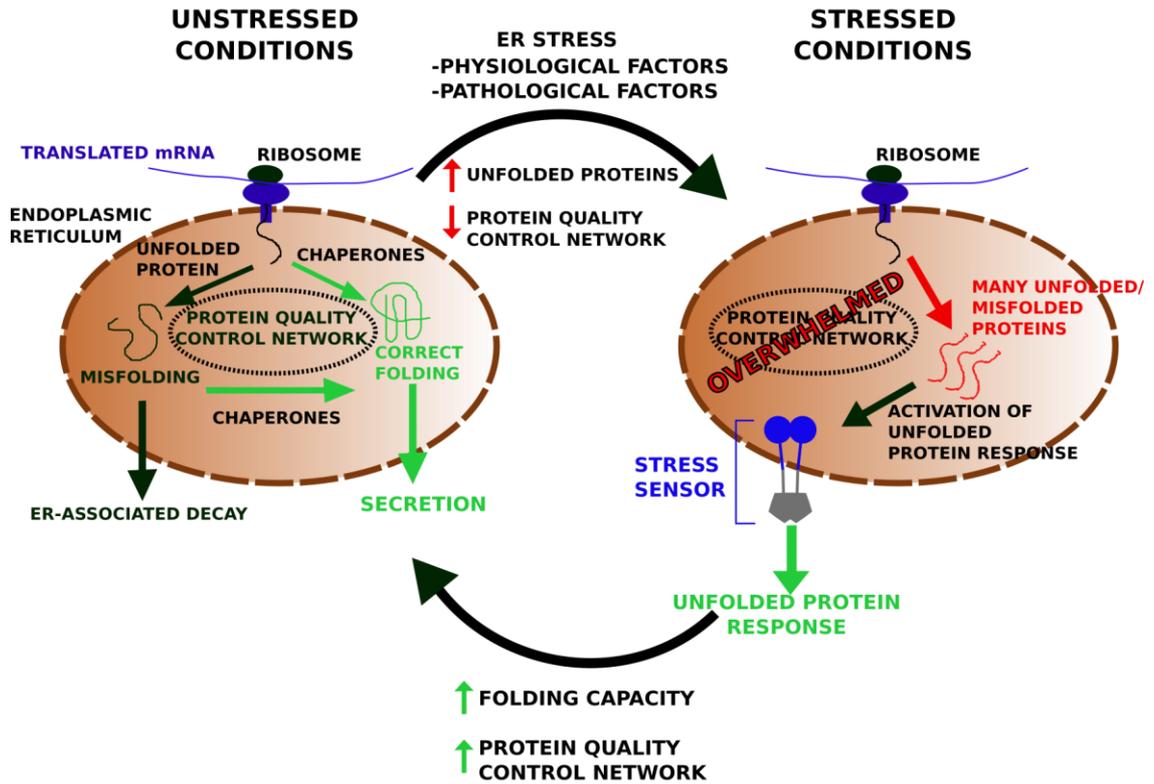


Figure 1.1

Representation of the endoplasmic reticulum in stressed and unstressed conditions. In unstressed conditions the protein quality control system ensures the correct folding of newly translated membrane and secretory proteins. Physiological and pathological factors can initiate endoplasmic reticulum stress, causing the protein quality control network to be overwhelmed. An abundance of unfolded and misfolding proteins in the endoplasmic reticulum promote activation of the unfolded protein response through stress sensors that transverse the endoplasmic reticulum membrane. The stress sensors produce a signal to adapt to the stress by increasing the folding capacity and machinery of the protein quality control network, returning the endoplasmic reticulum to non-stressed conditions.

1.1.2 The unfolded protein response

The unfolded protein response is activated by stress sensing proteins (Figure 1.2), of which there are three in mammalian cells; activating transcription factor 6 (ATF6), double stranded RNA-activated protein kinase (PKR)-like endoplasmic reticulum kinase (PERK) and inositol-requiring enzyme 1 (IRE1) (Walter and Ron, 2011). All three stress sensors transverse the endoplasmic reticulum membrane and have a domain inside the endoplasmic reticulum to sense stress, activation of this domain occurs due to an abundance of unfolded proteins in the endoplasmic reticulum which subsequently promotes activation of the cytoplasmic domains of the sensors

(Bertolotti et al., 2000). The cytoplasmic domains of the sensors signal to the cell for a response through divergent pathways. Initially the signal to the cell is to adapt to the increased burden on the endoplasmic reticulum. This adaptive response includes but is not limited to reducing the amount of newly synthesised proteins (Harding et al., 1999; Koumenis et al., 2002), increased autophagy (Yorimitsu et al., 2006), increased availability molecular chaperones (Ron and Walter, 2007), more machinery for endoplasmic reticulum-associated decay (Travers et al., 2000) and to increase the size of the endoplasmic reticulum (Schuck et al., 2009).

The stress sensors act in a negative feedback loop; once the stress is resolved the unfolded and misfolded proteins are cleared from the endoplasmic reticulum and therefore the signal is terminated. However, if the signal is not alleviated the adaptive response can turn to an apoptotic one. This includes activation of classical apoptotic machinery by the stress sensors such as the C/EBP homologous protein (CHOP) and the c-Jun Kinase (JNK) pathway (Urano et al., 2000; Marciniak et al., 2004). Each of the three endoplasmic reticulum stress sensing proteins have divergent activation mechanisms and initial signalling pathways but act together to accomplish these general cellular outcomes, the signalling of each sensor is briefly described in the following section.

1.1.2. Stress sensors

1.1.2.1. ATF6

After sensing stress, the ATF6 protein is transported to the golgi apparatus membrane mediated by its two golgi localisation sequences (GLS1 and GLS2; Figure 1.2) (Shen, J. et al., 2002). Once in the golgi apparatus, ATF6's endoplasmic reticulum luminal domain is cleaved by serine protease site-1 protease and metalloprotease site-2 protease (SP1 and SP2) (Chen et al., 2002; Haze et al., 1999; Shen, J. and Prywes, 2004; Ye et al., 2000). This cleavage allows the cytoplasmic domain of the protein to migrate to the cell nucleus where it acts as a transcriptional promoter for endoplasmic reticulum stress response elements (ERSE) to promote the unfolded protein response (Kokame et al., 2001; Yoshida et al., 1998).

1.1.2.2. PERK

In the case of the stress sensor PERK, activation promotes dimerisation and autophosphorylation of its cytoplasmic kinase signalling domain (Kebache et al., 2004). Once activated, among other activities, PERK can phosphorylate eukaryotic translation initiation factor 2 α (eIF2 α ; Figure 1.2), this modification prevents eIF2 α from exchanging GDP; a process critical for translation initiation

(Merrick, 2004; Gebauer and Hentze, 2004; Harding et al., 1999). Thus, activation of PERK inhibits translation of new proteins, giving the endoplasmic reticulum respite and therefore the opportunity for other components of the adaptive response to act.

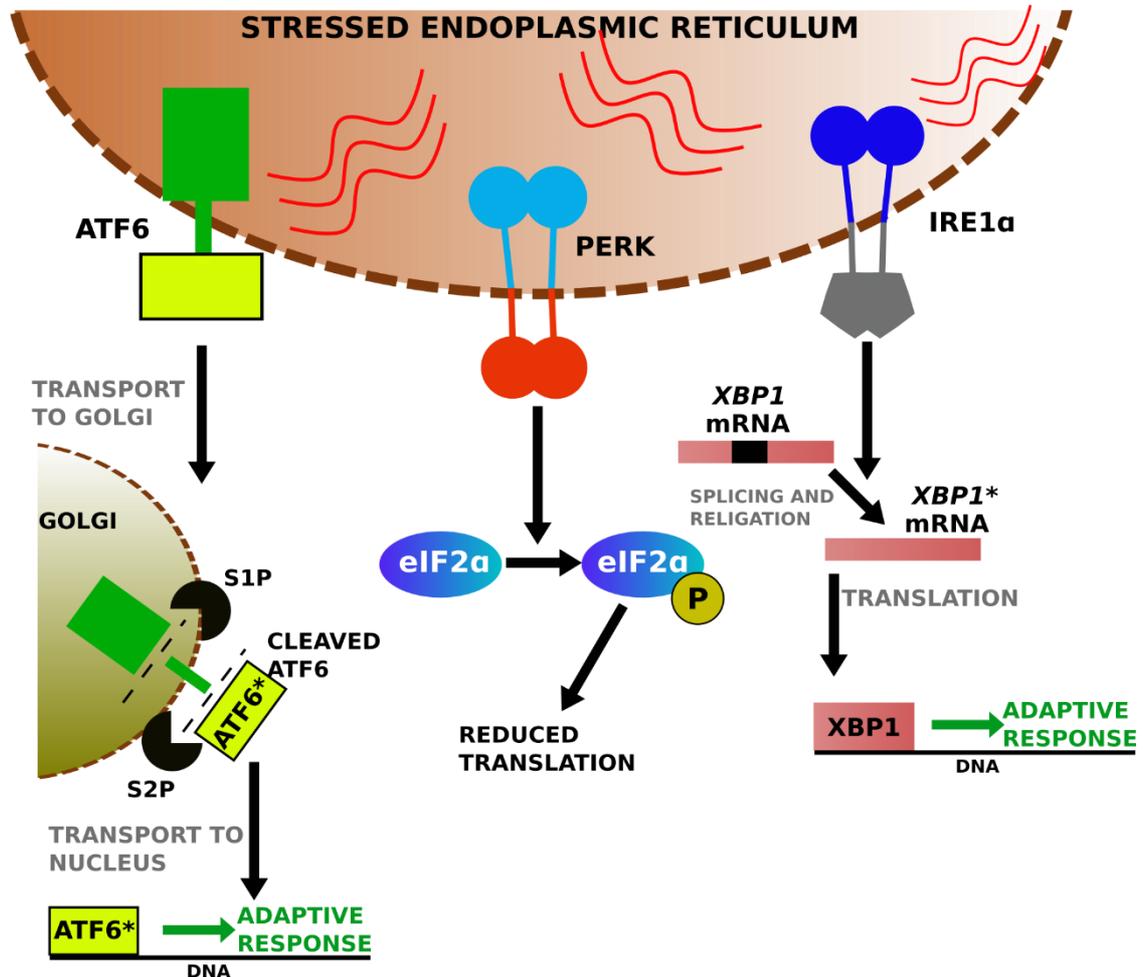


Figure 1.2

Simplified representation of the activation and roles of the three stress sensors ATF6, PERK and IRE1 α . Upon endoplasmic reticulum stress all three proteins are activated. ATF6 is translocated to the golgi apparatus where it is cleaved, it's cytoplasmic domain then acts as a transcription factor for unfolded protein response genes. PERK acts to phosphorylate eIF2 α which inhibits translation of new proteins. IRE1 α acts to splice XBP-1 mRNA, which is then re-ligated and translated to act as a transcription factor for unfolded protein response genes.

1.1.2.3. IRE1

IRE1 is the most conserved stress sensor and although its responses interact with the other stress sensors in human cells, it is the only sensor present in all eukaryotes (Mori, 2009). There are two homologues in humans, IRE1 α and IRE1 β . IRE1 β is only expressed in intestinal epithelial cells

and is thought to provide specialisation for the responses of these cells; IRE1 α on the other hand is expressed in the majority of cells and is the focus of this thesis (Wang, X.Z. et al., 1998; Imagawa et al., 2008). Put briefly, activation of IRE1 α involves autophosphorylation and clustering of the cytoplasmic domain which activates its endoribonuclease activity. This activity is initially used to splice the *XBP-1* mRNA, which is then re-ligated to form mRNA for the active XBP-1 transcription factor, which promotes the transcription of unfolded protein response genes (Figure 1.2) (Walter and Ron, 2011), the signalling and activation of IRE1 α will be discussed in more detail in the following sections.

1.2. IRE1 α 's signalling

1.2.1. IRE1 α 's upstream activation

IRE1 α is a transmembrane protein with a domain inside the endoplasmic reticulum and one in the cell cytoplasm. The stress sensing domain is the endoplasmic reticulum luminal domain which is tethered to a single transmembrane helix by a long linker region, on the other side of the membrane is the signalling cytoplasmic domain (Wang, P. et al., 2018; Zhou et al., 2006). The precise activation mechanism of IRE1 α 's luminal domain is still a contentious subject and will be discussed in more detail in Section 1.4. However, upon induction of stress IRE1 α 's luminal domain dimerises and oligomerises (Figure 1.3) (Bertolotti et al., 2000; Gardner and Walter, 2011; Amin-Wetzel et al., 2017). The clustering of the luminal domain causes an increase in the local concentration of the cytoplasmic domain on the other side of the membrane, the cytoplasmic domain contains both kinase and endoribonuclease activities. The increase in the cytoplasmic domain's local concentration promotes dimerisation and autophosphorylation, which in turn activates its endoribonuclease activity which dictates IRE1 α 's downstream response. There are, what can be grouped as two opposing outcomes to IRE1 α activation; the adaptive and apoptotic responses.

1.2.2. IRE1 α 's adaptive response

IRE1 α 's endoribonuclease activity will initially target, and splice *XBP-1* mRNA as mentioned previously. Upon IRE1 α splicing of *XBP-1*, 26 nucleotides are removed from its sequence, the spliced mRNA fragments are then ligated by the RtcB protein (Figure 1.3) (Lu et al., 2014). The ligated form of *XBP-1* mRNA can then be translated into the transcriptional activator XBP-1, which delivers a strong adaptive response to endoplasmic reticulum stress. This includes promoting the production of post-translational modification enzymes which promote the correct folding of newly synthesised proteins, machinery involved in the decay of misfolding proteins

and the previously described molecular chaperones, all of which are part of the protein quality control machinery and act to cope with the unfolded protein burden on the endoplasmic reticulum (Poothong et al., 2017; Yoshida et al., 2001; Calfon et al., 2002; Lee, A.H. et al., 2003).

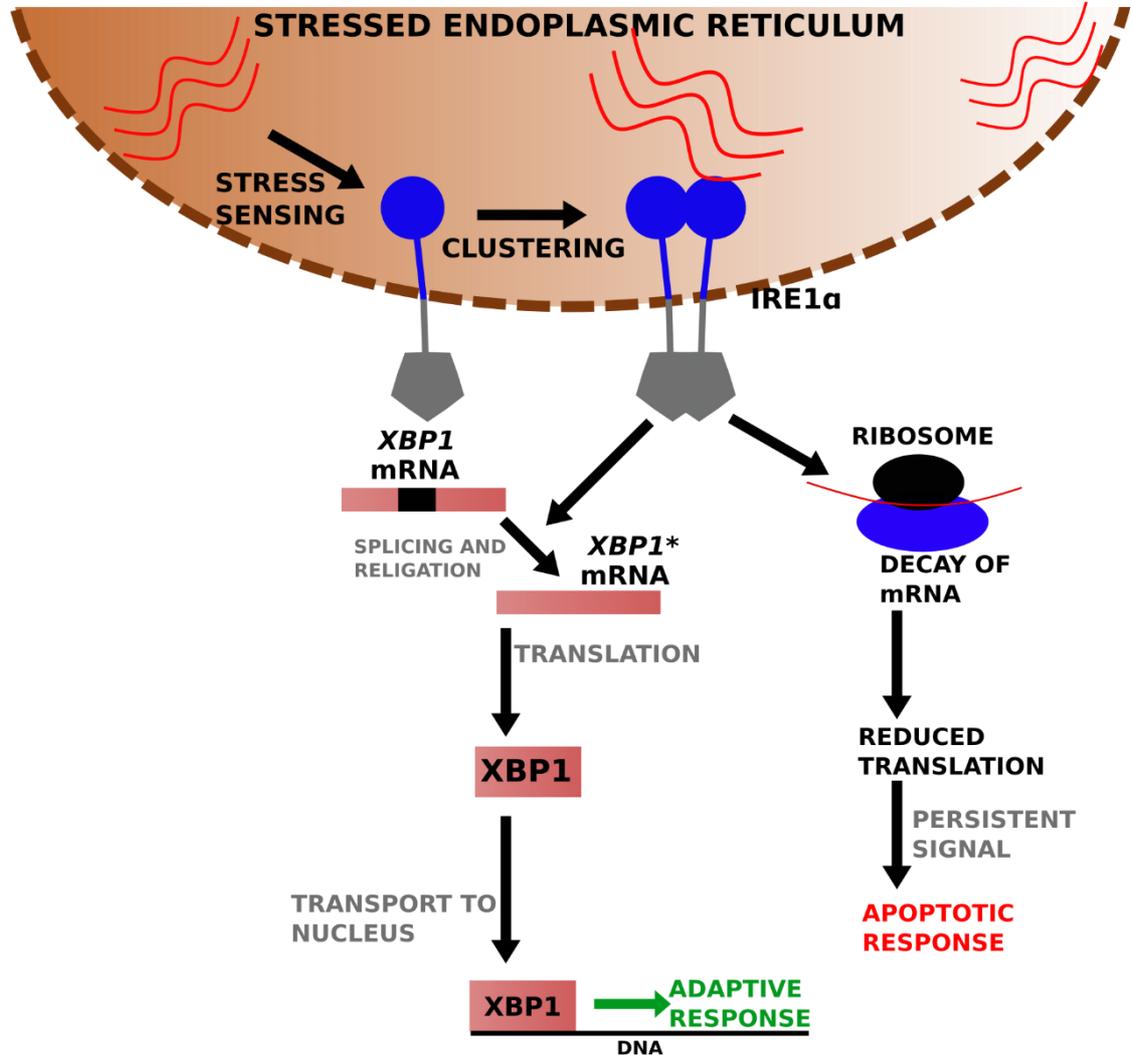


Figure 1.3

IRE1 α senses endoplasmic reticulum stress through its luminal domain, this domain then clusters to increase the local concentration of the protein's cytoplasmic domain. The cytoplasmic domain's endoribonuclease activity is activated through dimerisation and autophosphorylation and has roles in splicing XBP-1 mRNA and the decay of many different mRNA targets.

The adaptive activity of IRE1 α is the initial response to stress and is associated with the clustering of IRE1 α . Formation of clusters of IRE1 α and splicing of XBP-1 are observed from the onset of stress signals with the peak activity appearing to be at around four hours, although there are various studies with different conditions and therefore differing values for this (Li, H. et al., 2010; Tam et al., 2014; Lin et al., 2007; Prischi et al., 2014; Sundaram et al., 2017). After this time,

XBP-1 splicing declines and degradation of other mRNAs by IRE1 α begins to increase, this is a different IRE1 α endoribonuclease activity that is promiscuous and represents a move away from the adaptive response to an apoptotic response of the protein, it is known as regulated IRE1-dependent decay (RIDD; Figure 1.3).

1.2.3. IRE1 α 's apoptotic response

IRE1 α 's endoribonuclease domain's RIDD activity has been observed from the onset of stress, however, some studies suggest that this activity occurs predominantly only after *XBP-1* splicing is in decline (Tam et al., 2014; Hollien et al., 2009; Han et al., 2009). IRE1 α 's RIDD activity does not appear to be in competition with its *XBP-1* splicing activity, although interestingly the RIDD mRNA targets identified are enriched with similar consensus sequences and predicted secondary structure to *XBP-1* mRNA (Oikawa et al., 2010; Coelho and Domingos, 2014). RIDD activity acts to degrade mRNAs preventing their translation and subsequent increased burden on the stressed endoplasmic reticulum. In one study the levels of two hundred mRNA's were found to be reduced upon RIDD induction, these mRNA's were enriched with the mRNA of secretory pathway proteins (Coelho and Domingos, 2014), additionally some RIDD targets have been associated with regulation of metabolism and immunity (Lee, A.H. et al., 2011; Osorio et al., 2014). RIDD targets also include mRNA beneficial to the cell during times of ER stress, such as mRNA of the molecular chaperone; BiP and other proteins involved in correct folding of endoplasmic reticulum proteins (Han et al., 2009). Not only degrading mRNA of beneficial proteins, RIDD activity also targets anti-apoptotic microRNAs during high stress conditions (miR -17, -34a, -96 and -125), these microRNAs repress the proapoptotic caspase 2 (Upton et al., 2012). Therefore IRE1 α 's RIDD activity appears to have both a positive and deleterious effect on the cell, occurring after the adaptive response and in high stress conditions to promote an apoptotic response.

IRE1 α also forms downstream responses independent of its endoribonuclease activity. One such interaction is with TNFR-associated factor 2 (TRAF2) and apoptosis signal-regulating kinase 1 (ASK1) to eventually trigger the JNK pathway which promotes apoptosis (Urano et al., 2000; Nishitoh et al., 2002; Mauro et al., 2006). IRE1 α 's opposed signalling pathways mean that when perturbed, its activity can influence disease outcome by causing undesired cell death or promoting cell survival where apoptosis is required.

1.3. IRE1 α in health and disease

IRE1 α 's differential activities allow it react to endoplasmic reticulum stress in an appropriate manner and therefore maintain homeostasis. However, due to its strong signalling capabilities an imbalance in its response can be instrumental in pathogenesis. Prolonged disturbance of normal cell or endoplasmic reticulum function can lead to a detrimental apoptotic response, such as in type I diabetes. On the other hand, the adaptive response can be hijacked by some viruses and cancers to aid in abhorrent protein production and cell proliferation. Some such cases of each of IRE1 α 's activities promoting pathogenesis are discussed in the following sections.

1.3.1. The adaptive response in disease

(IRE1 α 's adaptive response is the first prominent signal following endoplasmic reticulum stress and promotes an increase in the protein folding capacity of the cell. The IRE1 α and *XBP-1* signalling pathway is involved in inflammation and therefore implicated in a range of disease where inflammation has a detrimental effect such as cystic fibrosis (Ribeiro and Lubamba, 2017). In Alzheimer's disease there is a causal link between IRE1 α activity and prognosis and although IRE1 α 's *XBP-1* activity appears to increase production of the Alzheimer's precursor protein (APP), sustained RIDD activity from age related sustained stress also appears to have a role in pathogenesis (Duran-Aniotz et al., 2017; Hollien et al., 2009; Cescon et al., 2016). Interestingly the unfolded protein response pathways can be hijacked by viruses to supply strong adaptive signals so that the endoplasmic reticulum isn't overwhelmed by production of the viral capsid, one example of this is in hepatitis B (Li, B. et al., 2007). IRE1 α 's adaptive response is also hijacked in some cancers, which will be discussed in Section 1.3.3.

1.3.2. The apoptotic response in disease

Numerous diseases can promote endoplasmic reticulum stress, if the stress signal is strong enough and persistent then IRE1 α will signal for an apoptotic outcome which can have severe deleterious effects to the organism as a whole. The unfolded protein response is involved in both type I and II diabetes, pancreatic β -cells produce the secreted protein insulin upon processing of food to glucose, this insulin production can cause endoplasmic reticulum stress. In this case prolonged activation of the unfolded protein response perturbs β -cell function and therefore causes type II diabetes (Scheuner and Kaufman, 2008). Specifically, IRE1 α activation can lead to activation of the JNK pathway, this causes phosphorylation of insulin receptor substrate 1 (IRS1), which inhibits insulin action (Ozcan et al., 2004; Chakrabarti et al., 2011).

Some forms of retinitis pigmentosa occur due to a mutant form of rhodopsin that cannot achieve its native fold, in this case IRE1 α and the unfolded protein response have been implicated in disease progression by promoting apoptosis leading to blindness (Lin and Lavail, 2010). IRE1 α 's signalling to TRAF2 has been implicated in a link between endoplasmic reticulum stress and nucleotide-binding oligomerisation domain containing protein (NOD) 1 and 2 mediated inflammation which is involved in Crohn's disease (Keestra-Gounder et al., 2016).

1.3.3. IRE1 α in cancer

As per the previous examples, there are many cases where IRE1 α 's activity influences disease outcome, where IRE1 α activity is mainly a product of distal effects in the cell or body. However, cancer is a mutation driven disease, where mutations that benefit cancer progression are selected for in a hostile environment with increasingly unstable DNA quality control mechanisms (Jeggo et al., 2016; Kaufman et al., 2002). Creating an imbalance in IRE1 α 's activity can be particularly important in cancer pathogenesis as the nature of cancer development induces significant endoplasmic reticulum stress; cancer cells typically have a high consumption of glucose, high proliferation rates and outgrow their environment's nutrient resources causing hypoxia (Kaufman et al., 2002; Chalmers et al., 2019).

Therefore it seems unsurprising that IRE1 α is mutated in 3% of all human cancers (examples in Table 1.1), and that there are plentiful examples of IRE1 α 's adaptive response being increased in different tumour types to give a worse prognosis (Tameire et al., 2015; Chen et al., 2014; Romero-Ramirez et al., 2009; Cuevas et al., 2017; Chalmers et al., 2019).

However, there are also contradictory cases where *XBP-1* splicing and IRE1 α activity improve prognosis (Chalmers et al., 2019; Denoyelle et al., 2006; Kaser et al., 2008). It therefore appears that IRE1 α 's activities being beneficial or detrimental to cancer development can be tumour and progression stage specific. A recent study of glioblastoma characterised tumours based on IRE1 α 's adaptive and RIDD activities. Here tumours with high adaptive and low RIDD activities were more aggressive, lowering prognosis than those groups of tumours with lower adaptive and higher RIDD activities, therefore the specificities of IRE1 α activity in cancer is beginning to be understood in different cancer types (Lhomond et al., 2018).

The nature of cancer's ability to commandeer IRE1 α 's activities through mutations to create an imbalance offers the unique opportunity to potentially separate IRE1 α 's adaptive and apoptotic activities using these mutations. Study of the mutations can be used to not only provide future

therapeutic targets against cancer, but also for other diseases and more generally to understand how IRE1 α switches between its opposing activities during normal function.

Although one may expect mutations of IRE1 α to be concentrated in the endoribonuclease domain or at the TRAF2 interaction site, the mutations are observed throughout the protein suggesting that it may not only be the cytoplasmic domain that dictates the cellular outcome after stress sensing and activation. Some cancer-associated mutations have been initially characterised, with mutations of the cytoplasmic domain inhibiting stress-induced apoptosis and IRE1 α 's adaptive activity (Ghosh et al., 2014). Two glioblastoma mutations of the luminal domain have recently had their activity characterised also. Although both mutants promoted increased oligomerisation and *XBP-1* splicing, one causes a more invasive phenotype and appeared to be a cancer driving mutation (A414T) whilst the other promoted apoptosis by increasing degradation of miR-17 and appeared to prevent tumour formation (P336L). It's likely that P336L requires a mutant p53 background to aid cancer development, but this study highlights how two mutations that increase oligomerisation of IRE1 α lead to divergent outcomes (Lhomond et al., 2018). Therefore, a more in depth structural and functional characterisation of such cancer-associated mutations is required.

Table 1.1

Examples of cancer-associated mutations of IRE1 α with their location in the protein and observed outcomes.

Mutation	Cancer type	Area of protein	<i>In vivo/ vitro</i> outcome
N244S	Clear cell carcinoma (Greenman et al., 2007).	β -sandwich domain.	Unknown.
S296F	Cutaneous squamous cell carcinoma, metastatic melanoma (Sanborn et al., 2015; Pickering et al., 2014).	Conserved anti-parallel β -sheet.	Unknown.
P336L	Glioblastoma (Lhomond et al., 2018; Parsons et al., 2008).	Long loop (loop 2 region).	Increased oligomerisation. Increased RIDD activity, requires other mutations to cause cancer (Lhomond et al., 2018).
A414T	Glioblastoma (Lhomond et al., 2018).	Luminal linker region.	Increased oligomerisation. Increased <i>XBP-1</i> splicing, lower RIDD activity. Cancer

			driving mutation (Lhomond et al., 2018).
V418M	Parathyroid carcinoma (Greenman et al., 2007; Pandya et al., 2017).	Luminal linker region.	Unknown.
L474R	Adenocarcinoma (Greenman et al., 2007).	Cytoplasmic linker region.	Unchanged <i>XBP-1</i> splicing. Reduced apoptosis (Ghosh et al., 2014; Xue et al., 2011).
R635W	Adenocarcinoma (Greenman et al., 2007).	N-terminal kinase lobe.	Unchanged <i>XBP-1</i> splicing. Reduced apoptosis (Ghosh et al., 2014; Xue et al., 2011).
N700S	Germ-line variant (Greenman et al., 2007).	Middle of kinase domain.	Unknown.
S769F	Glioblastoma (Greenman et al., 2007).	Between kinase and endoribonuclease domain.	Reduced phosphorylation and apoptosis. Contradictory results on <i>XBP-1</i> splicing (Xue et al., 2011; Ghosh et al., 2014).
P830L	Serous carcinoma (Greenman et al., 2007).	Between kinase and endoribonuclease domain.	Reduced activity, instability of endoribonuclease domain (Xue et al., 2011; Ghosh et al., 2014).

To understand the effect of specific cancer-associated mutations on IRE1 α 's activation cascade we must first have an understanding of the mechanisms that drive activation. Groups have studied IRE1 α 's activation mechanism using cellular and *in vitro* assays and have managed to make connections between this data and structural models obtained. However, there are still discrepancies and unknowns in the activation pathways suggested. Although understanding IRE1 α 's activation mechanism can allow appreciation of how it is perturbed by cancer-associated mutations, study of the cancer-associated mutations can also elucidate aspects of the activation mechanism which are not fully understood. Described in the following sections is the body of work carried out to understand the activation pathway of IRE1 α . Due to long linker regions separating and structurally decoupling the protein's luminal and cytoplasmic domains, it is common for *in vitro* study to adopt a 'divide and conquer' approach and separate the two domains, thus the activation pathways described here have also been separated into the luminal and cytoplasmic domains.

1.4. Luminal domain activation

1.4.1. Luminal domain architecture

IRE1 α 's endoplasmic reticulum luminal domain is responsible for sensing endoplasmic reticulum stress, it responds to stress by dimerising and oligomerising upon an abundance of unfolded proteins in the endoplasmic reticulum. Here the luminal domain's architecture will first be described before information about its activation mechanism. The N-terminal globular portion of the domain precedes a ~60 residue long linker region tethering it to its endoplasmic reticulum transmembrane helix. The domain has a high proportion of unstructured regions, 29% as predicted by the PONDR webserver (Romero et al., 2001) and 35% unresolved in the solved crystal structure (Figure 1.4) (Zhou et al., 2006).

A dimerisation interface was identified through crystal contacts in the solved crystal structure and confirmed by mutational studies, dimerisation was found to be essential in activation of downstream activities of the cytoplasmic domain (Zhou et al., 2006). When dimerised the conformation includes an anti-parallel β -sheet floor which exists as a potential unfolded protein binding groove. This region is highly conserved between the yeast and human proteins, particularly in residues that face upwards out of the groove (shown in cyan in Figure 1.5). The groove is major histocompatibility complex (MHC)-like and mutational studies have shown that conserved residues in the groove are required for substrate binding, the importance of which will be described later (Credle et al., 2005; Zhou et al., 2006). However, it has been argued that this MHC-like groove is only important for the yeast protein as the space is too minimal to allow unfolded protein binding in the human crystal structure (Zhou et al., 2006). More recently, studies have provided evidence that human IRE1 α also uses this groove to bind unfolded proteins (Karagoz et al., 2017; Kono et al., 2017).

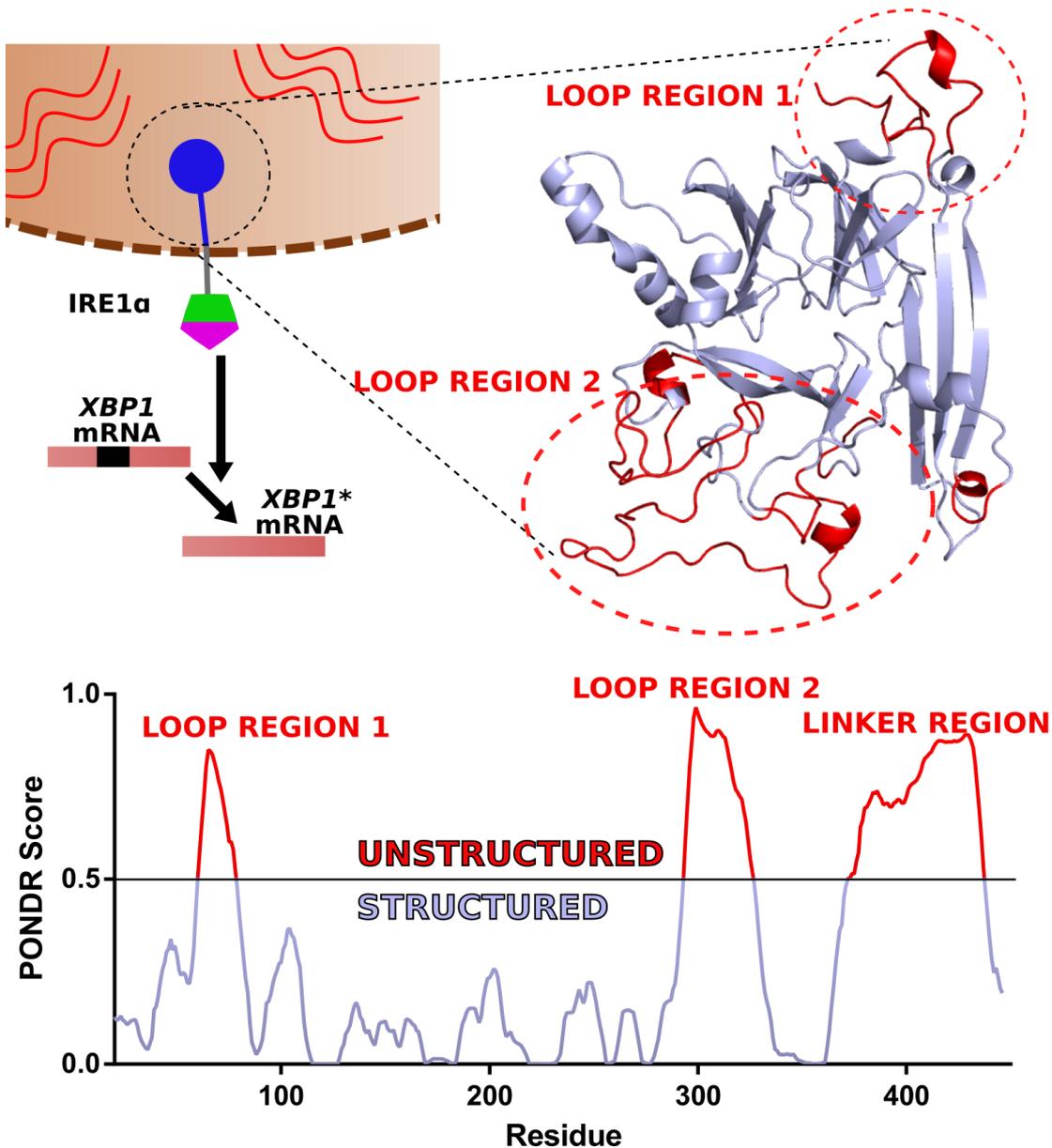


Figure 1.4

The resolved and unresolved residues of human IRE1 α 's luminal domain crystal structure (PDB: 2HZ6) (Zhou et al., 2006) with modelled loop regions through use of I-TASSER (Yang et al., 2015) (Figure prepared using PyMOL version 1.7) and a PONDNR webserver plot of regions predicted to be structured and unstructured (Romero et al., 2001). The unstructured regions on the PONDNR plot are from loop region 1, loop region 2 and the long transmembrane linker region, these are also not resolved in the crystal structure, as shown in red on the reconstructed model.

Other regions of importance identified include the β -sandwich and $\alpha\beta$ -helix motifs (Figure 1.5), these are suggested to propagate changes from unfolded protein binding to the MHC-like groove to the protein's oligomerisation interface, promoting a more active structure for oligomerisation (Karagoz et al., 2017). The oligomerisation interface itself is not conserved between the yeast and

human protein. In the yeast protein structure this region is solved in the crystal structure and has been confirmed by mutational studies (Credle et al., 2005). The human protein oligomerisation interface was identified through crosslinking studies and is suggested to consist of residues near to an unresolved region in the crystal structure (Figure 1.5). Four residues (359-362 WLLI) that were suggested to be important in the oligomerisation interface were mutated (359-362 WLLI to GSSG) and disfavoured the protein from forming oligomers and splicing *XBP-1* (Karagoz et al., 2017).

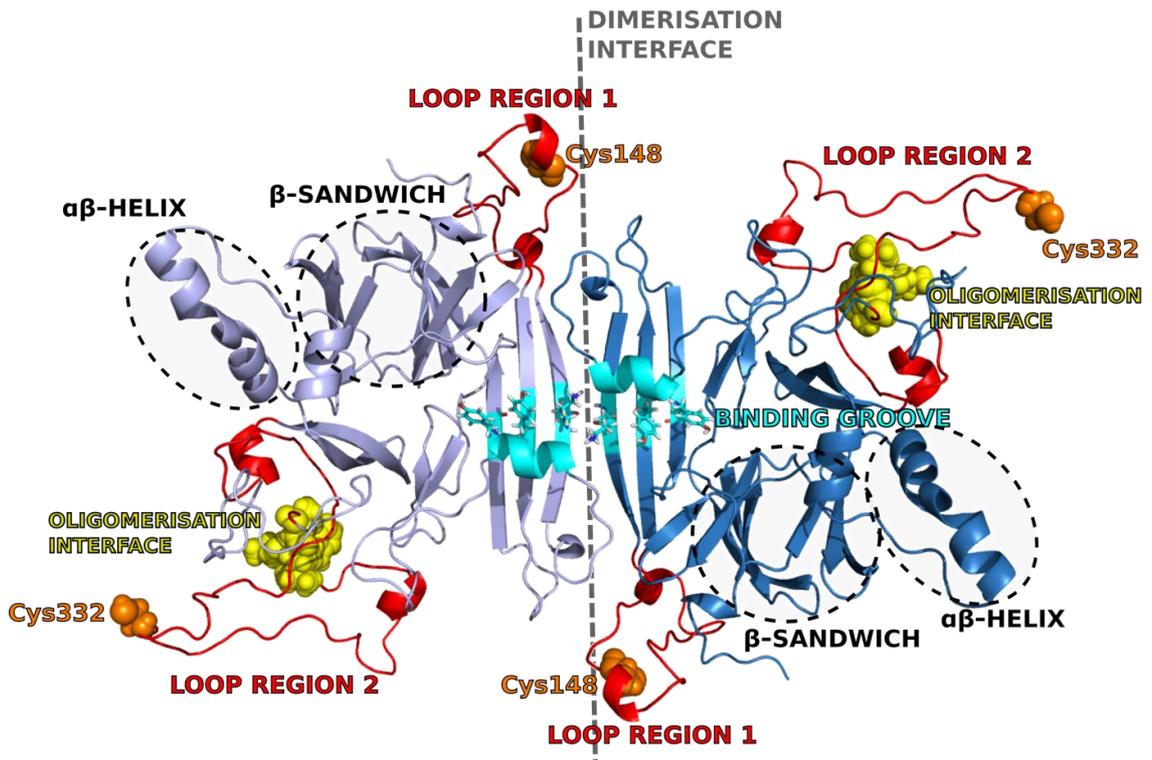


Figure 1.5

Reconstructed model of the solved IRE1 α luminal domain crystal structure annotated with regions of importance in activation, the linker region is not shown (PDB: 2HZ6) (Zhou et al., 2006; Yang et al., 2015; Karagoz et al., 2017). Figure prepared using PyMOL version 1.7)

Surrounding the oligomerisation interface is loop region 2, this is the longer loop region in the protein and consists of ~50 residues (Figure 1.5). The loop region isn't resolved in the crystal structure and includes cysteine 322, a smaller loop region (loop region 1, 22 residues) also contains a cysteine residue (Cys148). These cysteines are suggested to have a role in IRE1 α activation, however, they are not required for dimerisation (Liu et al., 2003; Eletto et al., 2014).

C-terminal to the oligomerisation interface is the beginning of the linker region. The luminal linker region is often removed in *in vitro* studies of the protein by a stop codon after residue 389

and was removed in the crystallisation conditions used (Zhou et al., 2006; Karagoz et al., 2017). However, the linker region's most C-terminal region and transmembrane helix have been implicated in a membrane stress sensing functionality of the protein (Halbleib et al., 2017; Kono et al., 2017), which causes clustering without the requirement of the globular portion of the luminal domain (Volmer et al., 2013), interactions between this mode of stress sensing and those discussed in the following sections are not fully understood. Additionally, previous studies of both yeast and human IRE1 α suggested that the linker region may be bound by a molecular chaperone (binding immunoglobulin protein (BiP)) which acts to repress IRE1 α 's activity (Kimata et al., 2004; Oikawa et al., 2009). However, more recent studies have suggested that this molecular chaperone interacts with IRE1 α 's 'core domain', without this linker region present (Carrara et al., 2015; Amin-Wetzel et al., 2017). The molecular chaperone, BiP, and its role in luminal domain activation will be discussed in more detail in the next section.

The crystal structure of human IRE1 α in combination with functional and mutational assays has therefore given some understanding about the luminal domain's function (Figure 1.5), however, there are still conflicting ideas about how the protein is activated structurally and functionally. The conflicting ideas in the field can be summarised into two different models, the first of which involves the molecular chaperone, BiP, binding and repressing IRE1 α until it dissociates during times of stress and the other involves IRE1 α 's luminal domain directly binding to unfolded proteins via its MHC-like groove region (Preissler and Ron, 2019). The aspects of luminal domain activation are discussed in the following sections.

1.4.2. BiP repression model of IRE1 α luminal domain activation

There's a large body of evidence for IRE1 α being bound by the molecular chaperone, BiP, under non-stressed conditions with less BiP bound during stressed conditions, although there's conflicting evidence for where this binding site is on IRE1 α , as previously mentioned (Bertolotti et al., 2000; Carrara et al., 2015; Oikawa et al., 2009; Pincus et al., 2010). Before describing the role BiP plays in luminal domain activation it is first necessary to understand its architecture and function. BiP is the most abundant molecular chaperone in humans and is a 70kDa multidomain protein. It contains a nucleotide-binding domain with ATPase activity connected via a short linker to a substrate binding domain which contains a β -sandwich and a 'lid' region that bind to unfolded regions of proteins (Figure 1.6). When ATP is bound to the nucleotide binding domain the two domains of BiP dock together and the protein has a low affinity for unfolded protein substrate (domain-docked). When ATP is hydrolysed to ADP, BiP adopts a domain-undocked conformation, here its domains act separately and the substrate binding domain has a high affinity

for unfolded proteins (Mayer, 2013; Wieteska et al., 2017). In this way BiP acts as a molecular chaperone to proteins in the endoplasmic reticulum, binding to and retaining misfolded proteins in the endoplasmic reticulum to promote their correct folding or degradation and chaperoning correctly folded proteins to the next stage of the secretion pathway (Hendershot, 2004).

There's evidence that both domains of BiP are able to bind IRE1 α ; truncations of BiP have suggested that BiP's substrate binding domain is required and also that this interaction requires ATP hydrolysis, suggesting BiP's canonical cycle, as shown in Figure 1.6, is required for binding (Amin-Wetzel et al., 2017; Liu et al., 2003). However, research has suggested that BiP can also bind directly to IRE1 α through its nucleotide binding domain in a nucleotide independent manner, termed the non-canonical interaction. Research is ongoing into the interaction's effect on the unfolded protein response (Carrara et al., 2015; Kopp et al., 2018).

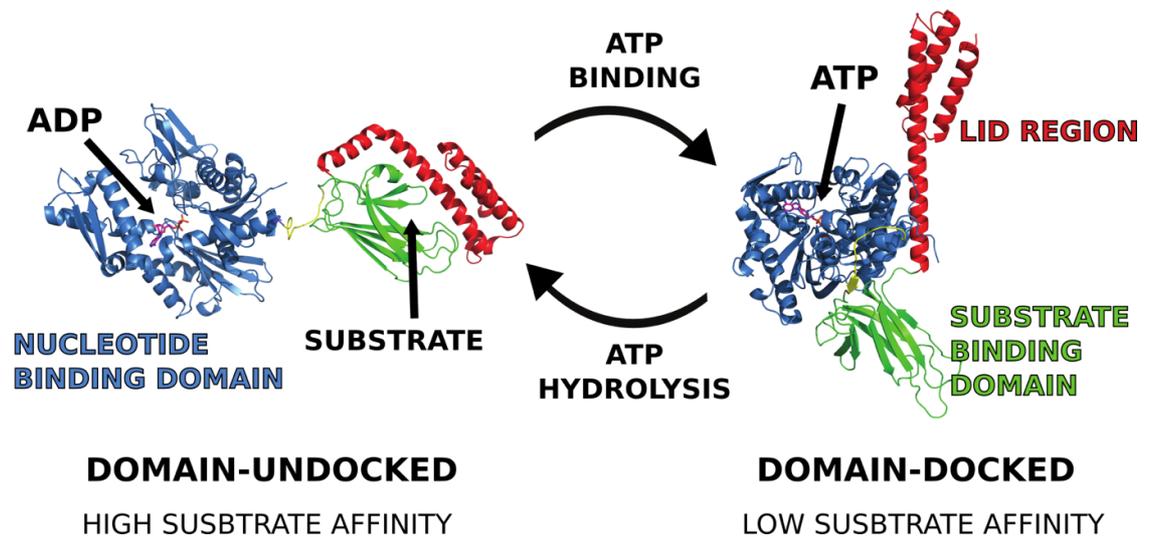


Figure 1.6

BiP's ATP dependent chaperone cycle. With ATP bound BiP has low substrate affinity and its domains are docked together. Upon ATP hydrolysis the domains of BiP undock and BiP has a high substrate affinity (Wieteska et al., 2017). Figure prepared using PyMOL version 1.7.

The BiP repression model of IRE1 α activation involves BiP binding to IRE1 α and inhibiting formation of dimers and thus an increase in the local concentration of the cytoplasmic domain, autophosphorylation and activation. However, it is not known which form of IRE1 α BiP binds to, earlier studies suggested monomers but more recently BiP has been shown to bind to dimers and its ability to bind oligomers has not yet been investigated (Amin-Wetzel et al., 2017; Bertolotti et al., 2000). The endoplasmic reticulum-localised DnaJ 4 (ERdj4) protein was recently suggested

to bind IRE1 α 's luminal domain during non-stressed conditions and promote BiP binding to IRE1 α , therefore repressing activity by preventing oligomerisation. This is in addition to the ERdj4's usual J-domain activity of stimulating BiP's nucleotide binding domain to hydrolyse ATP (Amin-Wetzel et al., 2017). In all cases it appears that BiP binding to IRE1 α represses activity and opposes IRE1 α dimerisation and subsequent clustering (Amin-Wetzel et al., 2017; Carrara et al., 2015; Bertolotti et al., 2000). BiP binds preferentially to unfolded proteins and during times of stress, unfolded proteins are abundant. Therefore, the BiP repression model suggests that upon binding to unfolded proteins, BiP dissociates from IRE1 α , which causes the luminal domain to dimerise and oligomerise in response to endoplasmic reticulum stress to promote a response, represented in Figure 1.7 (Carrara et al., 2015; Ron and Walter, 2007; Preissler and Ron, 2019).

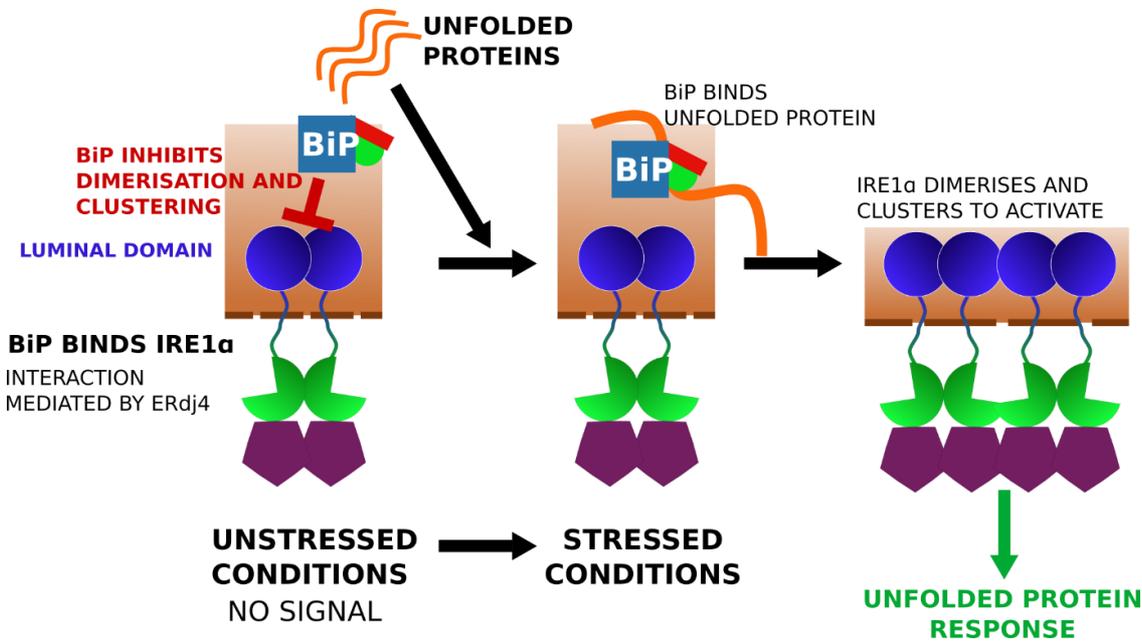


Figure 1.7

The BiP repression model of IRE1 α activation. BiP binds to the luminal domain during non-stressed conditions and opposes dimerisation. Upon endoplasmic reticulum stress there is an increase in unfolded proteins in the endoplasmic reticulum. BiP dissociates from IRE1 α to bind the unfolded proteins. IRE1 α 's luminal domain is now able to dimerise and cluster to promote activation of the cytoplasmic domain and the unfolded protein response (Amin-Wetzel et al., 2017; Bertolotti et al., 2000; Carrara et al., 2015).

Some of the studies supporting the BiP binding repression model suggest that unfolded protein substrate binding to IRE1 α directly can cause IRE1 α oligomerisation, but that BiP release from IRE1 α to bind the unfolded proteins is the initiating step in IRE1 α activation (Carrara et al., 2015; Amin-Wetzel et al., 2017; Pincus et al., 2010).

1.4.3. Unfolded protein binding model

The unfolded protein binding models don't directly refute BiP binding for repression of IRE1 α and BiP may affect the equilibrium of IRE1 α binding to unfolded proteins. However, they suggest that direct binding of unfolded proteins to IRE1 α is the principal cause of activation, as it promotes oligomerisation, crucial for maximal activity and the protein's observed clustering (Karagoz et al., 2017; Gardner and Walter, 2011; Li, H. et al., 2010). IRE1 α is suggested to adopt an oligomer-active state in solution transiently without the inclusion of unfolded proteins, allowing it to form tetramers at high concentrations (Karagoz et al., 2017). The inclusion of unfolded proteins promotes higher order species by stabilising this conformation through binding to IRE1 α . This binding causes a cascade of structural rearrangements leading to the oligomer-active state being stabilised (Figure 1.8) (Karagoz et al., 2017). However, there is no evidence for a pathway to terminate IRE1 α oligomers and stop the unfolded protein response signal other than the equilibrium of unfolded proteins, free BiP and unfolded protein bound BiP changing after stress is relieved.

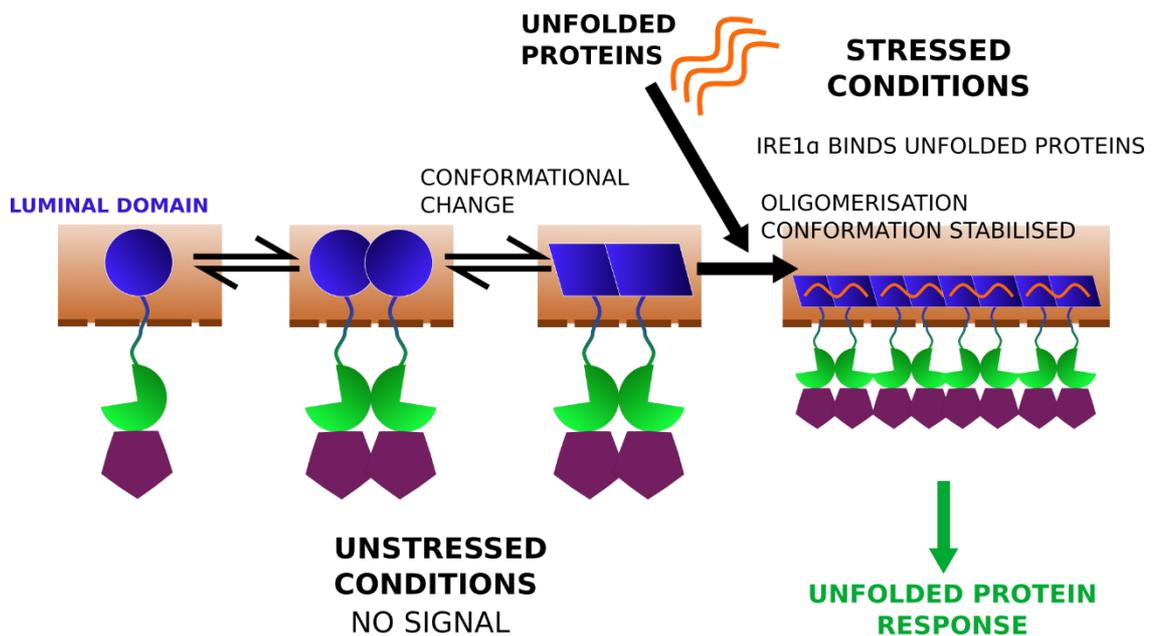


Figure 1.8

The unfolded protein binding model of IRE1 α activation. IRE1 α exists as an equilibrium between monomeric and dimeric protein, it transiently adopts the oligomerisation active state. Upon endoplasmic reticulum stress, IRE1 α binds unfolded proteins directly, stabilising the oligomerisation complex allowing the luminal domain to oligomerise and promote activation of the cytoplasmic domain (Karagoz et al., 2017).

1.4.4. Disulphide formation in IRE1 α activation

Another arm of IRE1 α activation is that the protein can form disulphide bonds within multimers, potentially stabilising the more active state (Figure 1.9). Studies have suggested that IRE1 α forms disulphide bonds with its conserved cysteine residues Cys148 and Cys332, and that IRE1 α interacts with protein disulphide isomerase A6 (PDIA6) (Liu et al., 2003; Eletto et al., 2014). PDI's in the endoplasmic reticulum act to reduce covalent disulphide bonds, meaning to break apart the thiol groups from the cysteine residues involved in the bond, in this manner PDIs promote correct folding of proteins in the endoplasmic reticulum (Wilkinson and Gilbert, 2004). PDIA6 has been shown to interact with IRE1 α 's Cys148, reducing its disulphide bond and terminating IRE1 α 's active state. Prevention of this interaction causes higher responsiveness to stress (Eletto et al., 2014; Groenendyk et al., 2014). Recent data from our group (carried out by Sam Dawes) suggests that IRE1 α forms disulphide bonds with Cys148 and 332 and the previously mentioned non-canonical interaction of BiP with IRE1 α 's luminal domain (independent of nucleotide binding) slows this formation and preformed disulphide bonds prevent the non-canonical BiP interaction. This possibly describes one facet for stabilisation and suppression of the IRE1 α signal.

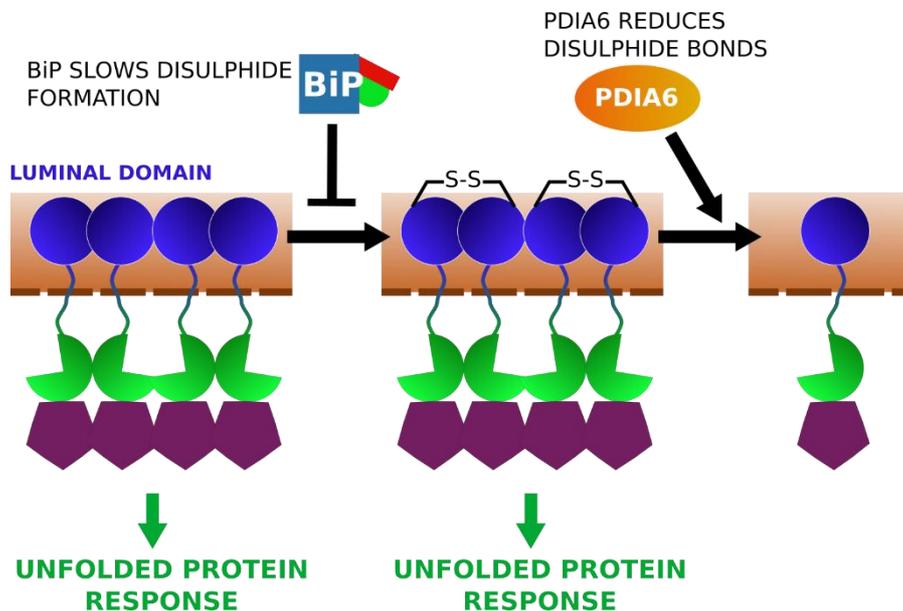


Figure 1.9

Active IRE1 α forms disulphide bonds over time. Formation of these bond appears to be in competition with non-canonical BiP binding (work carried out by Sam Dawes). After stress is relieved, PDIA6 reduces the disulphide bonds and terminates IRE1 α 's signal, thus allowing BiP to bind IRE1 α again (Groenendyk et al., 2014; Eletto et al., 2014).

1.4.5. Discrepancies in the models

It is clear that activation of IRE1 α involves multiple mechanisms in a complex process. From the information in the literature we can formulate a crude model for the luminal domain' activation (Figure 1.10). However, there are many questions that remain; firstly, whether BiP binds to the monomeric or dimeric form of IRE1 α to repress its activation, and if BiP has any interaction with the oligomeric form. It is also not clear whether IRE1 α clustering is driven by BiP dissociating from the luminal domain or whether the luminal domain binding directly to unfolded proteins is sufficient to form this response. Although the interaction of the luminal domain with unfolded proteins has been initially characterised with some structural detail (Karagoz et al., 2017), the process of oligomerisation and how large the oligomers can become is unknown. In addition to this the mechanism for termination of the signal and dissipation of the clustered IRE1 α molecules is not understood. Additionally, how disulphide bond formation is associated with this hypothesised pathway is not understood and what physiological outcome this process has. Lastly it is unknown how or if the luminal domain's activation can differentially promote the adaptive and apoptotic response, which also relates to the question of how cancer-associated mutations interfere with IRE1 α 's activation.

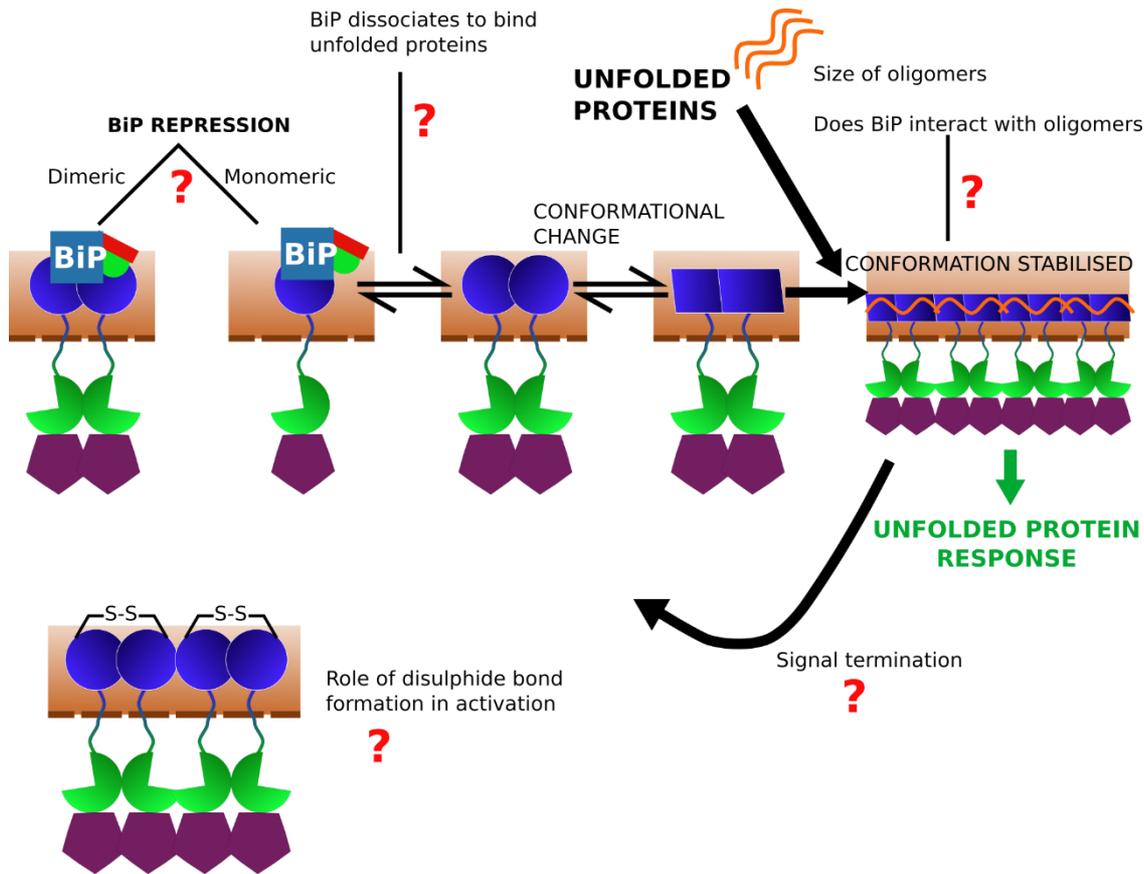


Figure 1.10

Combining information in the literature creates a hypothesised mechanism for activation of the luminal domain, however, there are still many discrepancies present. BiP represses dimer formation of IRE1 α , but it is unknown which form of IRE1 α it binds to. The dimeric protein undergoes conformational rearrangements as observed previously (Karagoz et al., 2017). Peptide binding stabilises oligomers of IRE1 α , however, it is unknown whether BiP interacts with these oligomers and how large the oligomers become. It is unclear how the formed oligomers dissipate after endoplasmic reticulum stress is relieved and also the role of disulphide bond formation in this activation pathway is not fully understood.

The luminal domain's activation cascade is therefore a complex and multi-step process, although questions remain over the precise mechanism, luminal domain activation causes clustering of the domain that causes an increase in the local concentration of IRE1 α 's cytoplasmic domain, thus promoting a cellular response to stress. The cytoplasmic domain undergoes autophosphorylation, dimerisation and oligomerisation in a complex multi-step activation pathway, in the following section the activation mechanism for the cytoplasmic domain in response to an increase in local concentration will be discussed.

1.5. IRE1 α 's cytoplasmic domain

1.5.1. Cytoplasmic domain architecture

The cytoplasmic domain is the larger of IRE1 α 's two domains and includes a long ~100 residue linker region from the transmembrane helix to the globular portion of the domain. The globular portion of the domain consists of a novel kinase extension endoribonuclease (KEN) fold including an activation loop that is common in autophosphorylation function (Figure 1.11) (Lee, K.P. et al., 2008). The protein's kinase, autophosphorylation and endoribonuclease activities are all required for correct cellular signalling activities.

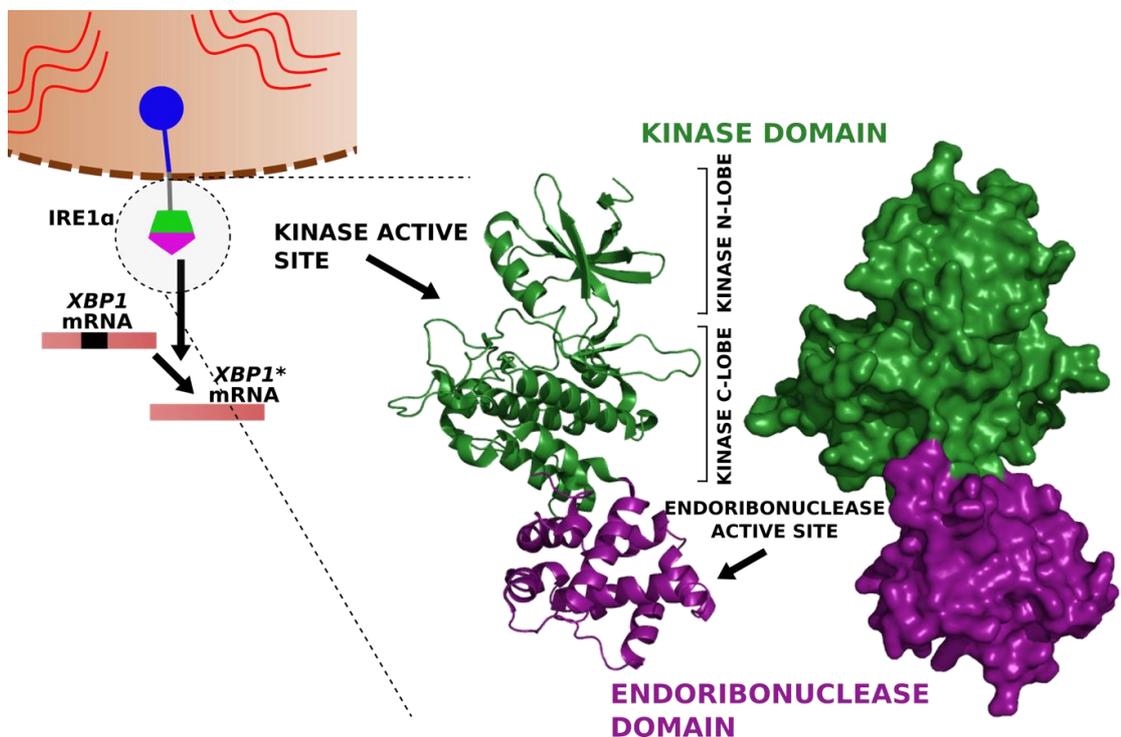


Figure 1.11

Figure displaying IRE1 α 's cytoplasmic domain architecture, the domain consists of a kinase extension endoribonuclease fold, with the kinase and endoribonuclease active sites shown on the solved monomeric structure of the protein (PDB: 4U6R) (Harrington et al., 2015). Figure prepared using PyMOL version 1.7.

There are currently 20 X-ray crystallography models solved for the cytoplasmic domain (Table 1.2). The protein has been captured in various conformers, including monomeric, dimeric and as an oligomer. Although these different forms have the same general architecture, there are more subtle structural rearrangements in the protein's kinase and endoribonuclease domains that appear to dictate the overall conformation and activity. The various structures and conformations solved

for the cytoplasmic domain provide an understanding about the protein's activation pathway, and how the protein's kinase and endoribonuclease activities influence this. Firstly, the architecture of the cytoplasmic domain's kinase domain is presented before the hypothesised activation pathway of the cytoplasmic domain is described, a summary of this pathway is presented in Figure 1.12. The cytoplasmic domain is considered to first exist as a monomer before adopting a dimeric conformation that promotes autophosphorylation, after autophosphorylation it is thought to adopt the 'back-to-back' conformation which promotes endoribonuclease activity, this conformation then oligomerises in repeating units to achieve maximal endoribonuclease activity.

Table 1.2

The crystal structures solved for IRE1 α 's cytoplasmic domain. Detailed are the states of the elements of the protein's kinase domain.

PDB ID	Organism	Molecule bound	αC-Helix	DFG-motif	R-spine	Reference
MONOMER						
4U6R	Human	Sulphamide inhibitor	Inactive	In	Disrupted	(Harrington et al., 2015)
6HV0	Human	Compound 33	N/A	Out (disrupted)	Disrupted	(Colombano et al., 2019)
6HX1	Human	Compound 2	N/A	Out (disrupted)	Disrupted (Only His686 aligned)	(Colombano et al., 2019)
FACE-TO-FACE DIMER						
3P23	Human	ADP	Inactive	In	Disrupted	(Ali et al., 2011)
4YZD	Human	ADP	Active (Slightly disrupted)	In (Slightly disrupted)	Heavily disrupted (only His686 aligned)	(Concha et al., 2015)
4PL3	Murine	MKC989 inhibitor + ADP	Inactive	In (Slightly disrupted)	Heavily disrupted (Only His686 aligned)	(Sanchez et al., 2014)

4PL4	Murine	OICR464 inhibitor + ADP	Inactive	In (Slightly Disrupted)	Heavily disrupted (only His686 aligned)	(Sanchez et al., 2014)
4PL5	Murine	OICR573 inhibitor + ADP	Inactive	In (Slightly Disrupted)	Heavily Disrupted (only His686 aligned)	(Sanchez et al., 2014)
CRYSTAL CONTACTS						
3LJ1	Yeast	Cdk1/2 Inhibitor III	Active (Slight disrupted)	In	Aligned	(Wiseman et al., 2010)
3LJ2	Yeast	JAK1 Inhibitor	Active (Slightly disrupted)	In	Aligned	(Wiseman et al., 2010)
BACK-TO-BACK DIMER						
4YZ9	Human	GSK2850 163A	Active	Out (Disrupted)	Heavily disrupted, Leu616 and Phe712 moved.	(Concha et al., 2015)
4Z7G	Human	Apo	Active	In (Slightly disrupted)	Phe712 misaligned	(Joshi et al., 2015)
5HGI	Human	Apo (Caesium ions)	Active	In	Aligned	(Feldman et al., 2016)
4YZC	Human	Staurospo rine	Active (not fully resolved)	In	Aligned	(Concha et al., 2015)
4Z7H	Human	Imidazop yridine compoun d 3	Active	In	Aligned	(Joshi et al., 2015)
3LJ0	Yeast	ADP and Quercetin	Active	In	Aligned	(Wiseman et al., 2010)

2RIO	Yeast	ADP	Active	In	Aligned	(Lee, K.P. et al., 2008)
OLIGOMER						
3FBV	Yeast	Oligonucl eotide	Active	In	Aligned	(Korenyk h, A.V. et al., 2009)
3SDJ	Yeast	Oligonucl eotide	Active	In	Aligned	(Korenyk h, A.V. et al., 2011b)
3SDM	Yeast	Oligonucl eotide	Active	In	Aligned	(Korenyk h, A.V. et al., 2011a)

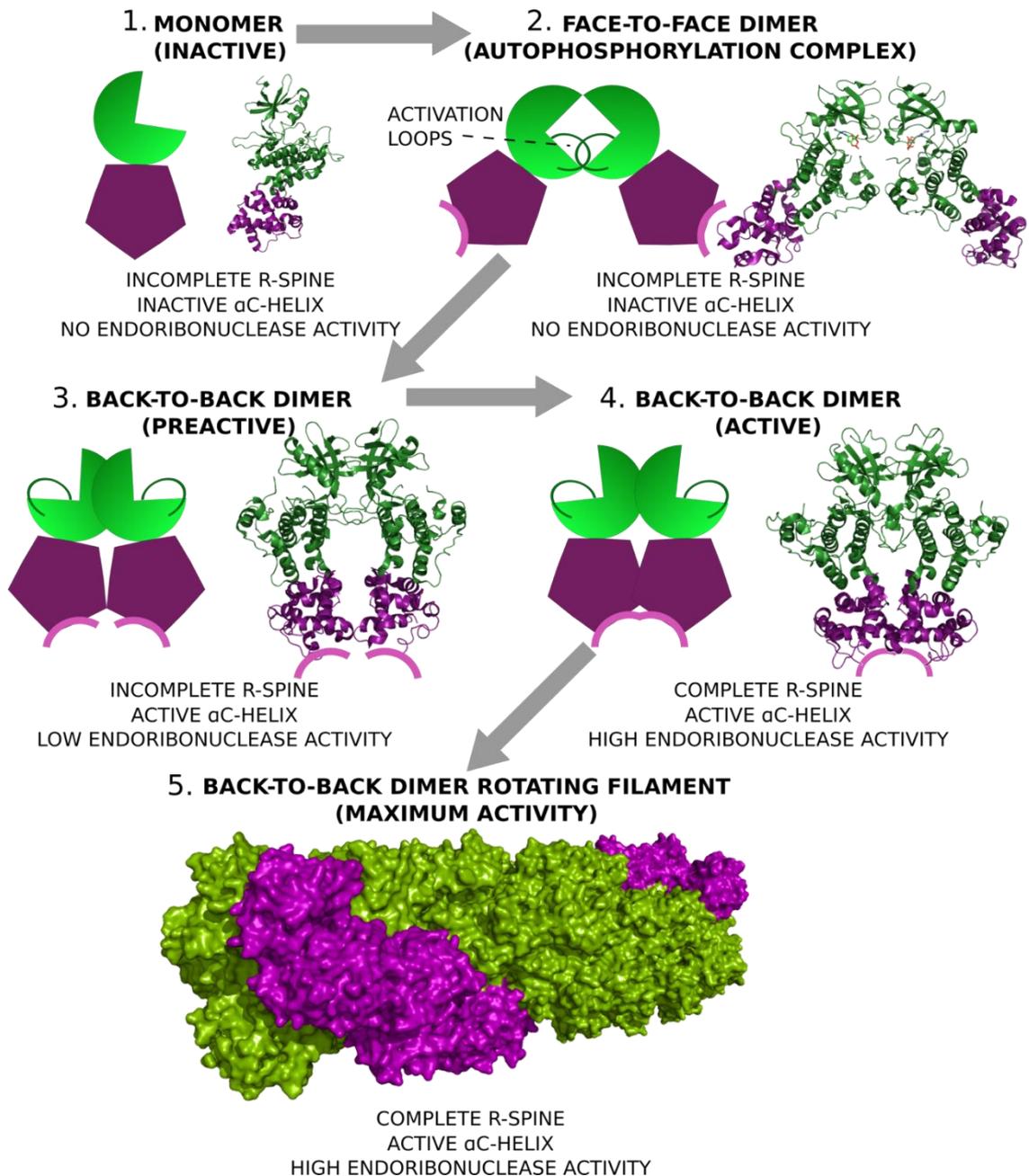


Figure 1.12

Schematic for the different conformations of the cytoplasmic domain solved by X-ray crystallography. Detailed is the state of the kinase domain's R-spine, α C-helix and the endoribonuclease activity of the complex. 1. PDB: 4U6R. 2. PDB 3P23. 3. PDB: 4Z7G. 4. PDB: 2RIO. 5. PDB: 3FBV (Harrington et al., 2015; Ali et al., 2011; Joshi et al., 2015; Concha et al., 2015; Korennykh, A.V. et al., 2009; Feldman et al., 2016). Figure prepared using PyMOL version 1.7.

In order to understand the conformations adopted by the cytoplasmic domain, it is essential to first gain an understanding about the architecture of the protein's kinase domain and how it is regulated. The protein's kinase domain contains a N- and C-terminal lobe surrounding the ATP

binding site, activation of the domain relies on three motifs present. Firstly, the DFG motif (consisting of Asp711, Phe712 and Gly713; Figure 1.13B) contains Asp711 which is critical for coordination of the ATP and magnesium molecules and therefore kinase activity (Figure 1.13C). The position of Asp711 is stabilised by Gly713, where an active conformation is referred to as a DFG-in conformation and a DFG-out conformation is inactive due to Asp711 moving out of the active site (Figure 1.13C and D). The Phe712 of the DFG motif allows coordination of the second motif, the α C-helix (Figure 1.13A) and a salt bridge interaction between Lys599 and Glu612 to stabilise this conformation (Figure 1.13B). Movement of the α C-helix into its active position promotes formation of the third motif, a hydrophobic regulatory 'R-spine' in the protein (consisting of residues tyrosine 628, leucine 616, phenylalanine 712 (from the DFG motif) and histidine 686) which is indicative of an activated kinase domain (Figure 1.13B-D). Phosphorylation of the activation loop stabilises and promotes the active formation of the kinase domain also (Kornev et al., 2006; Joshi et al., 2015). The state of these structural motifs is important in dictating IRE1 α conformation and activity. By using the conformations identified in the numerous crystal structures, the hypothesised activation pathway for the cytoplasmic domain will be described in the following sections (Figure 1.12).

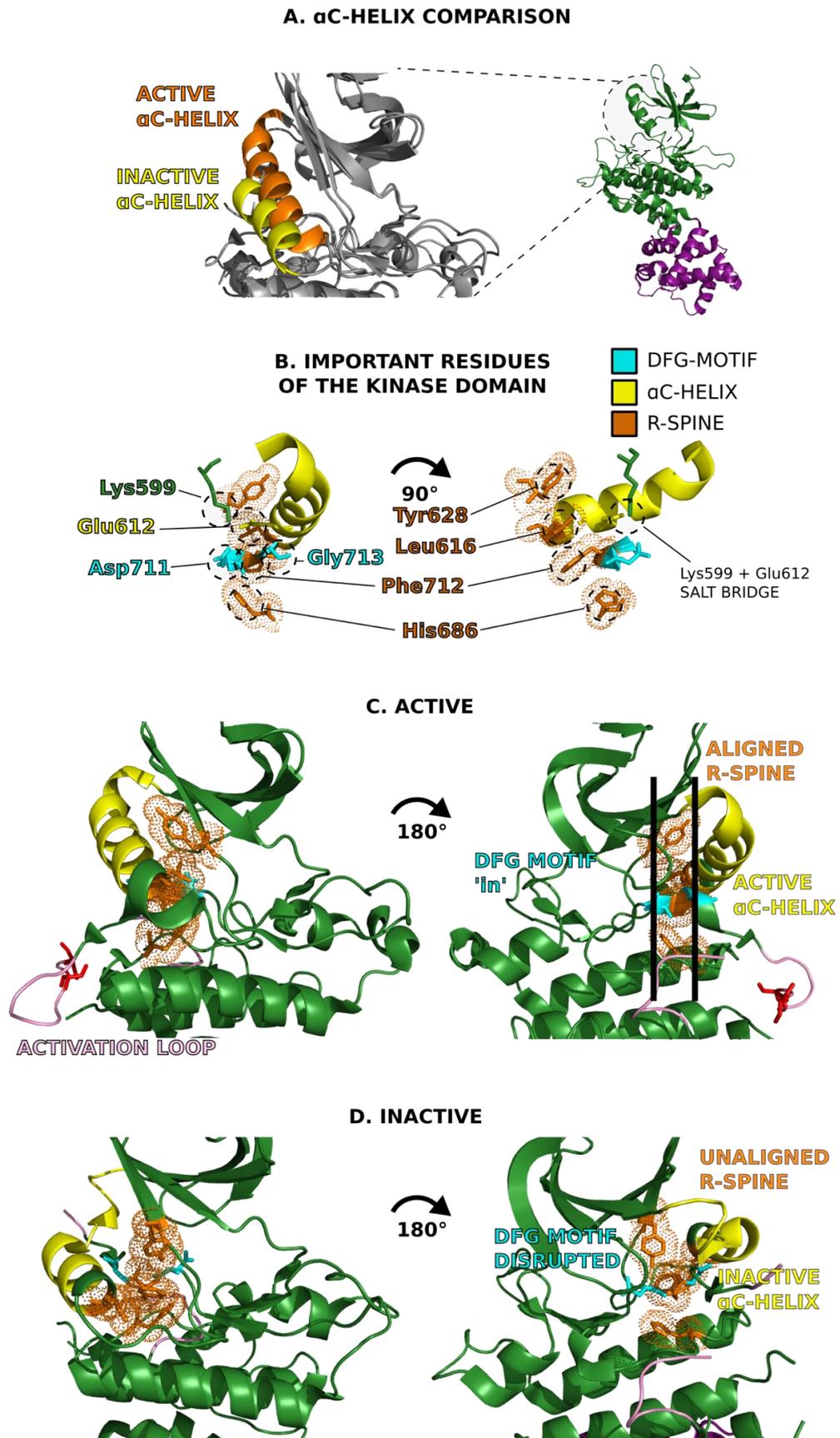


Figure 1.13

Crystal structures of the active and inactive kinase domains of IRE1 α 's cytoplasmic domain. Important motifs in kinase domain function are shown. A. Comparison of the α C-helix in active and inactive states. The inactive conformation causes steric clashes when forming the back-to-back dimer (Active PDB: 4Z7H, inactive PDB: 3P23). B. Residues of importance for kinase domain functioning. Residues of the R-spine, DFG-motif and Lys599 and Glu612 of the α C-helix which form a salt bridge (PDB: 4Z7H). C. The active structure (PDB: 4Z7H) has an aligned R-spine shown in orange coordinated by an active DFG motif and leading to an active α C-helix. D. The inactive kinase domain (PDB: 3P23) shows a disrupted R-spine unable to coordinate the α C-helix into an active conformation (Ali et al., 2011; Joshi et al., 2015). Figure prepared using PyMOL version 1.7.

1.5.2. Monomer

The cytoplasmic domain is thought to be inactive in its monomeric form, the human protein has been crystallised and its structure solved as a monomer with a potent kinase inhibitor compound bound (PDB: 4U6R), which showed a decrease in *XBP-1* splicing in cell culture (Harrington et al., 2015). *In vivo* and *in vitro* studies have also supported the idea of an inactive monomer, when comparing UPR responses in cells, GFP-linked IRE1 α clustered when active and dispersed when inactive, also increasing concentrations of the protein *in vitro* promote endoribonuclease activity suggesting dimerisation/oligomerisation for activation (Li, H. et al., 2010; Itzhak et al., 2014). This agrees with the luminal domain having to first oligomerise to increase the local concentration of the cytoplasmic domain to allow it to form active multimers.

1.5.3. Face-to-face dimer

Due to IRE1 α 's luminal domain clustering following a stress signal, the local concentration of the cytoplasmic domain increases. The unphosphorylated cytoplasmic domain is thought to first form a dimer that positions the adjoining monomers' kinase regions proximally to one another's activation loops, promoting trans-autophosphorylation; this is referred to as the 'face-to-face' dimer (Figure 1.14) (Ali et al., 2011). Phosphorylation significantly increases IRE1 α 's endoribonuclease activity and the specific phosphorylation sites have been identified to be three serines in the activation loop of the kinase domain. The phosphorylation of different serine residues alter the strength of the endoribonuclease activity of IRE1 α (Prischi et al., 2014).

The face-to-face conformation has been captured in a crystal structure of human IRE1 α bound to ADP and Mg²⁺ (PDB: 3P23). Mutations at key interaction sites in this orientation (Q636A and

F637A) cause a loss of autophosphorylation, suggesting the face-to-face dimer's role in this process (Figure 1.14) (Ali et al., 2011). Further crystal structures of IRE1 α 's face-to-face conformation were solved with murine IRE1 α (PDB: 4PL3, 4PL4 and 4PL5), which closely resemble the human conformation (RMSD = 1.28Å) but interestingly were phosphorylated, suggesting that the conformation may also exist before adoption of the next orientation in the proposed activation mechanism (the back-to-back dimer) (Sanches et al., 2014).

The face-to-face dimer represents an early stage in the activation mechanism of IRE1 α which allows for its autophosphorylation but does not facilitate endoribonuclease activity due to distal positioning of the endoribonuclease domains. Interestingly the kinase domain in the human face-to-face structure is in an inactive conformation (misaligned R-spine); which may be a result of ADP being bound or due to crystallisation conditions (Kornev et al., 2006; Ali et al., 2011; Sanches et al., 2014).

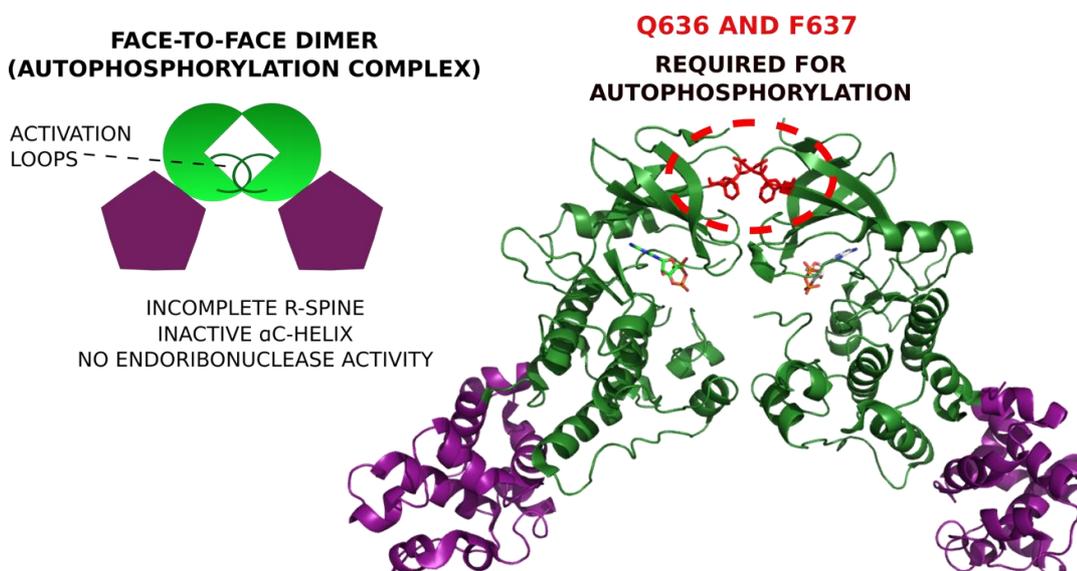


Figure 1.14

The face-to-face dimer autophosphorylation complex (PDB: 3P23). The face-to-face dimer complex was solved with ADP and Mg²⁺ bound. Important residues for this interaction are shown in red. Mutation of these residues to alanine prevent autophosphorylation of the protein (Ali et al., 2011). Figure prepared using PyMOL version 1.7.

1.5.4. Back-to-back dimer

IRE1 α 's back-to-back conformation has the two monomers rotated so that their kinase active sites now face away from one another, the endoribonuclease domains are positioned closely and allow

XBP-1 splicing (Figure 1.15) (Joshi et al., 2015). The crystal structures solved in this conformation have more active kinase domain arrangements, although there is variation between the structures. The kinase domain α C-helix is positioned differently than in the face-to-face structure; this positioning allows the back-to-back interface to form without steric hinderance (Figure 1.13A, 1.15) (Harrington et al., 2015; Ali et al., 2011; Joshi et al., 2015).

Mutations to residues in the back-to-back dimer interface inhibit the endoribonuclease but not kinase activity of the protein, further confirming the physiological importance for the back-to-back conformation in mRNA splicing and the conformation not being required for kinase activity (Sanches et al., 2014; Lee, K.P. et al., 2008). Throughout the crystallography structures solved for the back-to-back conformations there are numerous subtle differences such as rearrangements of the kinase domain, which appear to have effects on the endoribonuclease domain. A closer proximity of the endoribonuclease domains is thought to promote higher activity; thus the activation state of the kinase domain influencing the activity of the endoribonuclease domain is a common theme throughout the different structures solved (Concha et al., 2015; Feldman et al., 2016). Although many of the structures solved are influenced by unphysiological conditions and compounds bound, they can provide a sequential framework for this segment of IRE1 α activation.

The structure that can be considered as the least active of the back-to-back conformations is IRE1 α in apo-form (PDB: 4Z7G). As solved structures with ADP and Mg²⁺ bound form the face-to-face dimer, it is possible that ADP must dissociate from the protein to allow rearrangement of the α C-helix, removing steric hinderances that would prevent the back-to-back conformation. The apo structure and subsequent back-to-back structures have active kinase domains and a salt bridge formed between Lys599 and Glu612 stabilising the 'DFG' motif, and an active α C-helix orientation (Figure 1.13A-C). The apo form of the protein has a near fully-formed kinase domain R-spine and endoribonuclease domains that are positioned closer in space than in the face-to-face structures, but not close enough for high endoribonuclease activity (Joshi et al., 2015). Another human back-to-back structure was solved in apo state but with two caesium ions stabilising an active DFG motif and α C-helix conformation (PDB: 5HGI) (Feldman et al., 2016). The monomers are fully aligned and parallel in the construct which, similar to fully activated yeast structures solved (Korennykh, A.V. et al., 2009). However, the construct of IRE1 α used was fully dephosphorylated and therefore lacked further stabilisation of the kinase domain components causing minimal endoribonuclease activity when analysed.

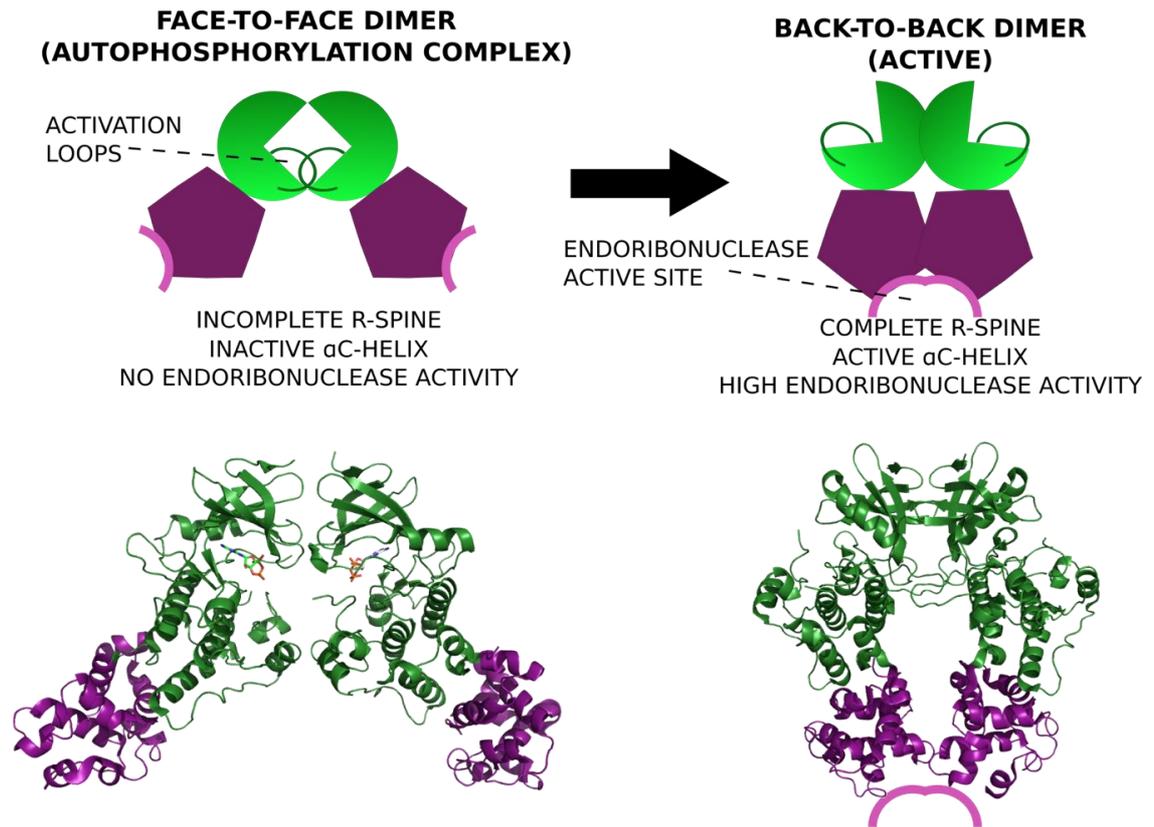


Figure 1.15

The face-to-face dimer is the autophosphorylation complex, autophosphorylation promotes higher endoribonuclease activity in the back-to-back dimer (PDB: 3P23, 4Z7H). The IRE1 α monomers rotate 180°C to face their kinase active sites away from each other in the back-to-back dimer (Ali et al., 2011; Joshi et al., 2015). Figure prepared using PyMOL version 1.7.

With the kinase and endoribonuclease inhibitory compound, GSK2850163, bound, a back-to-back crystal structure was solved (PDB: 4YZ9). This represents an inhibited structure of the protein and has a heavily disrupted R-spine and its DFG motif in an inactive conformation. Inclusion of this inhibitory compound mimics the apo state protein in activity and the structure also has the endoribonuclease domains positioned away from each other. Due to the inclusion of a synthetic inhibitor this structure is unlikely to be closer to the physiological state of IRE1 α immediately post-face-to-face conformation than the apo form (Joshi et al., 2015; Concha et al., 2015).

The crystals used for the first apo state structure were soaked with the imidazopyridine kinase inhibitor; compound 3 (PDB: 4Z7H). Likely due to the technique of soaking the solved structure is similar to the apo state. However, in this structure the kinase domain R-spine is fully formed, and the endoribonuclease domains are closer together, demonstrating the correlation between kinase and endoribonuclease domain conformations. Therefore, compound 3 successfully inhibits

kinase activity but promotes an active kinase structure to promote endoribonuclease activity (Joshi et al., 2015).

A similarly active human structure was solved bound to the kinase inhibitor staurosporine (PDB: 4YZC). Here the endoribonuclease domains are positioned more closely together than in previous structures with an active DFG motif position consolidating the kinase domain R-spine. *In vitro* assays show that the compound increases endoribonuclease activity of unphosphorylated protein; suggesting that the compound induces an active conformation in the protein, bypassing phosphorylation (Concha et al., 2015).

The structure of yeast IRE1 α was also solved in a back-to-back dimer (PDB: 2RIO), albeit with a 24-residue deletion to aid crystallisation (Lee, K.P. et al., 2008). The yeast protein has been extensively studied and is believed to behave in a similar way to the human protein, with very similar activity and structural motifs (Korennykh, A. and Walter, 2012). Although the structure is not bound to a synthetic compound, it is crystallised with ADP and Mg²⁺ which promoted formation of the face-to-face dimer in human IRE1 α (Ali et al., 2011). This conformation is considered to be a representation of the most active state of the back-to-back dimer with an active kinase domain, the expected activation loop phosphorylation marks and endoribonuclease domains positioned closely together with functional assays confirming its high endoribonuclease activity (Prischi et al., 2014; Korennykh, A.V. et al., 2009).

Through the numerous crystal structures and knowledge of homologues, the kinase domain of IRE1 α has been well defined. However, the endoribonuclease domain has much less known structurally, especially in terms of changing from conventional *XBP-1* splicing to promiscuous RIDD activity. The back-to-back dimer is required for *XBP-1* cleavage, however, yeast protein studies suggest only one monomer cleaves mRNA and the other's function is to orientate the substrate (Lee, K.P. et al., 2008). Important residues (residues 900-916) and motifs for the mRNA interface have been identified (Figure 1.16), including the helix-loop element (HLE), which is stabilised in more active oligomeric complexes. This region is conserved in the human protein, where the HLE was shown to have increased exposure upon binding of an endoribonuclease activator (Korennykh, A.V. et al., 2011b; Concha et al., 2015; Dong, B. et al., 2001).

XBP-1 and RIDD substrates do not appear to compete for the endoribonuclease domain's active site, and in the yeast protein some residues specific for only *XBP-1* splicing have also been identified. As mentioned previously, *XBP-1* splicing and RIDD activity appear to have different Hill coefficients also, but RIDD requirement of monomers or oligomers has been disputed (Tam et al., 2014; Han et al., 2009). And contradictory to the notion of *XBP-1* splicing and RIDD

activity being separate is that both targets contain similar consensus sequences and predicted structure (Hur et al., 2012; Oikawa et al., 2010; Coelho and Domingos, 2014).

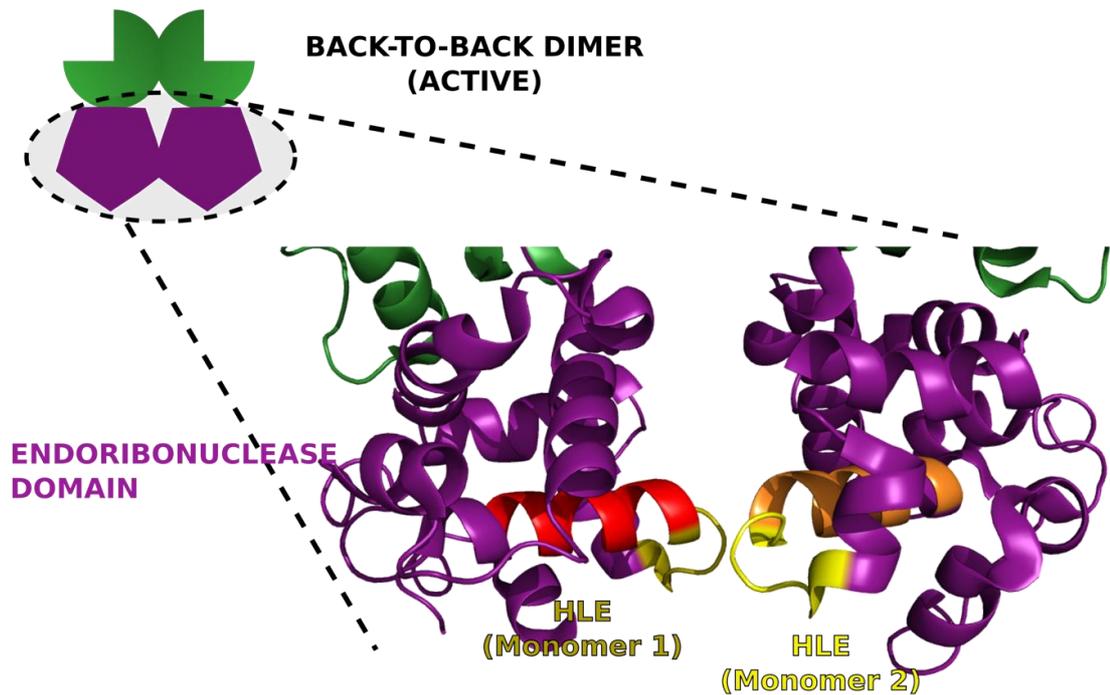


Figure 1.16

Important residues for endoribonuclease activity of the cytoplasmic domain. Residues 900-916 highlighted in each promoter that form the active site in the active back-to-back dimer (PDB: 4Z7H) (Joshi et al., 2015). In more active conformations of IRE1 α the HLE elements are less dynamic. Mutation to residues in the regions shown in red and orange abrogate endoribonuclease activity, suggesting their importance in the active site (Joshi et al., 2015; Korennykh, A. and Walter, 2012; Korennykh, A.V. et al., 2011b). Figure prepared using PyMOL version 1.7.

It is apparent that the back-to-back conformation is required for endoribonuclease activity in IRE1 α , and that rearrangements to the kinase domain have long range effects on this activity. These rearrangements can be caused by small molecules binding, phosphorylation and even from the inclusion of the long linker regions. The endoribonuclease activity can also be influenced by the concentration of IRE1 α 's cytoplasmic domain, this has been suggested to be due to the oligomerisation of the back-to-back conformation to achieve maximal endoribonuclease activity (Itzhak et al., 2014).

1.5.5. Oligomer

An oligomeric structure was solved with the yeast protein, with the back-to-back conformation being repeated as a rotating filament (PDB: 3FBV; Figure 1.12). The endoribonuclease domains

of the dimers in the filament are positioned very closely in space, suggesting high activity. The structure was solved with a short oligonucleotide bound to the endoribonuclease domain of the protein (assumed through electron density present). The structure represents the fully activated form of yeast IRE1 α , with activated kinase and endoribonuclease domains and a fully formed activation loop. The authors also observed that the oligomerisation of the protein increased its splicing activity (Korennykh, A.V. et al., 2011b; Korennykh, A.V. et al., 2011a).

Evidence for the importance of oligomerisation in activity for the human protein has also been shown by clustering of GFP-linked IRE1 α in cell assays after stress induction which corresponded with an increase in *XBP-1* splicing (Li, H. et al., 2010). A number of studies have observed the clustering of the yeast protein in a similar way (Aragon et al., 2009; Ishiwata-Kimata et al., 2013; Kimata et al., 2007; Shamu and Walter, 1996). Oligomerisation of the cytoplasmic has been hypothesised to drive higher endoribonuclease activity in the human protein, with phosphorylation promoting oligomerisation (Itzhak et al., 2014).

It has also been suggested that oligomers of IRE1 α promote RIDD activity as the oligomeric species identified in yeast protein have a larger active site for mRNA splicing, and therefore may promote promiscuity in mRNA targets (Korennykh, A.V. et al., 2009; Han et al., 2009). However, Tam et al. (2014) show foci of IRE1 α dissipating after eight hours in stressed HEK293 cells which coincides with a decline in *XBP-1* splicing and an increase in degradation of RIDD targets BLOS1 and SCARA3, peaking at around eight hours. This suggests that RIDD activity is not enhanced by oligomerisation. The study also showed a hill coefficient of 1.1 ± 0.29 for RIDD from IRE1 α compared to a coefficient of 3.07 ± 0.65 for cleavage of *XBP-1* mRNA. To support the data further, yeast IRE1 mutations preventing oligomerisation did not affect RIDD activity but did impact *HAC1* splicing activity (Tam et al., 2014).

It appears that the activation of IRE1 α 's cytoplasmic domain involves the transition of a monomer to a face-to-face dimer to promote autophosphorylation of the protein (Ali et al., 2011; Prischi et al., 2014). Phosphorylation promotes a more active kinase domain conformation which rearranges the α C-helix to allow formation of the back-to-back conformation, which have varying levels of endoribonuclease activity (Joshi et al., 2015; Concha et al., 2015). Once activated the back-to-back conformation can oligomerise to achieve maximal *XBP-1* splicing activity, this activation pathway through various crystal structures is represented in Figure 1.12. Although there are some conflicting reports, it appears that dissipation of these oligomers promotes IRE1 α 's RIDD activity (Korennykh, A.V. et al., 2009; Tam et al., 2014).

The structure to function relationship of the cytoplasmic domain's different responses are not yet fully understood. The same is true of the effect of the kinase domain on mRNA target selection. Further understanding of the conformations associated with each activity would be invaluable when designing therapeutic approaches to the problematic activation of IRE1 α in pathology, as described previously. The cytoplasmic domain therefore requires more in-depth structural analysis, especially to bridge the gaps between the activation cascade suggested through X-ray crystallography structures and to elucidate the intricacies in areas of importance and the conformational dynamics involved in selecting different mRNA targets.

1.6. Methods for the characterisation of protein interactions

As discussed, there are still features of IRE1 α 's activation pathway that are not fully understood, outlined in this section are methods that can provide more information about the activation pathway and interactions of the luminal domain of IRE1 α .

1.6.1. Techniques to monitor protein interactions and changes to protein conformation

Techniques to characterise the association and dissociation of molecules in solution can be used to better understand the conformational state of the luminal domain, including monomer to dimer transitions, binding to BiP and even to unfolded proteins. Three such techniques used in this thesis are fluorescence polarisation, microscale thermophoresis and nuclear magnetic resonance (NMR). Fluorescence polarisation and microscale thermophoresis both require a fluorescently labelled reporter molecule to be involved in the interaction being monitored and are able to report on global changes of the size of the complex that this molecule is in to give a precise binding affinity. NMR can report on similar conformational changes, giving site specific information about the regions of the protein that undergo changes in these interactions. These three methods for characterising protein interactions and changes in conformation are described in the following sections.

1.6.1.1. Fluorescence polarisation

Larger molecules tumble slower in solution due to the principles of Brownian rotation, this behaviour can be exploited by fluorescence polarisation. A fluorescent molecule can be excited by light of a specific wavelength to a higher energy level, subsequently the molecule returns to its resting energy level, releasing energy as light of a different wavelength. Use of plane polarised light filters allow use of polarised light to excite a fluorophore, the fluorophore will then emit

photons at the emission wavelength that are also polarised, fluorescence parallel and perpendicular to the plane of excitation can be detected. Between when the fluorophore is excited to when emitting light, larger molecules will tumble less, therefore more parallel light to the excitation plane will be emitted. On the other hand, smaller molecules will tumble more therefore having less correlation of the light polarisation and less fluorescence parallel to the excitation light (Figure 1.17). In this way an increase in size of a complex can be detected (Jameson and Ross, 2010).

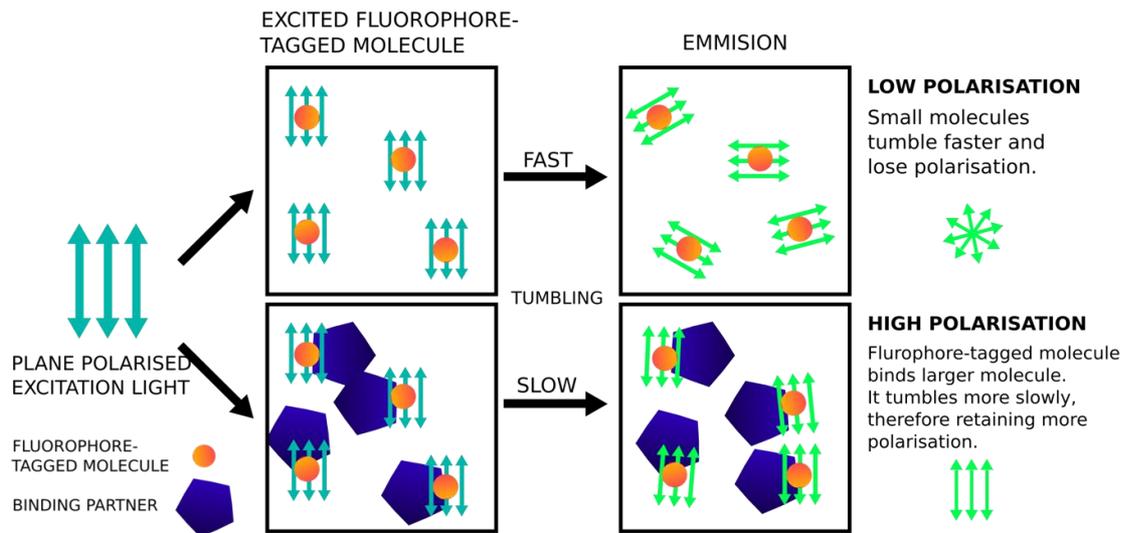


Figure 1.17

Representation of the principles of fluorescence polarisation. Fluorophore-tagged molecules are excited with plane polarised light. Molecules tumble quickly in solution, if they bind to a larger molecule, they will tumble slower. Molecules that tumble slower retain light polarisation in their emission, therefore having higher detectable polarisation than small molecules that tumble quickly and lose polarisation.

1.6.1.2. Microscale thermophoresis

Microscale thermophoresis also relies on the principles of Brownian motion for its effect. In solution larger molecules move more slowly than smaller molecules, in addition to this the temperature of the solution effects this movement, where higher temperature means a higher energy and therefore faster movement. Microscale thermophoresis utilises this effect by using an infrared laser to heat up a small area of a capillary with the protein sample in it (Figure 1.18). The reporter molecule is fluorescently labelled, and the fluorescence of this heated area is measured throughout the procedure. The increase in temperature causes molecules to diffuse away from the area, therefore a loss of fluorescence is detected, with smaller molecules diffusing away faster than larger molecules. The infrared laser is then turned off, returning the temperature of this area to the same as the rest of the capillary. This prompts molecules to diffuse back into the previously

heated area. Thus, there is an increase in fluorescence detected again and larger molecules will migrate back to the area more slowly than small molecules. Multiple capillaries are measured with different concentrations of the binding partner of the fluorescently labelled molecule. In this way the affinity of interaction between the reporter molecule and its binding partner can be measured, as when bound to its partner the molecule will be larger, move more slowly and the fluorescence signal will be different (Jerabek-Willemsen et al., 2014).

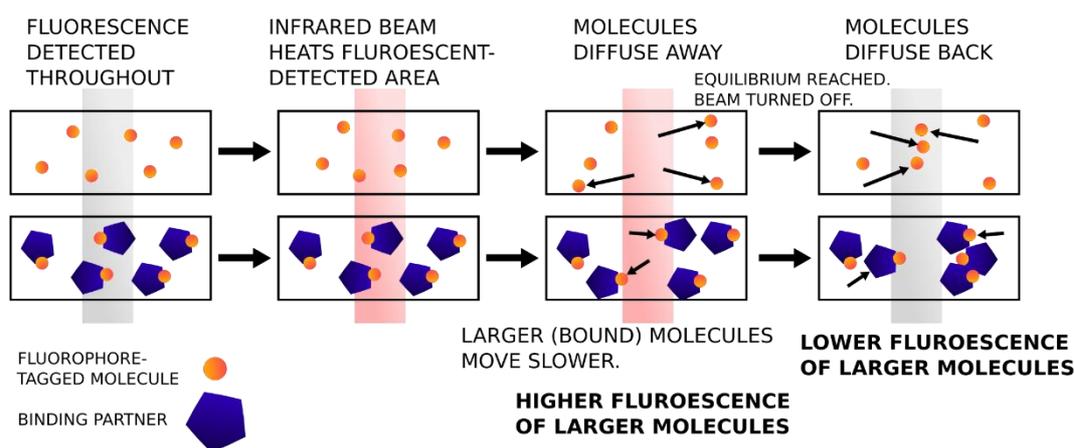


Figure 1.18

Schematic for the principles of microscale thermophoresis. Fluorescently-tagged molecules are mixed in separate capillaries with increasing concentrations of binding partner. The fluorescence of one area of the capillary is measured throughout. An infrared beam is applied to this area, heating it. Molecules diffuse away, larger molecules diffuse away more slowly, meaning a higher fluorescence. After equilibrium is reached, the infrared beam can be turned off, allowing the initial equilibrium of molecules in the fluorescently detected area to be restored. Larger molecules will diffuse back more slowly than smaller molecules, thus giving a lower fluorescence signal.

1.6.1.3. Nuclear magnetic resonance

Whereas fluorescence polarisation and microscale thermophoresis methods give global information about a protein's interactions, NMR can be used to give site specific information about these interactions and changes in conformation. The concept of NMR was first discovered in the 1940's and the first spectrum of a protein produced in 1957 (Saunders et al., 1957). Since then the method has been developed to be one of the most powerful structural tools for studying proteins and an ideal technique for observing disordered and dynamic regions of proteins and their interactions, which are often not resolved by methods such as X-ray crystallography. NMR is therefore ideal for observing the different conformations in the luminal domain activation cascade, how changes such as inclusion of binding partners or cancer-associated mutations affect

them and importantly, such a technique may allow for the eventual understanding of the conformational changes that dictate how the cytoplasmic domain changes its mRNA targets.

NMR reports on NMR active nuclei, these are atoms that have a property called spin with a value of $\frac{1}{2}$; such atoms include ^1H , which is abundant in proteins. A 1D NMR spectrum of a protein would theoretically show a peak for each different ^1H atom of the protein. The peaks for each different ^1H atom will have a position (x-axis, frequency) determined by its chemical environment. However, in proteins ^1H atoms are so abundant that the signals for each proton would overlap in position, meaning that site-specific information would be lost. For this reason, 2D and 3D NMR can be used. In this case, the signal for one NMR active nuclei (such as a proton) is coupled with the signal of another, covalently bonded NMR-active nuclei (such as isotopic ^{15}N nitrogen). This is usually by use of an 'INEPT' pulse sequence, that allows the passage of signal from one covalently bonded atom to another (Morris and Freeman, 1979).

Therefore, in a 2D experiment of this kind, each peak of the spectrum reports on the chemical environment of one pair of covalently bonded ^1H and ^{15}N atoms. The signal retains information about the ^1H and ^{15}N atoms' chemical environments and so the spectral peaks are separated by nuclei 1 (x-axis) and nuclei 2 (y-axis) giving better dispersion of the peaks so that less site-specific information is lost due to overlapping positions.

However, the additional NMR active nuclei (other than ^1H) must be incorporated into the protein of interest. Commonly used NMR active nuclei to couple the ^1H signal to are ^{15}N and ^{13}C isotopes, as nitrogen and carbon atoms are abundant in proteins. Only 1.1% of carbon and 0.4% of nitrogen atoms are naturally occurring NMR active isotopes, so methods have been developed for the introduction of these isotopes into recombinantly produced proteins. These methods allow proteins to be labelled in many different ways, from relatively simple ubiquitous isotopic labelling of ^{15}N atoms to specific labelling of amino acid sidechains (Tugarinov et al., 2006). One caveat of the labelling techniques is that many of the cost-effective schemes are specific for recombinant *E. coli* expression systems, meaning that the protein of interest must be readily producible by *E. coli* expression. The luminal domain of IRE1 α has an optimised *E. coli* expression protocol (Liu et al., 2002) whereas the cytoplasmic domain does not and is often expressed in insect expression systems instead (Joshi et al., 2015).

After gaining an isotopically labelled protein of interest, 2D NMR experiments can be used. Different sequences of radio-frequency pulses can be used to optimise the signal gained from large proteins. For the study of large proteins, such as IRE1 α , a commonly used experiment is the ^1H ^{15}N transverse relaxation optimised spectroscopy (TROSY) experiment, this experiment is

employed in the majority of cases in this thesis (Pervushin et al., 1997). In the TROSY experiments used each peak in the spectra gives information about a covalently bonded HN group in the backbone of the protein's polypeptide chain, and so the position of the peak reports on the chemical environment of the individual backbone HN group through separation by a ^1H and ^{15}N axis (Figure 1.19).

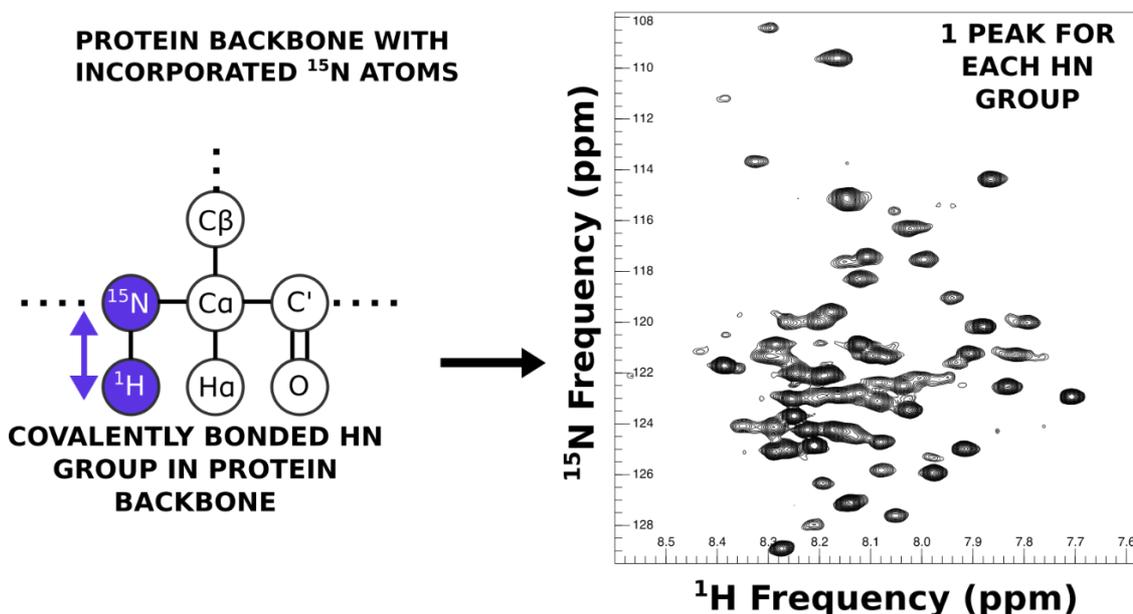


Figure 1.19

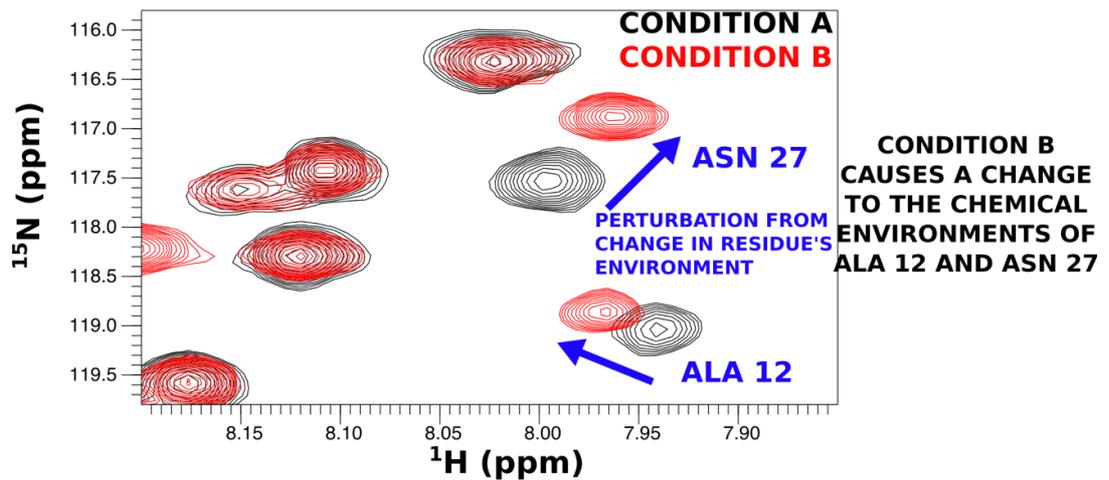
The INEPT pulse sequence allows magnetisation to be transferred between covalently bonded ^1H and incorporated ^{15}N atoms in the protein amide backbone. The signal can then be Fourier transformed and plotted on 2-dimensional spectra where peaks have a ^1H and a ^{15}N frequency domain, thus separating otherwise overlapping chemical shifts.

After obtaining a 2D spectrum of the protein of interest it is possible to change the conditions of the experiment and observe the changes in the spectrum. In most cases solution NMR allows for experimental conditions to be changed similarly to accompanying *in vitro* assays to give site specific information processes observed in those assays, such as interactions with binding partners, changes in conformation or inclusion of cancer-associated mutations. This is visualised by peaks (or chemical shifts) in the spectrum changing position or intensity, therefore indicating a change in the chemical environment of that peak's corresponding residue in that condition (Figure 1.20A). In this way regions of importance in the protein for these processes can be identified, although to obtain site-specific information the peaks in the spectrum must be assigned to the residues in the protein that they report on. Assignment of a protein can be carried out using a set of 3D NMR experiments (Frueh, 2014) or by mutation (Siivari et al., 1995;

Wieteska et al., 2017), where a mutation or truncation of a specific part of the protein causes peaks in the spectrum to disappear, which can then be assigned to the altered region of the protein.

NMR can also be used to give structural information, such as solving a protein structure, although this is not always possible and is dependent on the protein's characteristics. However, other experiments can be used to give structural information about a protein of interest, one such technique is temperature correlation, which gives information on the secondary structure present in the protein (Figure 1.20B). As before, HN TROSY experiments are used. By changing the temperature of the experiment, the chemical shift of each peak migrates along the ^1H axis, in a linear relationship to temperature. The magnitude of this change in position gives information about the secondary structure of those residues. This is because secondary structure motifs have many well-defined hydrogen bonds from the backbone of the polypeptide chain, whereas unstructured regions have hydrogen bonds with the solvent, that have less well-defined lengths. When modulating temperature, the energy of these hydrogen bonds is changed. The hydrogen bonds in rigid secondary structural elements are less affected by the temperature change than the unstructured regions, and therefore structured regions exhibit a smaller shift in the ^1H axis of the NMR spectrum (Okazaki et al., 2018; Tomlinson and Williamson, 2012).

A.



B.

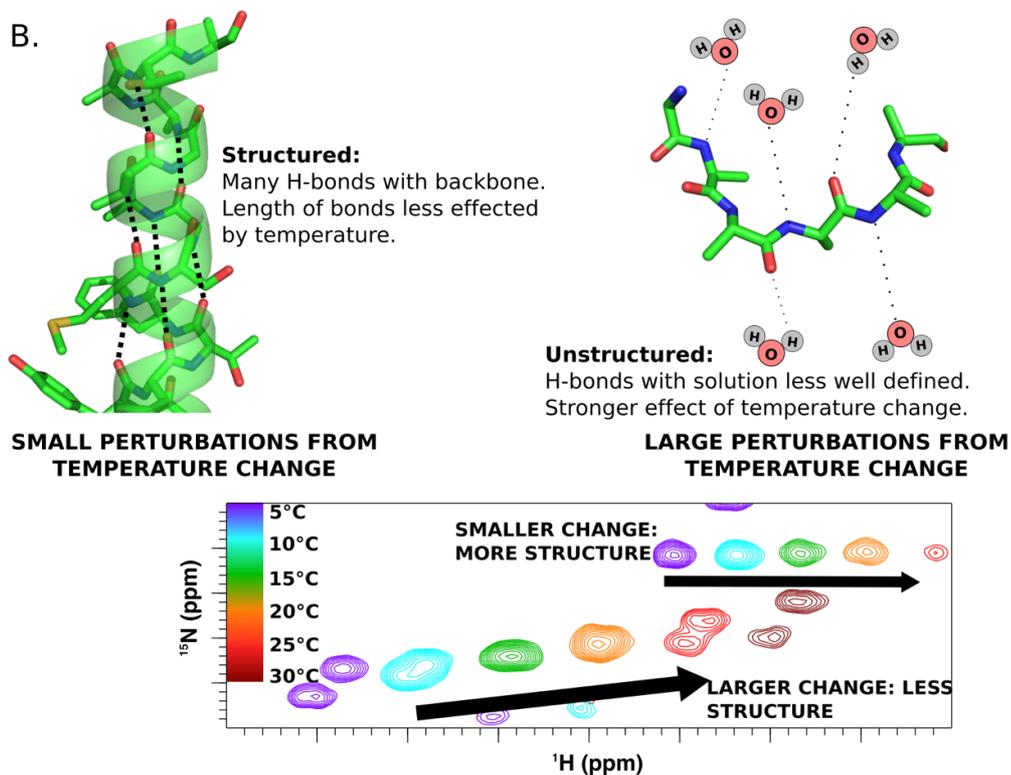


Figure 1.20

By modulating conditions of a protein NMR experiment information about the protein can be gained. A. Upon changing conditions, the peaks of residues that have had their chemical environment altered will exhibit a shift. Such conditions can be binding to another protein, interaction with a small molecule or changes in the protein's conformation due to mutation. B. Temperature correlation experiments can be used to report on hydrogen bonds existing in the backbone of the protein and therefore give information on the presence of secondary structural elements. More structured regions give a smaller ^1H temperature-based perturbation than less structured regions. Figure prepared using PyMOL version 1.7.

More complex experiments can also be used to understand protein dynamics on timescales of milli-seconds to pico-seconds (Kleckner and Foster, 2011). Experiments such as these would be invaluable in the study of the cytoplasmic domain's activation pathway, to create a bridge in information between the obtained crystal structures and to provide precise information for the nuances of differential mRNA target selection. However, as mentioned a protocol for production of the cytoplasmic domain in *E. coli* must first be optimised. NMR can therefore provide complimentary information to the other assays discussed here about the luminal domain's conformational changes.

1.6.2. Size exclusion chromatography

Although giving information about conformational changes, the assays described are unable to give specific information about the size of the species that they are reporting on, this can be complimented by use of size exclusion chromatography. Size exclusion chromatography involves a column packed with a matrix of particles (Figure 1.21). Aqueous protein samples can be passed through the column's matrix, larger molecules are too big to interact with the matrix and will simply flow through the column and therefore be eluted first. Smaller molecules on the other hand will be influenced by the matrix and therefore take longer, requiring larger volumes of buffer to elute them from the column. The protein's interaction with the column is dependent on its hydrodynamic radius, a calibration curve can be constructed from known protein standards. In this way size exclusion chromatography can be used to determine the size (through hydrodynamic radius) of protein conformers, as proteins existing in multimers will remain in complex as they pass through the column. It is also possible to lower the concentration of these protein complexes in order to break the complex and use this information to determine the affinity of interaction, however, this information is not as precise as with fluorescence polarisation and microscale thermophoresis (Fekete et al., 2014). Therefore, size exclusion chromatography can give information about the size of high affinity multimeric states of the luminal and cytoplasmic domains *in vitro* and how different conditions affect these multimers; this is of particular use when characterising oligomeric species of the constructs.

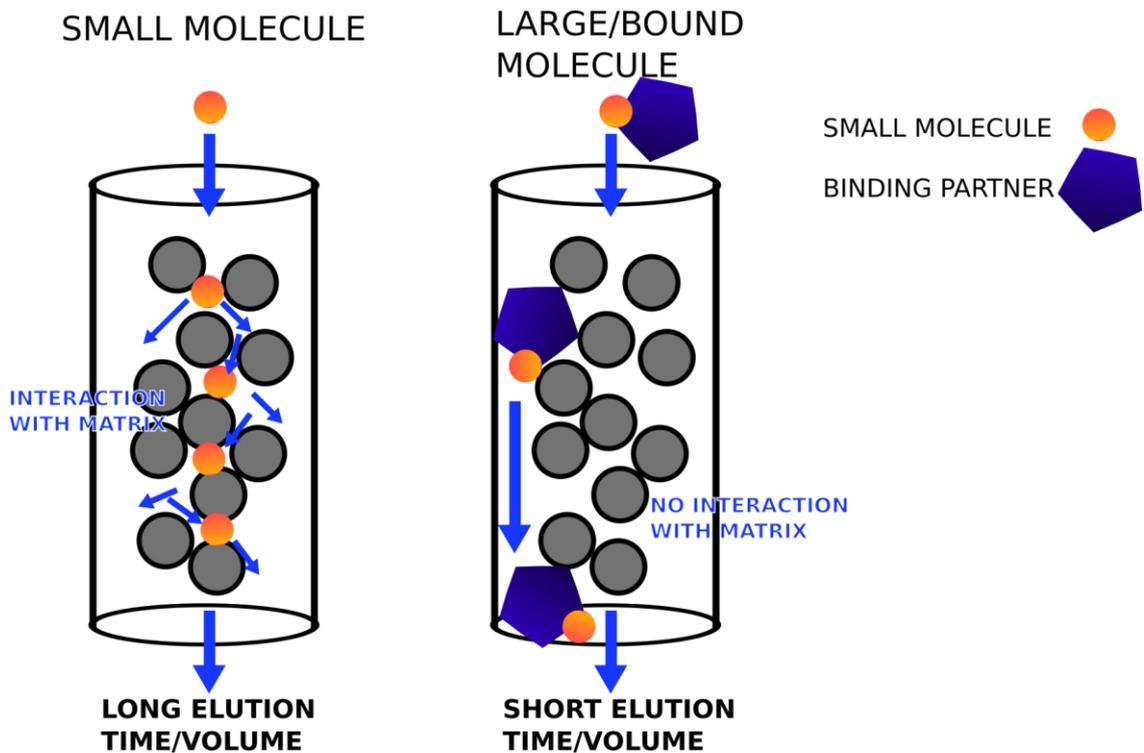


Figure 1.21

Simplified representation of the principle of size exclusion chromatography. The column includes a matrix. The matrix allows smaller molecules to enter and interact with it, the smaller the molecule the more interactions it will have, impeding its flow through the column. Therefore, smaller molecules require more solution to be passed through the column to elute. Larger molecules have less interactions with the column and therefore elute in a smaller volume of solution.

1.6.3. Turbidity assays

Turbidity assays represent a reproducible method for observing the formation of protein multimers, using absorbance at 400nm (Figure 1.22). Turbidity assays are used widely in the study of amyloid proteins as they effectively report on the extension of the amyloid fibril through light scattering (Zhao et al., 2016) and have also been used to observe oligomerisation of IRE1 α 's cytoplasmic domain (Korennykh, A. and Walter, 2012). They can also be applied to the activation mechanism of IRE1 α 's luminal domain, which forms oligomers upon interaction with unfolded protein mimics (Gardner and Walter, 2011; Karagoz et al., 2017). The turbidity assays allow a distinction of the kinetics of the formation of large species and how these species are affected in different conditions (Borgia et al., 2013). Large protein oligomers will often precipitate in *in vitro* assays, this makes the previously discussed techniques in this section redundant as they are not compatible with precipitated protein. However, turbidity assays can be coupled with solubility

assays to create a calibration curve to relate the optical density at 400nm (OD400) absorbance reading to the amount of insoluble protein present (due to formation of large oligomers).

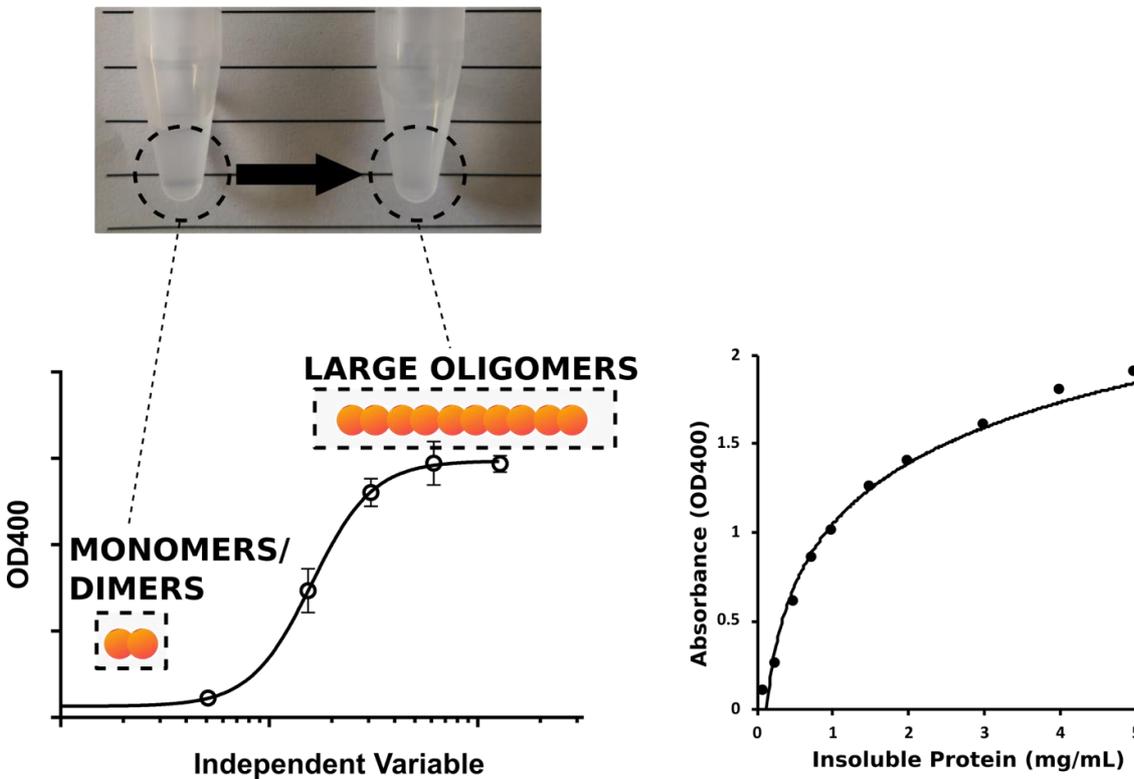


Figure 1.22

Turbidity assays report on the scattering of light at 400nm. Shown is a low and high turbidity sample, where visible light is able to pass through the low visibility sample but less so in the high turbidity sample alongside an example of a calibration curve for conversion of OD400 absorbance values to amount of insoluble protein formed. Turbidity assays can give kinetics of oligomerisation/precipitation of proteins over time and based on an independent variable.

1.7. Project aims

The aim of this thesis is to increase the understanding of IRE1 α 's activation and elucidate how selected cancer-associated mutations of the luminal domain perturb the protein's function. Firstly, this will involve producing and optimising a method for the recombinant *E. coli* expression of the protein's cytoplasmic domain. *E. coli* expression of the cytoplasmic domain can offer higher yield expression and make available isotopic labelling strategies for the protein. This would allow for accelerated study of the protein's structure and dynamics by high powered techniques such as NMR. Such study of the domain may lead to a greater understanding of the protein's activation mechanism, how it selects different mRNA targets and therefore how it can be therapeutically targeted to control function.

The second focus of this thesis will be to further characterise conformations adopted by the luminal domain in activation and how the protein's signal is terminated. This will be accomplished by designing assays to separate individual processes and conformations of the protein and to study them using the methods described in Section 1.6. Further understanding of the interactions and regions of the protein important in these interactions will allow validation of the different models for luminal domain activation and to understand the current discrepancies in them.

The third focus of the research presented in this thesis will be to gain an understanding of how cancer-associated mutations of the luminal domain perturb the conformational landscape of the protein. This will be carried out by study of the cancer-associated mutants using assays developed for the study of individual conformations and interactions of the *wild-type* protein. Understanding of how cancer-associated mutations affect the protein's conformation can aid in therapeutic design against the studied mutations but can also elucidate structural regions of importance in the protein's activation as potential drug targets to control the activity of the protein in the numerous associated diseases of IRE1 α .

2. Materials and methods

2.1. List of general reagents and kits

Table 2.1

A list of reagents used here and their supplier.

Reagents and apparatus	Supplier
10kDa Amicon Ultra centrifugal filter units	Merck
2-mercaptoethanol	Fisher Scientific
4x Laemmli sample buffer	Bio-Rad
Agar	Fisher Scientific
Agarose	Invitrogen
Ammonium chloride	AnalaR
Ammonium chloride (¹⁵ N)	CK Isotopes
Arginine	Sigma Aldrich
Bradford Assay BSA Standards	Bio-Rad
Bradford Reagent	Bio-Rad
BugBuster Master mix	Merck
Calcium chloride	Sigma Aldrich
COmplete EDTA free protease inhibitor tablets	Roche
Custom DNA oligonucleotides	IDT
Custom RNA oligonucleotides	Stratech
D ₂ O	CK Isotopes
DMSO	Alfa-aesar
DNase I	Merck
DTT	Sigma Aldrich
EDTA	Fisher Scientific
Ethanol	Sigma Aldrich
Ethidium bromide	ThermoFisher Scientific
Eukaryotic produced cytoplasmic domain	SinoBiological
Filters (0.22 and 0.45μM)	Merck
Glucose	Fisher Scientific
Glycerol	Fisher Scientific
Guanidine hydrochloride	Sigma Aldrich
HEPES	Fisher Scientific

His-Trap FF 1mL column	GE Healthcare Life Sciences
HMW Gel Filtration Calibration Kit	GE Healthcare Life Sciences
Imidazole	Sigma Aldrich
IPTG	Generon
Kanamycin sulfate from Streptomyces	Sigma Aldrich
LB broth	Fisher Scientific
Loading Dye (agarose gel)	New England Biolabs
Lysozyme	Merck
Magnesium chloride	Sigma Aldrich
Magnesium sulfate	Fisher Scientific
Microplates	Grenier Bio-One
Molecular Weight Marker (agarose gel)	New England Biolabs
NAP5 column	GE Healthcare Life Sciences
Nickel chloride	Sigma Aldrich
PCR purification kit	Invitrogen
PD SpinTrap G-25 column	GE Healthcare Life Sciences
Plasmids (1B, 1G, 1M)	Addgene
Potassium chloride	Merck
Potassium hydroxide	Fisher Scientific
Potassium phosphate monobasic	MP Biomedicals
Protein weight marker	Bio-Rad
Pur-a-lyzer Maxi Dialysis Tubes	Sigma Aldrich
PureLink PCR Purification Kit	ThermoFisher Scientific
PureLink Quick Plasmid Miniprep Kit	ThermoFisher Scientific
Q5 High-fidelity 2x Master mix	New England Biolabs
Qiaquick gel extraction kit	Qiagen
Quikchange lightning enzyme	Agilent
RNase A	ThermoFisher Scientific
SDS-PAGE precast gels	Bio-Rad
Sodium azide (NaN ₃)	Sigma Aldrich
Sodium carbonate	AnalaR VWR
Sodium chloride	Fisher Scientific
Sodium hydroxide	Fisher Scientific
Sodium phosphate dibasic	Acros Organics
Sodium phosphate monobasic	Sigma Aldrich

SspI restriction enzyme	ThermoFisher Scientific
Sterile filter units (stericups)	Merck
Superdex 200 10/300 GL column	GE Healthcare Life Sciences
T4 DNA polymerase reagents	Invitrogen
TAE buffer	ThermoFisher Scientific
TCEP	Fluorochem
Thiamine	Sigma Aldrich
Tris	Fisher Scientific
Triton X-100	Fisher Scientific
U- ¹³ C6 D-glucose	CK Isotopes
Urea	AnalaR VWR

2.2. *E. coli* strains

Genotypes of *E. coli* strains used:

E. coli DH5 α : F- U169 recA1 endA1 hsdR17(rk-, mk+) phoA supE44 thi-1 gyrA96 relA1 λ -

E. coli BL21 DE3: F- dcm ompT hsdSB (rB-, mB-) gal λ (DE3)

E. coli Rosetta 2: F- ompT hsdSB(rB- mB-) gal dcm (DE3) pRARE2 (CamR)

E. coli C3013: MiniF lysY lacIq(CamR) / fhuA2 lacZ::T7 gene1 [lon] ompT gal sulA11 R(mcr-73::miniTn10--TetS)2 [dcm] R(zgb-210::Tn10--TetS) endA1 Δ (mcrC-mrr) 114::IS10

E. coli Origami 2: Δ (ara-leu)7697 Δ lacX74 Δ phoA PvuII phoR araD139 ahpC galE galK rpsL F'[lac+ lacIq pro] (DE3) gor522::Tn10 trxB (StrR, TetR)

2.3. Molecular biology methods

2.3.1. Polymerase chain reaction

2.3.1.1. PCR reaction for DNA insert

The DNA insert for the cytoplasmic domain of IRE1 α (548-977) was created using the ligation independent cloning (LIC) insert primers:

Forward: 5' - TACTTCCAATCCAATGCAAGCAGCCCCTCCCTGGAACAAGACG -3'

Reverse: 5' - GGATTGGAAGTAGAGGTTCTCGAGGGCGTCTGGAGTCACTGG -3'

The PCR was carried out with Quikchange lightning enzyme as per the manufacturer's protocol.

The PCR reaction consisted of 18 cycles of:

95°C 20 seconds

60°C 10 seconds

68°C 3 minutes

The PCR reaction was carried out using a pBS vector which contained the full IRE1 α sequence (donated by Prof. David Ron, University of Cambridge).

2.3.1.2. PCR clean up reaction

The PCR clean up reaction was carried out after PCR using the PureLink PCR Purification kit as per the manufacturer's instructions.

2.3.2. SspI Digest

Digestion of plasmids (1B, 1M and 1G) by SspI restriction enzyme were carried out at 37°C for one hour with shaking at 200rpm. 2 μ L of SspI 10x buffer, and 1mg plasmid DNA total, Milli-Q pure water to 19 μ L and 1 μ L of the SspI enzyme.

2.3.3. Agarose gel

2.3.3.1. Running agarose gels

Agarose gel were created by dissolving 1g of agarose in 100mL of 1x TAE buffer with 3 μ L of ethidium bromide added and left to set in gel housing. Molecular weight ladders used in all gels run and loading buffer was added to each DNA sample to be run on the gel.

2.3.3.2. Gel extraction

Bands of DNA excised from gel using a sterile blade and purified using a QIAquick gel extraction kit, following the manufacturer's instructions.

2.3.4. Ligation independent cloning reaction

2.3.4.1. T4 Polymerase Digestion

Gel purified vector was added to 2.5mM of dGTP, 5mM DTT and T4 polymerase buffer before T4 polymerase was added. The same was also mixed with the gel purified cytoplasmic insert, but with dCTP replacing dGTP in the reaction.

Both reactions were incubated at room temperature for 30 minutes before 75°C for 20 minutes to stop the reaction.

2.3.4.2. LIC Annealing Reaction

The separate vector and insert T4 polymerase reactions were mixed to allow for annealing at a ratio of 6:1 insert to vector. This reaction was incubated for one hour at room temperature.

2.3.5. Transformation

Transformations were carried out with the same protocol for different *E. coli* strains used and all in sterile conditions. *E. coli* competent cells were thawed on ice for 5 minutes before plasmid DNA was added (typically 100ng for plasmid transformations and 8 μ L of mutagenesis reactions) to the competent cells, and then left on ice for 20 minutes. The cells were then heat shocked for 45 seconds at 42°C before being placed on ice again for 2 minutes. 500mL of LB media (25g/L) was added to the cells and incubated with shaking (200rpm) for 30 minutes at 37°C before being centrifuged at 3500xg for 5 minutes. The pelleted cells were resuspended in 50 μ L LB media and spread onto LB agar plates with the appropriate antibiotic.

2.3.6. Plasmid extraction

A Single colony from successfully transformed DH5 α competent cells was taken, grown overnight in 5mL of LB media with the appropriate antibiotic shaking (200rpm) at 37°C. The culture had its plasmid DNA isolated using the PureLink Quick Plasmid Miniprep kit, as per the manufacturer's instructions. The isolated DNA plasmid had its concentration calculated using A260 values.

2.3.7. Sequencing

For sequencing plasmids were diluted to 100ng/ μ L and sent for the Source BioScience Sanger Sequencing Service to be sequenced using the T7 promoter forward and reverse stock primers provided by Source BioScience.

2.3.8. Mutagenesis

For mutagenesis Q5 High-fidelity 2x Master mix was used as per the manufacturer's instructions. Primers used are listed below.

Cytoplasmic domain mutants:

I642G

Forward: 5'- G TAC ATT GCC GGC GAG CTG TGT G -3'

Reverse: 5'- C ACA CAG CTC GCC GGC AAT GTA C -3'

Luminal domain mutants:

N244S

Forward: 5'- GGG TCT CCA CAG CGA CGC TGA TGT GCA TCA CCT TC -3'

Reverse: 5'- GAA GGT GAT GCA CAT CAG CGT CGC TGT GGA GAC CC -3'

S296F

Forward: 5' - ACC ACG CTC TAT GCC TTT CCC TC -3'

Reverse: 5' - GTG TAC CAT TGA GGG AAA GGC ATA GAG G -3'

A414T

Forward: 5' - CAG ACT TCA GAA AAC ACC CCT ACC ACC -3'

Reverse: 5' - G AGA CAC GGT GGT AGG GGT GTT TTC -3'

V418M

Forward: 5' - CCT CCA CAT CCC GAG ACA TGG TGG TAG GTG CG -3'

Reverse: 5' - CGC ACC TAC CAC CAT GTC TCG GGA TGT GGA GG -3'

WLLI – GSSG

Forward: 5' - GCT CAA CTA CTT GAG GAA TTA CGG GAG TTC GGG AGG ACA CCA
TGA AAC CCC ACT G -3'

Reverse: 5' – CAG TGG GGT TTC ATG GTG TCC TCC CGA ACT CCC GTA ATT CCT CAA
GTA GTT GAG C -3'

L406-stop

Forward: 5' - TTTGAGGAAGTTATCAAC TAG GTTGACCAGACTTCAG -3'

Reverse: 5' - CTGAAGTCTGGTCAAC CTA GTTGATAACTTCCTCAAA -3'

S411-stop

Forward: 5' - GTTGACCAGACT TAA GAAAACGCACCTACC -3'

Reverse: 5' - GGTAGGTGCGTTTTTC TTA AGTCTGGTCAAC -3'

T416-stop

Forward: 5' - GAAAACGCACCT TAA ACCGTGTCTCGGGATG -3'

Reverse: 5' - CATCCCGAGACACGGT TTA AGGTGCGTTTTTC -3'

D421-stop

Forward: 5' - CGTGTCTCGG TAA GTGGAGGAGAAG -3'

Reverse: 5' - CGTGTCTCGG TAA GTGGAGGAGAAG -3'

P426-stop

Forward: 5' - GGAGGAGAAG TAA GCCCATGCCCC -3'

Reverse: 5' - GGGGCATGGGC TTA CTTCTCCTCC -3'

2.4. General bacterial methods

2.4.1. Sterilisation

Sterilisation was carried out by autoclaving at 123°C for 40 minutes; in cases where compounds in solution were heat sensitive sterilisation was carried out through filtration with a 0.22µM filter in a sterile environment.

2.4.2. LB media

LB media used was 25g/L of lysogeny broth in dH₂O and was sterilised by autoclaving. After cooling, the desired antibiotic was added.

2.4.3. LB agar plates

LB agar plates were produced with 25g/L lysogeny broth and 15g/mL agar in dH₂O. The solution was sterilised using the autoclave before being left to cool. The required antibiotics were added to the media and it was then poured into petri dishes.

2.4.4. M9 Media

M9 media was used for isotopic labelled growth. M9 media was produced by adding per 1L: 6.5g Na₂HPO₄, 3g KH₂PO₄, 0.5g NaCl, 4g D-glucose (for ¹³C labelling 2g of U-¹³C₆ D-glucose was used), 2.5g LB broth, 120mg MgSO₄, 11mg CaCl₂, 10µg/mL of biotin and thiamine and 1g of NH₄Cl (or ¹⁵NH₄Cl for ¹⁵N labelling).

The media was stirred at room temperature for 10 minutes before being filtered using sterile filter units and stored at 4°C until required.

2.4.5. Protein expression

2.4.5.1. 30 and 37°C expression

For 30°C and 37°C expression in *E. coli*, a single *E. coli* colony was taken from the transformation of the expression plasmid into the BL21 DE3 strain of *E. coli* (unless strain is stated otherwise) and cultured in 5mL LB media with kanamycin at 37°C overnight whilst shaking at 200rpm. In the morning the OD₆₀₀ of this culture was measured. This culture was used to give an OD₆₀₀ of 0.1 in a larger culture (50 – 100mL) of LB media with kanamycin. This culture was then grown at 37°C whilst shaking at 200rpm.

For small scale expressions IPTG was added at a final concentration of 1mM (unless stated otherwise) to the 50mL culture once it had reached an OD₆₀₀ of 0.8. The culture was then grown for four hours (at 30 or 37°C) with shaking at 200rpm before samples were taken for solubility tests/the whole culture was pelleted (unless stated otherwise).

For larger expressions, the culture was grown until the OD₆₀₀ value was 1.0. At this point this grown culture was added to 500-1000mL LB media with kanamycin to give an OD₆₀₀ of 0.1. This culture was then grown until its OD₆₀₀ value reached 0.8. At this point IPTG was added to

a concentration of 1mM (unless stated otherwise) and the culture was grown for four hours at 30/37°C with shaking at 200rpm. After four hours the culture was pelleted by centrifugation at 6000xg for 30 minutes.

2.4.5.2. 20°C expression

For 20°C expression in *E. coli*, as before, a single *E. coli* colony was taken from the transformed expression strain and cultured in 5mL LB media with kanamycin at 37°C from 9am whilst shaking at 200rpm. After ~five hours of growing the OD600 value was measured and this culture was used to give an OD600 of 0.1 in a larger culture (50 – 100mL) of LB media with kanamycin. This culture was then grown at 37°C whilst shaking at 200rpm.

For small scale expressions IPTG was added at a final concentration of 1mM to the 50mL culture once it has reached OD600 of 0.8. The culture was then grown overnight at 20°C with shaking at 200rpm before samples were taken for solubility tests/the whole culture was pelleted (unless stated otherwise).

For larger expressions, the culture was grown until the OD600 value was 1.0. At this point this grown culture was added to 500-1000mL LB media with kanamycin to give an OD600 of 0.1. This culture was then grown until its OD600 value reached 0.8. At this point IPTG was added to a concentration of 1mM and the culture was grown overnight at 20°C with shaking at 200rpm. In the morning the culture was pelleted by centrifugation at 6000xg for 30 minutes.

2.4.5.3. Labelled expression

For labelled expression for NMR study of the desired protein the same protocols as for 30/37°C or 20°C expression were used. The starting 5mL culture was LB media with kanamycin, but all larger cultures used (50-100mL and 500-1000mL) utilised M9 media with the selected labelled compounds added with kanamycin. The cells were then centrifuged as before, 6000xg for 30 minutes.

2.4.5.4. Different construct's expression conditions

Unless stated otherwise Table 2.2 shows the expression temperatures used for each construct and mutant used.

Table 2.2

The bacterial overexpression conditions used for each construct.

37°C Expression	20°C Expression
Cytoplasmic domain (all constructs and mutants)	
BiP	
Luminal domain constructs:	Luminal domain constructs:
<i>Wild-type</i>	D123P
N244S	WLLI-GSSG
S296F	D123P+A414T
V418M	WLLI-GSSG+A414T
L406stop	Core luminal domain
S411stop	
T416stop	
D421stop	
P426stop	

2.4.6. Preparation of expression culture pellet for purification

2.4.6.1. Cytoplasmic domain

After centrifugation the pellet of the bacterial culture was resuspended in inclusion body preparation buffer (100mM Tris pH 7.5, 150mM NaCl, 10mM EDTA and 10% glycerol), 35mL to resuspend 1L of culture. The resuspended pellet could then be frozen at -80°C.

The resuspended pellet was thawed when required. DNase I (40mg/1000mL), lysozyme (10mg/100mL) and protease inhibitor tablets were added to the thawed resuspended pellet. The cells were then sonicated in a cycle of 6 seconds sonication and 6 seconds rest so that the total sonication time was 6 minutes.

2.5% Triton X-100 was then added to the disrupted cells and the solution was stirred at 4°C for one hour to solubilise membrane fragments. The inclusion bodies present were then sedimented by centrifugation at 20,000rpm for 20 minutes. The pellet was then washed three times with inclusion body preparation buffer. The pellet was then resuspended in 100mL of cytoplasmic buffer A (50mM HEPES pH 7.5, 6M guanidinium, 1mM TCEP and 10% glycerol) and the solution was stirred at 4°C for >1 hour. After this the solution was taken and centrifuged at 20,000rpm for 20 minutes. The supernatant was taken and filtered through a 0.45µM filter, ready for purification.

2.4.6.2. Luminal domain and BiP

After centrifugation the pellet of the bacterial culture was resuspended in luminal domain buffer A (20mM HEPES pH 8.0, 400mM NaCl), 35mL to resuspend 1L of culture. The resuspended pellet could then be frozen at -80°C.

The resuspended pellet was thawed when required. DNase I (40mg/1000mL), lysozyme (10mg/100mL) and protease inhibitor tablets were added to the thawed resuspended pellet. The cells were then sonicated in a cycle of 6 seconds sonication and 6 seconds rest so that the total sonication time was 6 minutes. The cells were then centrifuged at 20,000rpm for 20 minutes. The supernatant was then taken and filtered with a 0.45µM filter ready for purification.

2.5. General protein methods

2.5.1. Bacterial overexpression solubility tests

Solubility tests were used after small scale (50mL) bacterial expression cultures. After induction with IPTG for the desired time a 1mL and a 0.5mL sample were taken. Both were centrifuged at 4500rpm for 3 minutes.

The 1mL culture sample had its supernatant removed and the pellet resuspended in 250µL of BugBuster master mix and was incubated at room temperature with gentle shaking for 20 minutes. After this the sample was centrifuged at 14,000rpm for 5 minutes. 20µL of the supernatant was taken and added to 8µL Laemmli sample buffer and 12µL 8M urea, this is the soluble protein fraction SDS-PAGE gel sample. The pellet was washed with LB media three times, before being resuspended in 16µL of Laemmli sample buffer and 24µL of 8M urea, this was then boiled before 2µL was taken and added to 8µL Laemmli sample buffer, 3µL of dH₂O and 12µL 8M urea. This is the insoluble protein fraction SDS-PAGE gel sample.

The 0.5mL culture sample's supernatant was removed and the pellet was resuspended in 20µL of Laemmli sample buffer and 30µL of 8M urea. This was boiled before 5µL was taken and added to 8µL of Laemmli sample buffer and 12µL 8M urea. This is the whole protein fraction SDS-PAGE gel sample.

2.5.2. IMAC purification of proteins

In all cases the protein constructs contained a N-terminal 6His-tag. This was used to purify the protein using immobilised metal affinity chromatography (IMAC). For all purifications a 1mL HisTrap Ni²⁺ column was used in combination with an AKTA liquid chromatography system (GE

healthcare). In all cases the HisTrap column was first equilibrated with the relevant 'buffer A' for the construct being purified.

2.5.2.1. Cytoplasmic domain purification

For purification of the cytoplasmic domain, the cell lysate was loaded onto the HisTrap column at 0.1mL/min, due to its viscosity and to avoid the column being over-pressured. Cytoplasmic buffer B was used to elute bound protein, this was cytoplasmic buffer A with 500mM imidazole added. All steps in the protocol were done so with a flow rate of 0.1mL/min.

After all cell lysate was loaded onto the column, the column was washed with 8 column volumes of buffer A. A 1% gradient of buffer B was then used (5mM imidazole) for 8 column volumes. After this a 100% gradient (500mM imidazole) was used to clean the column for 8 column volumes. 1mL fractions were collected throughout the elution steps.

2.5.2.2. Luminal domain purification

2.5.2.2.1. Basic protocol

For purification of the luminal domain constructs, the cell lysate was loaded onto the HisTrap column at a rate of 1mL/min. Luminal buffer B was used to elute bound protein, this was luminal buffer A with 1000mM imidazole added. All steps in the protocol were done so with a flow rate of 1mL/min.

After all cell lysate was loaded onto the column, the column was washed with 15 column volumes of buffer A with a 2% gradient of luminal buffer B (20mM imidazole) to remove non-specifically bound species. A gradient of luminal buffer B was then used, from 2% to 50% (20mM to 500mM imidazole) over 20 column volumes to elute the bound luminal domain. 1mL fractions were collected over this gradient. After this a 100% gradient of luminal buffer B (1000mM imidazole) was used to clean the column for 10 column volumes. Purification profiles for the luminal domain and its mutants are shown in Figure 2.1.

2.5.2.2.2. S296F purification

For purification of S296F luminal domain buffer A and B contained 1mM TCEP. The same protocol as the basic protocol for purification of the luminal domain was followed otherwise. Purification of the S296F mutation is shown in Figure 2.1.

2.5.2.3. BiP purification

For purification of BiP constructs, the cell lysate was loaded onto the HisTrap column at 1 mL/min. Luminal buffer B was used to elute bound protein, this was luminal buffer A with 1000mM imidazole added. All steps in the protocol were done so with a flow rate of 1mL/min.

After all cell lysate was loaded onto the column, the column was washed with 15 column volumes of buffer A with a 4% gradient of luminal buffer B (40mM imidazole) to remove non-specifically bound species. A 50% gradient of luminal buffer B (500mM imidazole) over 15 column volumes to elute the bound BiP. 1mL fractions were collected over this time. After this a 100% gradient (1000mM imidazole) was used to clean the column for 10 column volumes. The purification chromatogram for BiP is shown in Figure 2.1.

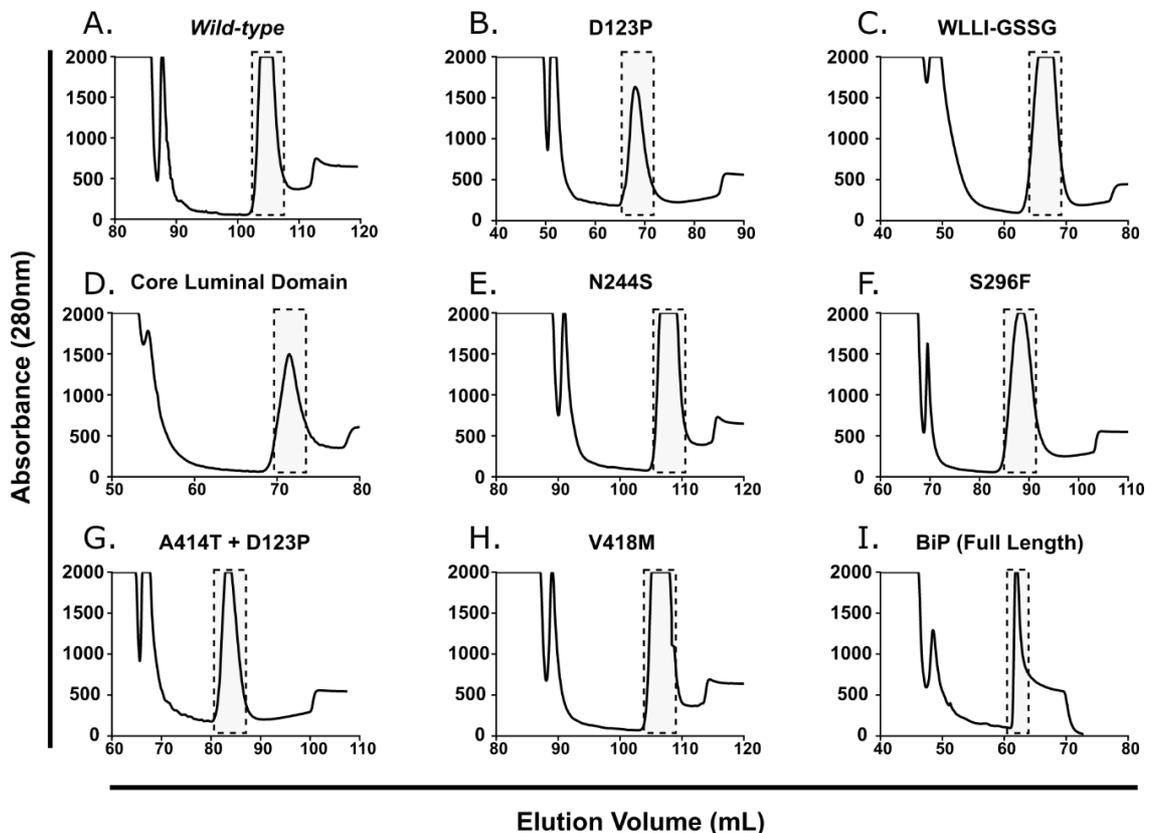


Figure 2.1

The IMAC purification chromatograms for the luminal domain constructs and BiP, as purified using the methods described here. Highlighted are the areas of the chromatogram that were collected, as these contained the purified protein.

2.5.3. A280 concentration

For the luminal domain of IRE1 α and for BiP, where no ATP/ADP was present in the solution the protein concentration was measured using ultraviolet (UV) spectroscopy at 280nm. The protein was diluted 20-fold in 6M guanidinium and measured. The Beer-lambert equation

(Equation 1) was then used to calculate the concentration of the protein. The extinction coefficient (ϵ) was calculated using the Expasy web server (Gasteiger et al., 2003). For IRE1 α luminal domain constructs this was 66935 M⁻¹ cm⁻¹, and for BiP constructs used 17420 M⁻¹cm⁻¹.

$$A = \epsilon lc$$

Equation 1

2.5.4. Bradford assay

Bradford assays were carried out adding 980 μ L Bradford reagent to 20 μ L sample. BSA protein standards at 1, 0.75, 0.5, 0.25, and 0.125mg/mL were used to form a calibration curve for protein concentration. Bradford assays were used for the cytoplasmic domain and for the luminal domain and BiP where there were significant amounts of ATP/ADP in the solution.

2.5.5. SDS-PAGE gels

Samples were prepared by adding 4x Laemmli sample buffer and 8M urea and boiled for 5 minutes. The sample were then run on precast SDS-PAGE gels.

2.5.6. Buffer exchange protocols

2.5.6.1. Dialysis

Dialysis was used after protein purification of the luminal domain and BiP to remove imidazole and to put the protein into a more favourable buffer. The purified fractions of protein were pooled and put into a dialysis tube. The dialysis tube was then put into 500mL of the desired buffer for one hour at 4°C with stirring. After this, the 500mL of buffer was changed to a fresh 500mL of buffer and the dialysis tube was left in this overnight at 4°C with stirring. The next morning the 500mL buffer was changed again, this time to 1L of buffer, and the dialysis tube was added for one hour at 4°C with stirring. After this, typically protein samples were stored after having their concentrations measured. The luminal domain and BiP constructs were dialysed into HMK buffer (20mM HEPES pH 7.6, 100mM KCl, 5mM MgCl₂ with 0.02% NaN₃).

2.5.6.2. NAP5 columns

NAP5 columns were used to exchange the buffer of small amounts of protein at a time, such as before activity assays of the cytoplasmic domain or to remove labels/reducing agents for the luminal domain. The NAP5 column was first equilibrated with a total of 9mL of the desired final buffer. After this 500 μ L of the protein sample was added to the column and allowed to flow into the column. Then 1mL of the desired buffer was added, and a 1mL fraction of the protein flowing through the column was collected.

2.5.7. Concentrating protein

Proteins were concentrated where required using a 10kDa molecular weight cut-off Amicon centrifugal filter unit. These were used as per the manufacturer's instruction until protein was at the concentration desired.

2.6. General analytical protein methods

2.6.1. Mass spectrometry

2.6.1.1. Sequence identification

For sequence determination by mass spectrometry, samples were first run on an SDS-PAGE gel and stained using Coomassie blue (provided by the University of Leeds Mass Spectrometry facility). The band corresponding to the protein of interest was cut out, and the sample submitted to the University of Leeds Mass Spectrometry facility. Here the contents of the gel band were solubilised and incubated with trypsin. The protein fragments from this trypsin digest were analysed by ESI-MS and the fragments used to determine the protein sequence.

2.6.1.2. Molecular mass

To carry out molecular mass determination by mass spectrometry, samples were prepared at 20 μ M and submitted to the University of Leeds Mass Spectrometry facility for analysis by ESI-MS. Shown in Figure 2.2 are the mass spectrometry molecular mass results for the mutant constructs of the luminal domain.

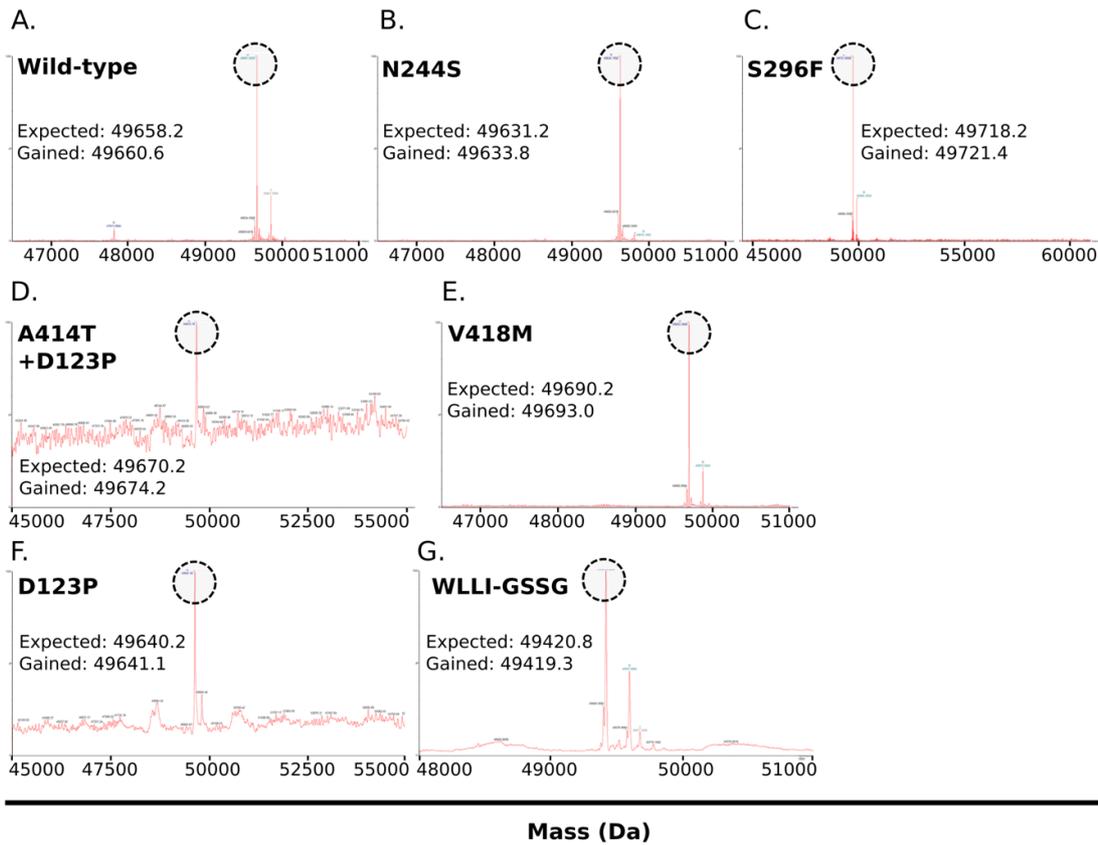


Figure 2.2

Mass spectrometry molecular mass results for each of the mutations of the luminal domain used.

2.6.2. Size exclusion chromatography

2.6.2.1. Methodology

For the luminal domain all size exclusion chromatography was carried out in HMK buffer with 1mM TCEP unless states otherwise. For the cytoplasmic domain all size exclusion chromatography was carried out in the refolding buffer unless stated otherwise. In all cases a Superdex 200 10/300 column was used (GE healthcare Life Sciences). 300 μ L of each sample were injected at the stated concentration. Samples were eluted at 0.2mL/min in the case of the cytoplasmic domain constructs, and at 0.3mL/min for luminal domain and BiP constructs. All experiments were carried out at 14°C.

2.6.2.2. Calibrations

Protein molecular weight standards were used to calibrate the elution profile of different species through the column. These molecular weight standards included ovalbumin (4mg/mL), conalbumin (3mg/mL), aldolase (4mg/mL), ferritin (0.3mg/mL), thyroglobulin (5mg/mL) and blue dextran (0.5mg/mL). These were included in a HMW calibration kit (GE Healthcare Life Sciences). Calibration curves were constructed in each condition used (refolding buffer and HMK

buffer). The calibration curve for the cytoplasmic domain condition is shown in the cytoplasmic domain chapter and the calibration for the luminal domain is shown in Figure 2.3.

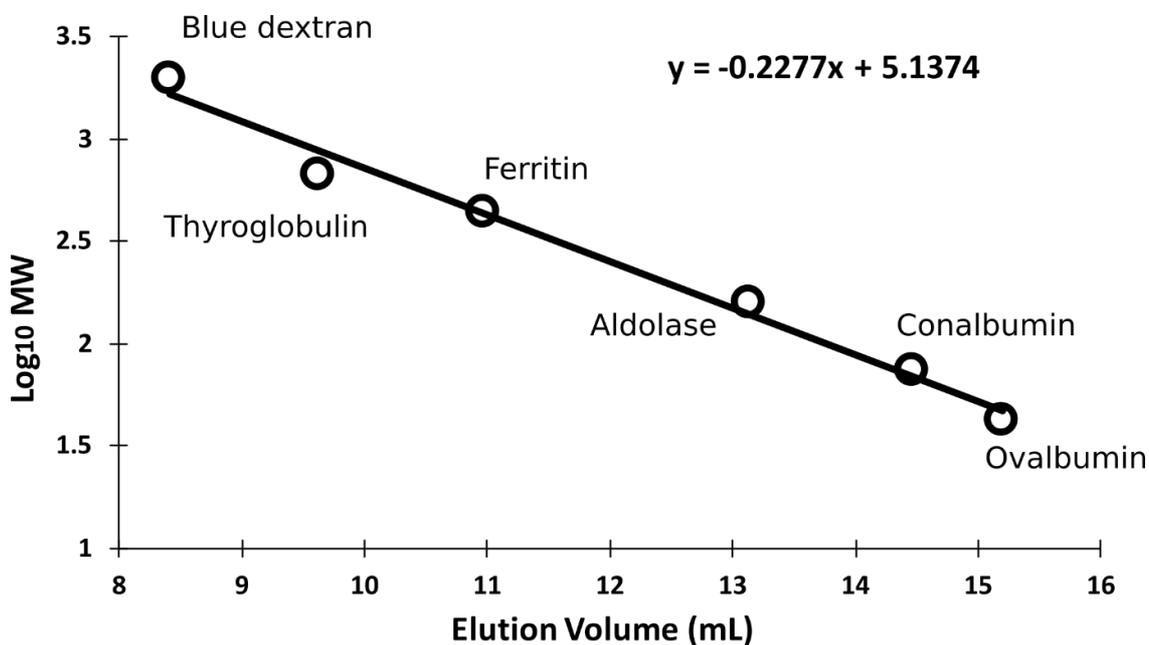


Figure 2.3

Size exclusion chromatography molecular weight calibration in HMK buffer + 1mM TCEP. Log₁₀ molecular weights used for a linear calibration. Annotated are the species used to construct the calibration.

The elution volume of the D123P luminal domain construct was assumed to give monomeric protein and so the calibration curve in Figure 2.3 was used to calculate the apparent mass of this species, the size of other luminal domain multimers were calculated from this mass. In order to determine the concentration of luminal domain protein on the size exclusion column equation 2 was used. [Protein] is the concentration of protein eluted, V_i is the volume of protein injected onto the column, V_{eluted} is the volume of the eluted peak and C_i is the concentration of protein injected onto the column.

$$[Protein] = \frac{V_i}{V_{eluted}} \times C_i$$

Equation 2

2.6.3. Circular Dichroism

For circular dichroism analysis of proteins, the protein was first buffer exchanged using a NAP5 column to the circular dichroism buffer (10mM sodium phosphate dibasic pH8.0, 150mM NaCl, 1mM TCEP) and diluted to a concentration of 0.2mg/mL. The Circular dichroism spectra was measured from 180nm to 260nm with 0.5nm intervals.

2.7. Nuclear magnetic resonance

Unless stated all experiments were carried out using a 950MHz Bruker Ascent Aeon magnet with a TCI-cryoprobe (3mm). 5% of D₂O was added to samples before they were added to the NMR tube. All experiments were run at 288K unless stated otherwise. In all cases the collected spectra were processed using NMRPipe and analysed using CCPN analysis version 2 (Delaglio et al., 1995; Vranken et al., 2005).

2.7.1. Amide TROSY experiments

In all cases ¹⁵N ¹H BEST TROSY experiments were performed (Pervushin et al., 1997). Chemical shift perturbations were calculated from *wild-type* samples to the condition used with Equation 3 (Williamson, 2013), where H and N represent the nitrogen and hydrogen chemical shifts.

$$\text{Chemical shift perturbation} = \sqrt{(\Delta H)^2 + (0.14\Delta N)^2}$$

Equation 3

Significant chemical shift perturbations from a standard chemical shift are defined as chemical shift perturbations $>2\sigma$ of the average chemical shift perturbations for all peaks (All chemical shift perturbations shown in Section 9, Figures 9.7 and 9.8).

Significant changes in protein dynamics are defined through comparison of the peak intensities in units of signal to noise (S/N), where $>2\sigma$ or $>25\%$ from the reference peak is defined as a significant change. S/N calculated by use of the recorded signal and the noise level as determined by NMRpipe (Delaglio et al., 1995).

2.7.1.1. Cytoplasmic domain denatured and refolded

The cytoplasmic domain at 57μM in 6M urea and also after being refolded and concentrated to 60μM were analysed. The cytoplasmic domain samples were analysed using a TXO-cryoprobe (5mm) in a shaped NMR sample tube (Bruker).

2.7.1.2. Luminal domain constructs

For analysis of luminal domain constructs, the desired constructs were made to 50μM with HMK with 1mM DTT and 5% D₂O.

2.7.1.3. Truncation mutant assignment

For assignment of the truncation mutants, all constructs used were made to a concentration of 40μM. Analysis through CCPNMR allowed for the partial assignment of the linker region (Vranken et al., 2005).

2.7.1.4. Temperature correlation

For temperature correlation experiments a 750MHz Oxford NMR spectrometer was used with a TCI-cryoprobe. Here *wild-type* luminal domain was at a final concentration of 50 μ M in a shaped NMR sample tube (Bruker). Experiments with the same sample were carried out at 5°C, 10°C, 20°C, 25°C and 30°C sequentially. Data was analysed using CCPNMR version 2 (Vranken et al., 2005) to find the ^1H chemical shift of each peak at each temperature. These were then plotted against the temperature and a linear fit calculated to give the chemical shift temperature correlation value (Okazaki et al., 2018).

2.7.1.5. Luminal domain with peptide experiments

For the experiments with the addition of peptide the luminal domain was run at a concentration of 36 μ M alone. For three hours 36 μ M of the luminal domain was incubated at 15°C with 36 μ M of ΔEspP , then spun down at 14,000rpm for 1 minute and its concentration measured before it was analysed using NMR. The signal/noise of the spectra was adjusted to the measured concentration. For the WLLI-GSSG experiments, 50 μ M of protein and each peptide were used in separate experiments, otherwise using the same conditions as before.

2.7.2. Solid-state NMR

^{15}N and ^{13}C IRE1 α luminal domain was prepared at 42 μ M in HMK buffer with 5mM DTT in two separate samples. To both samples 205.5 μ M of ΔEspP peptide was added. The samples were incubated at room temperature for three hours. To one sample BiP protein was added to a final concentration of 40 μ M and to the other an equal volume of buffer. Both samples were left overnight at 4°C. The samples were then centrifuged at 14,000rpm for 2 minutes to pellet the insoluble fractions. The supernatant was removed and fresh HMK buffer with 5mM DTT was added over the pellet. The samples were then shipped on ice to József Lewandowski at the University of Warwick who ran all experiments on the samples.

2.8. Cytoplasmic domain methods

2.8.1. Refolding techniques

2.8.1.1. Dilution refolding (final protocol)

The cytoplasmic domain construct purified from inclusion bodies in guanidinium was refolded through dropwise addition to refolding buffer (50mM HEPES pH 8.0, 500mM arginine, 300mM NaCl, 5mM TCEP and 10% glycerol) at 4°C to a concentration of 2.5-5 μ M and subsequently gently stirred for one hour.

2.8.1.2. Column refolding

For column refolding a NAP5 column was used to buffer exchange the unfolded cytoplasmic domain in cytoplasmic buffer A/B into the desired refolding buffer with a starting concentration of 5 μ M and therefore an expected elution concentration of 2.5 μ M.

2.8.1.3. Dialysis refolding

For dialysis refolding of the cytoplasmic domain, the unfolded cytoplasmic domain in cytoplasmic buffer A/B was added to a dialysis tube and submerged in 1L of the desired buffer overnight at 4°C, in the morning the buffer was replaced by 500mL of fresh buffer and the protein was dialysed for a further two hours at 4°C.

2.8.2. Optim 1000

A sample of the refolded cytoplasmic domain buffer exchanged to circular dichroism buffer at 3 μ M was provided to the University of Leeds Biomolecular Interactions facility. Here Iain Manfield carried out the Optim 1000 protocol. The samples were diluted into the various buffer conditions trialled in a 16-place array (Unchained Labs). These included MES 100mM (pH 5.5), MES 100mM (pH 6.5), HEPES 100mM (pH 7.5), bicine 100mM (pH 8.5), N-Cyclohexyl-2-aminoethanesulfonic acid (CHES) 100mM (pH 9.5), addition of glycerol at 5%, 7.5%, 15%, 20% and 30%, NaCl to a concentration of 75mM, 150mM, 250mM, 500mM, 1000mM and 2000mM. Unfolding and aggregation of the samples was assessed between 15°C and 90°C with 1°C steps and 30 seconds incubation using an Optim 1000 (Unchained Labs). The instrument settings were 1000ms exposure time with 380nm centre wavelength and 100 μ m slit width. Results were analysed using Optim analysis software version 6.30.

2.8.3. Dynamic light scattering

For dynamic light scattering experiments the cytoplasmic domain was refolded and buffer exchanged to the desired buffer for use. The protein was then centrifuged at 14,000rpm for 1 minute and the concentration was measured. The sample was then filtered through a 0.22 μ M filter. A Wyatt miniDawnTreos system was used. Buffer was injected, and equilibrium was reached over 5 minutes, then the sample was injected, and the following three minutes of data was used for analysis. Between measurements the system was washed with 1M nitric acid and dH₂O. Astra software was used to analyse the data using the cumulants model (Dobson et al., 2017).

2.8.4. Kinase assay

Kinase activity assays were carried out using the ADP-Glo kinase assay kit (Promega). The cytoplasmic domain was first refolded as per the optimised method before being buffer exchanged using a NAP5 column to circular dichroism buffer. After this the protein concentration was measured using a Bradford assay. The protein was diluted to 500nM in kinase buffer (20mM Tris pH 7.4, 500mM NaCl, 10% glycerol and 1mM TCEP) with 500 μ M ATP in a volume of 5 μ L. Reactions were quenched after the desired time as per the manufacturers' instructions with 5 μ L of ADP-Glo Reagent, the reactions were then incubated for 40 minutes at room temperature before 10 μ L of Kinase Detection Reagent was added. The reactions were then incubated at room temperature for one hour before the luminescence of the reactions was measured. A calibration of luminescence to ATP hydrolysed was created as per the manufacturer's protocol. The calibration is shown in Figure 2.4.

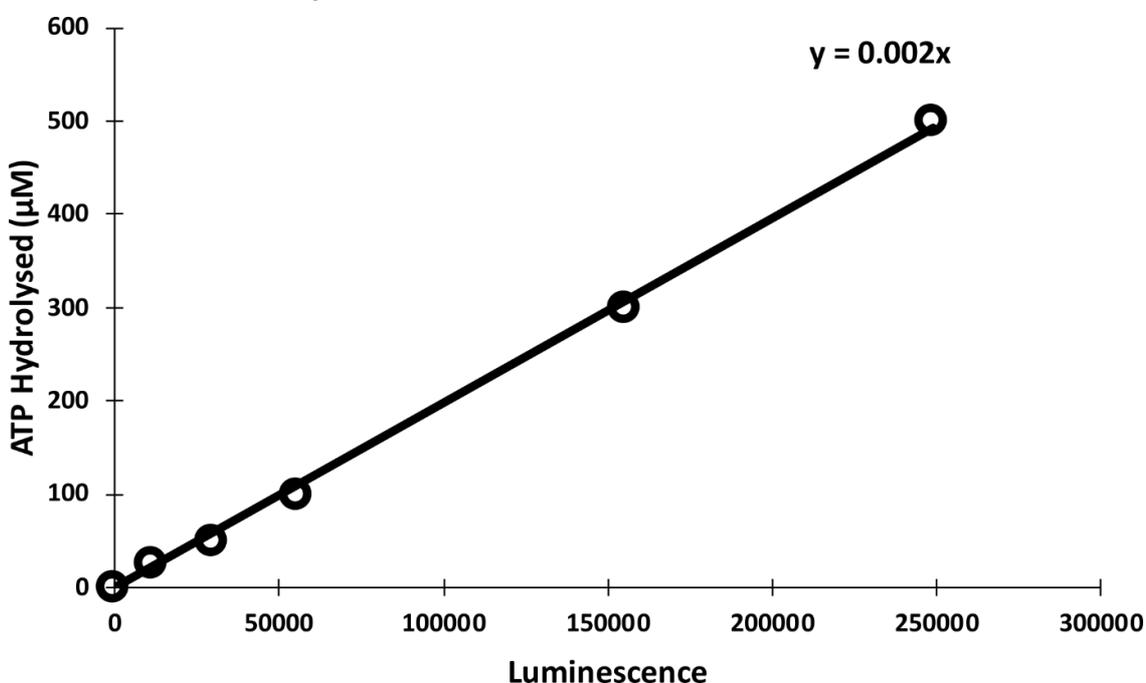


Figure 2.4

Calibration for the ADP-Glo kinase kit. ATP hydrolysed is plotted against luminescence.

2.8.5. Autophosphorylation assay

For autophosphorylation assays, the cytoplasmic domain was refolded before being incubated with 5mM ATP and 5mM MgCl₂ at room temperature for one hour to promote autophosphorylation. After this the sample was frozen and sent to the University of Leeds Mass Spectrometry facility for molecular mass determination.

2.8.6. Endoribonuclease assays

For all assays measuring endoribonuclease activity of the cytoplasmic domain, the protein was refolded, and buffer exchanged to circular dichroism buffer. After this the protein was diluted in endoribonuclease buffer (20mM HEPES pH 8.0, 50mM KOAc, 0.5mM MgCl₂, 10% glycerol and 1mM TCEP) to the desired concentration. Where ADP or ATP were required, the protein was incubated with 5mM of ATP or ADP at room temperature for one hour with a final concentration of 5mM MgCl₂, and fresh ATP/ADP was used in the final endoribonuclease reaction. An RNA oligomer was used with fluorescent and black hole quencher 1 groups (FAM-GAA CAA GAU AUC CGC AGC AUA UAC AGU UC-BHQ1). Cleavage of the RNA substrate was measured by fluorescence (excitation 494nm, emission 520nm).

2.8.6.1. Time dependent assays

The time dependent assays included 100nM protein and 100nM FRET substrate, they were run over night at 25°C. Here the splicing activity of the refolded *E. coli* protein was compared to eukaryotic produced protein in APO, ATP or ADP states.

2.8.6.2. Concentration dependent assays

The concentration dependent assays were measured at varying time points with different amounts of refolded cytoplasmic domain incubated with 25nM of the FRET RNA substrate at 30°C. RNase A was used to construct a calibration curve of the amount of the FRET RNA substrate cleaved.

2.8.7. TEV cleavage reaction

TEV protease was produced and purified in the lab by previous lab member Sam Dawes. To cleave the Histidine tag from the cytoplasmic domain, TEV protease was incubated with the refolded cytoplasmic domain at an A280 ratio of 1:100 TEV protease to cytoplasmic domain. The reaction included 0.5mM EDTA and 1mM DTT and was carried out overnight at 4°C. The reaction was subsequently concentrated, including dilution steps to remove EDTA from the solution.

2.9. Luminal domain methods

2.9.1. Differential scanning calorimetry

A MicroCal DSC calorimeter used. Protein at 25µM in HMK +1mM TCEP buffer was measured, with the reference cell filled with HMK +1mM TCEP only. A temperature scan from 12°C to 90°C was used and data analysed using Origin.

2.9.2. Preparation of peptides

The peptides used:

Δ EspP- MKKHKRILALCFLGLLQSSYSAAKKKK (Gardner and Walter, 2011)

MPZ1- LIRYCWLRRQAALQRRISAME (Karagoz et al., 2017)

All peptides here were purchased from BioMatik and resuspended in HMK buffer at a concentration of 2mM.

2.9.3. FITC-labelling of the luminal domain

To label the luminal domain constructs of IRE1 α with FITC, the protein was prepared at 25 μ M in 500 μ L of HMK buffer. 50 μ M of 1M sodium carbonate at pH 8.0 was then added. 1mg of FITC powder was dissolved in 1mL DMSO to give a 1mg/mL solution. 9.6 μ L of this was added to the luminal domain protein sample and left on ice for 60 minutes. A NAP5 column was then used with HMK to remove excess FITC. The concentration of the eluted protein was then measured by spectroscopy, measuring absorbance at 280nm and 495nm. The final concentration of labelled protein was calculated using the Beer Lambert equation after a correction for the A_{280} value using Equation 4.

$$A_{280_{Corrected}} = A_{280} - 0.35(A_{495})$$

Equation 4

2.9.4. Fluorescence polarisation

50nM of FITC-labelled luminal domain with serial dilutions of the desired peptide were used in HMK buffer +1mM TCEP, samples were incubated for 30 minutes at room temperature before being measured. Fluorescence polarisation was measured using a BMG POLARstar OPTIMA plate reader with excitation 485nm and emission 525nm wavelengths. Polarisation was calculated using Equation 5.

$$Polarisation = \frac{(Fluorescence_{parallel} - Fluorescence_{perpendicular})}{(Fluorescence_{parallel} + Fluorescence_{perpendicular})}$$

Equation 5

2.9.5. Microscale thermophoresis

FITC-labelled luminal domain at 500nM had a dilution series of the desired protein added to it, all in HMK buffer +1mM TCEP. Each condition was then added to a capillary and measured using a NanoTemper Monolith NT 1.15. Data was analysed using NanoTemper Analysis software.

2.9.6. Turbidity assays

2.9.6.1. OD400

In these assays, a final concentration of 20 μ M of the luminal domain construct was used in HMK buffer +1mM TCEP. The desired concentration of peptide was added, and the reaction incubated at 30°C with OD₄₀₀ being measured every three minutes for 12 hours.

2.9.6.2. Soluble concentration determination

To determine the soluble concentration of the luminal domain reactions were started with 20 μ M of IRE1 α and different concentrations of peptide. After three hours (or different amounts of time if stated) the reaction was centrifuged at 14,000rpm for 1 minute. A Sample of the supernatant was then taken, and the concentration measured. Gel samples were also taken of the supernatant and the pellet (after washing three times in HMK buffer). The amount of insoluble protein was calculated for each condition and compared to the turbidity assay of the same conditions in order to create a calibration curve (Borgia et al., 2013).

2.9.7. BiP chaperone activity assay

The BiP chaperone activity assay was carried out as with the turbidity soluble concentration determination assay. 20 μ M of the IRE1 α construct was incubated initially with 169 μ M of Δ EspP for three hours at room temperature. After this 24 μ M (unless stated otherwise) of the desired BiP construct was added and incubated at room temperature for three hours. Then 40mM of ATP was added and the reaction was incubated at room temperature for three hours. Each condition was carried out in triplicate and each the samples were centrifuged at 14,000rpm for 1 minute before the protein concentration of the supernatant was measured using a Bradford assay. Gel samples were also taken for the supernatant and pellet. Control measurements of BiP alone were used to calculate the amount of luminal domain in the soluble fraction.

2.9.8. Negative stain electron microscopy

For electron microscopy (EM) the luminal domain at a final concentration of 20 μ M was incubated with 75 μ M of Δ EspP for two hours, or with Δ EspP for one hour and with 24 μ M BiP and 40mM ATP for one hour. The samples were then negative stained, first by being diluted 10-fold. 3 μ L of the diluted reaction was taken and added to a glow discharged carbon coated copper grid (produced at the University of Leeds Astbury BioStructure Laboratory) for 30 seconds, before being washed with 5 μ L of dH₂O for 3 seconds, with excess liquid being blotted away in between steps. 5 μ L of 2% uranyl acetate was then added to the grid and left for 3 seconds, this step was repeated once more. Lastly 5 μ L of 2% uranyl acetate was added to the grid again, but for 30

seconds this time before being blotted away extensively. Images were taken in the University of Leeds Astbury BioStructure Laboratory using a FEI T12 microscope with a Lab6 filament and Gatan UltraScan 4000 CCD camera.

3. Expression and characterisation of hIRE1 α 's cytoplasmic domain in *E. coli*

In order to further understand the cytoplasmic domain's activation, in particular how the protein's dynamic endoribonuclease domain (Concha et al., 2015) is able to change from an adaptive to an apoptotic response, different conformations and the dynamic rearrangements of the protein must be characterised. This can be achieved using methods such as X-ray crystallography, hydrogen deuterium exchange experiments, single molecule fluorescence resonance energy transfer (FRET), cryo-EM and NMR. These methods can provide information about conformational states of the protein and dynamic processes occurring for the protein to adopt these states, bind ligands and carry out enzymatic activity. NMR is especially apt for characterising protein ms-ps timescale dynamics because the technique monitors the excited states of nuclei which are preserved in the ms timescale, changes in the environment of these nuclei from faster dynamic processes cause them to return to equilibrium sooner, and these changes can be monitored on a residue specific basis (Kleckner and Foster, 2011). NMR is therefore an ideal technique to study the cytoplasmic domain. However, even with sensitivity advances in the field, the technique requires a large amount of isotopically labelled protein and therefore prokaryotic expression systems such as *E. coli* are required for improved protein yields and labelling strategies (Tugarinov et al., 2006), as described in Section 1.6.1.3.

There is currently no published protocol for the expression of human IRE1 α 's cytoplasmic domain in *E. coli*. Producing recombinant proteins in prokaryotic systems is not a trivial task and requires much optimisation especially for a protein including both kinase and endoribonuclease activities as in this case. However, there are numerous ways to create favourable conditions for soluble expression of proteins from *E. coli* cultures including changing the expression temperature, the level of expression induction and the *E. coli* strain used among others. Ways to achieve soluble expression of recombinant proteins in *E. coli* expression systems have been extensively reviewed but the overarching conclusion is that although certain conditions appear beneficial in some cases, there is no formula to predict the perfect condition and many conditions will have to be trialled for each specific protein (Peti and Page, 2007; Rosano and Ceccarelli, 2014). Where it appears impossible to find the condition for soluble recombinant expression of a protein, it is possible to purify the protein from the insoluble fraction and attempt to refold the protein in favourable conditions (Palmer and Wingfield, 2004). Buffer additives and techniques for refolding proteins have also been extensively reviewed, and as with recombinant protein solubilisation, conditions for the correct refolding are protein dependent (Yamaguchi and Miyazaki, 2014; Burgess, 2009).

In this chapter a cytoplasmic domain construct suitable for *E. coli* expression will be created and its expression subsequently optimised. Due to the protein being produced in inclusion bodies, the objective will be to isolate the construct, purify it and refold the protein to its native state. After refolding, as is standard in such studies, the enzymatic activities of the refolded cytoplasmic domain can then be verified.

3.1. Subcloning of the cytoplasmic domain construct

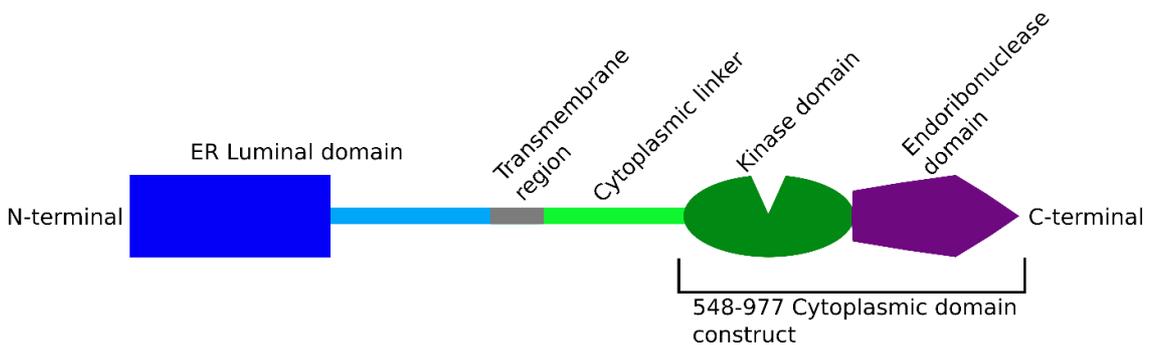


Figure 3.1

Representation of human IRE1 α . The domains of the protein in the schematic are labelled. Shown is the cytoplasmic domain construct used in this chapter is shown (residues 548 – 977). Full sequence of the cloned construct in Appendices Figure 9.2.

Genetic subcloning has become a necessity in the study of proteins, made feasible by the invention of the polymerase chain reaction (PCR) (Mullis, 1990). Gene sequences are now manipulated in numerous ways, such as single point mutations, truncating proteins, extending proteins and importantly the desired sequences can be cloned into plasmid vectors. Once the sequence of interest is in the desired plasmid vector, the plasmid can be added to a host organism, in this case *E. coli*, and the manipulated DNA sequence can be transcribed for protein expression.

As described in Section 1.2.1, IRE1 α is a transmembrane protein, which presents further challenges in characterisation and use of the protein *in vitro* techniques and so the protein has been separated into its luminal and cytoplasmic domains. This has been made possible using PCR to extract the desired sequences for each domain from the complete protein sequence. Figure 3.1 shows a schematic of IRE1 α and the cytoplasmic domain construct selected for use here (residues 548-977). As the aim of this chapter is to express the cytoplasmic domain in a bacterial system and compare it to protein expressed in a eukaryotic system, the domain's linker region is excluded as the majority of structural and functional studies thus far have excluded it also (Concha et al., 2015; Joshi et al., 2015).

3.1.1. Ligation independent subcloning principle

Commonly, restriction enzymes are used to cut specific regions of DNA to leave overhangs that can be joined by a ligase in order to insert genes of interest into plasmid vectors. This method requires information about the vector and insert in each case, and it can be challenging to find the correct restriction enzyme sites in the target DNA to facilitate insertion. The ligase reaction can also produce false positives through ligation of empty vectors. Due to these disadvantages, various methods that do not use a ligation reaction have been created. One such method is ligation independent cloning (LIC) (Figure 3.2) (Hsiao, 1993; Aslanidis and de Jong, 1990).

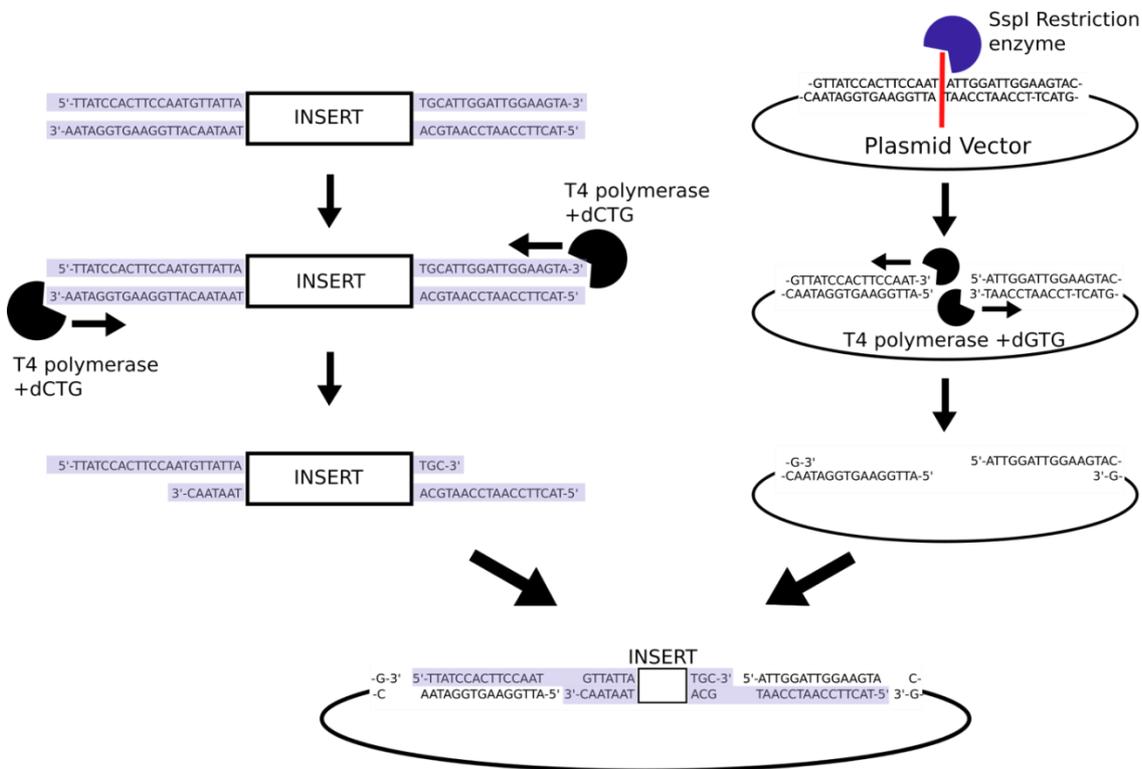


Figure 3.2

A schematic for ligation independent cloning. Shown are the processes applied to the 'insert' which is the gene of interest and the plasmid vector which the insert is ligated into for expression. Shown at the foot of the figure is the eventual ligation reaction, where the complementary vector and insert 'sticky ends' are brought together (Section 2.3.4).

For LIC, the DNA fragment of interest is created through PCR with the addition of 18 nucleotide bases 3' and 5' of the desired DNA sequence. The 3' – 5' exonuclease activity of T4 DNA polymerase is then able to remove bases from the 3' end. If dCTP is present in the reaction, then the T4 polymerase will stop removing bases when it encounters cytosine, as its polymerase activity will be in equilibrium with its exonuclease activity. Using this method, stable and

complimentary sticky ends can be formed between the insert and vector sequences by addition of only dGTP to the vector T4 polymerase reaction and dCTP with the insert reaction. The vector and insert can then be mixed to allow annealing (Figure 3.2), before transformation into *E. coli* cells.

LIC confers the ability to clone the fragment of interest into multiple different plasmid vectors, providing those vectors are LIC compatible. In this way multiple different constructs can be produced simultaneously from the same DNA region of interest, which is ideal when optimising a protein's expression and therefore this method was selected for the subcloning of the cytoplasmic domain.

3.1.2. Optimising the LIC reaction for cytoplasmic domain constructs

In order to perform LIC with the human cytoplasmic domain, primers were first designed for the isolation of the DNA sequence of residues 548-977 of the cytoplasmic domain in a PCR reaction with the correct end sequences to facilitate ligation independent cloning (Section 2.3.1.1). The plasmids selected were all pET plasmids that include a N-terminal 6His tag and a TEV cleavage site in the reading frame, allowing for immobilised metal affinity chromatography (IMAC) purification of the construct, and cleavage of purification and solubility tags by TEV protease. Plasmid 1B did not include a solubility tag, plasmid 1M included a MBP solubility tag N-terminal to the TEV cleavage site and 1G included a GST tag here (Appendices Figure 9.1). For all plasmids a ratio of 6:1 insert to plasmid DNA allowed successful ligation (Section 2.3.4.2). Shown in Figure 3.3 is an agarose gel with the DNA insert and plasmid vectors alone and after successful ligation reactions. Each sample was incubated with the *SspI* restriction enzyme (Section 2.3.2), if the insert was successfully ligated, the restriction enzyme is unable linearise the plasmid. If the plasmid is linearised, the gel band will run as a single line on the agarose gel. If the vector remains circular, it means that there is no *SspI* restriction site, as the insert was successfully annealed, and the gel band will appear curved and will not run at its correct size, when comparing to molecular weight standards. In this case the completed annealing reactions have a different appearance to the vector plasmids alone, suggesting that the reactions were successful.

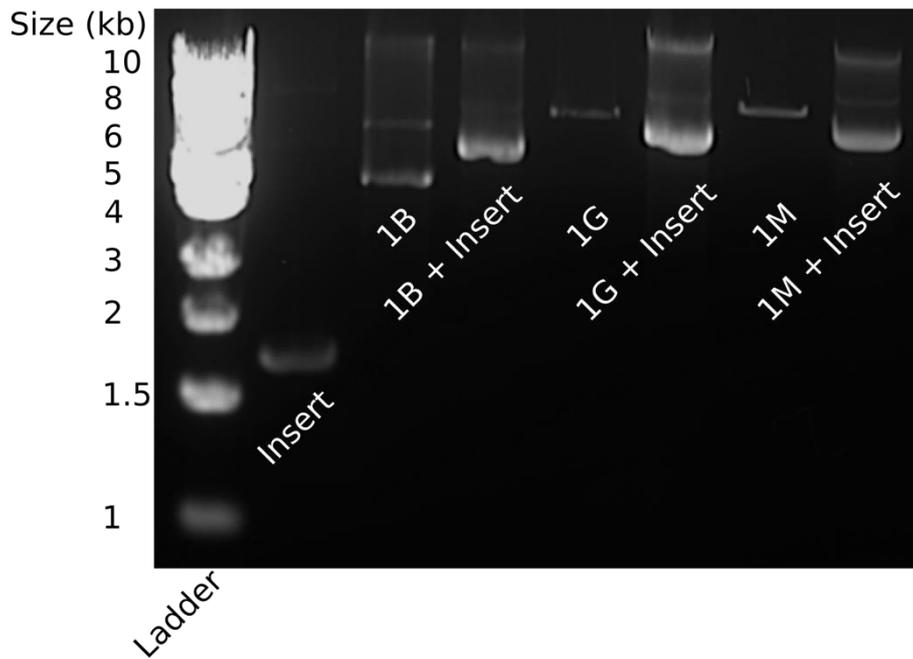


Figure 3.3

Agarose gel for the constructs of the cytoplasmic domain created. The cytoplasmic domain insert (548-977) and the plasmid vectors used (1B, 1G, 1M; Appendices, Figure 9.1) with and without the insert successfully cloned into them are analysed on the agarose gel (Section 2.3.3). All plasmids have been incubated with the *SspI* restriction enzyme (Section 2.3.2). Without the inserts the plasmids are linear and run as a single line. Plasmids containing the insert are not linearised, as they do not possess *SspI* cleavage sites and therefore don't appear as a single line.

The presumed successful reactions were validated by sequencing reactions displaying the expected cytoplasmic domain construct (Appendices Figure 9.3) before their expression in *E. coli* could be trialled.

3.2. Optimisation *E. coli* recombinant overexpression of the cytoplasmic domain

3.2.1. Strategy to optimise protein solubility and yield

The cytoplasmic domain construct was initially expressed in *E. coli* at 37°C, however, all expressed cytoplasmic domain protein was found to be in the insoluble fraction, as determined by a solubility assay (Section 2.5.1). This isn't unexpected as only a small portion of recombinant proteins are expressed as soluble and correctly folded species in *E. coli* (Braun et al., 2002). Although insoluble, it was promising that the construct was expressed as some recombinant proteins cause toxicity to expression systems leading to bacterial cell death (Dumon-Seignover et al., 2004). Therefore, with the aim of optimising expression yield and solubility multiple different growth conditions were trialled (Figure 3.4).

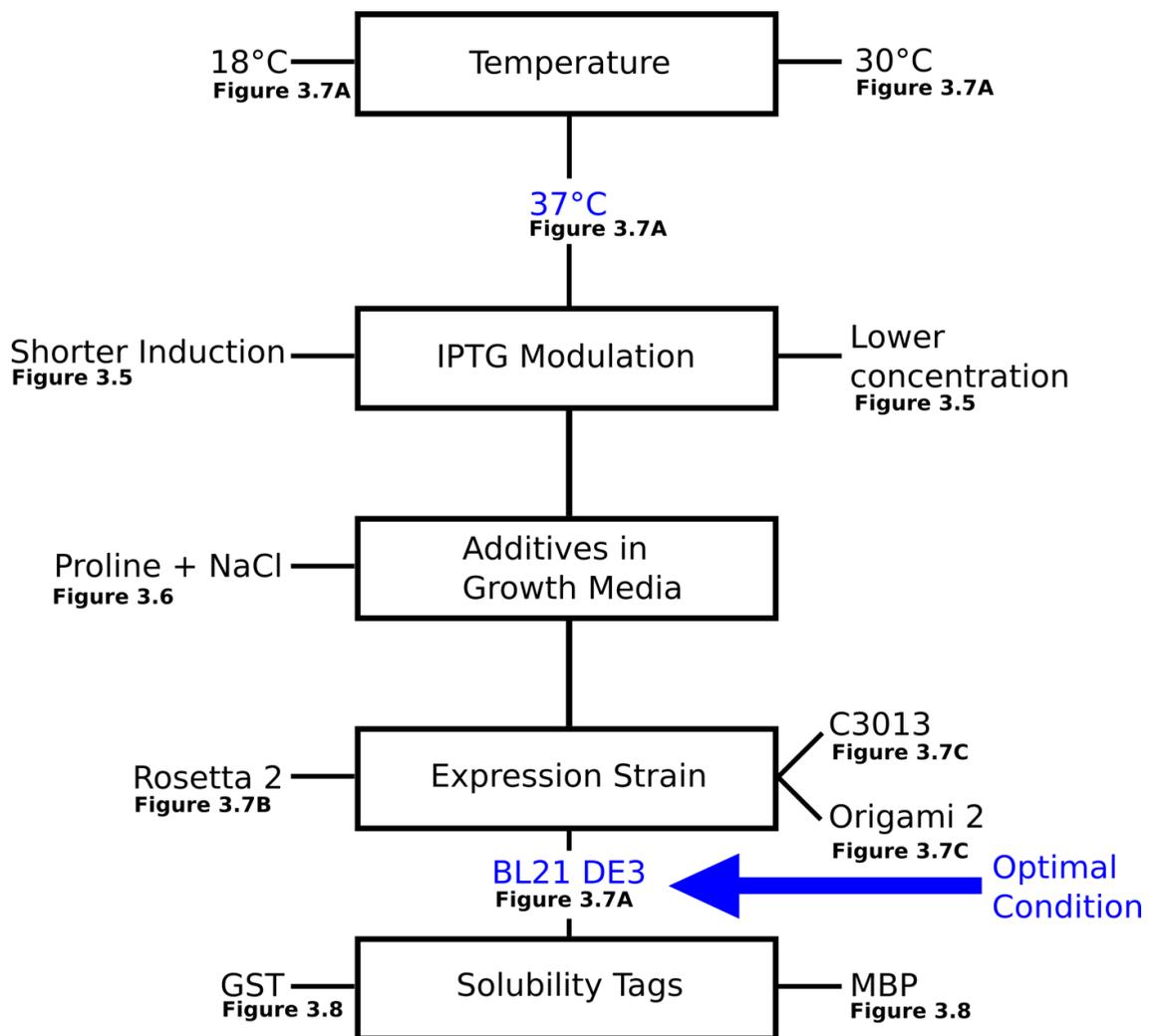


Figure 3.4

Schematic for expression conditions trialled to achieve a high yield of protein expression. Different conditions are labelled with 'Figure x', corresponding which SDS-PAGE gel figure their solubility test is displayed in.

For each condition a small-scale expression and solubility test were therefore carried out, the results of which are visualised in a series of SDS-PAGE gels where the cytoplasmic domain appears close to the 50kDa molecular weight marker. The amount of protein produced and if any condition produces soluble IRE1 α can be observed. Here numerous conditions were used, none of which produced soluble protein, however, the condition to give optimal protein yield could be identified.

3.2.2. Assessment of *E. coli* expression conditions

3.2.2.1. Temperature

Lower temperatures during expression of recombinant proteins can slow the rate of expression and therefore improve solubility by reducing the metabolic burden on the bacteria (Hoffmann and Rinas, 2004; Schein, 1989). For this reason, the different expression conditions trialled were mostly carried out at 37, 30 and 18°C (Figures 3.6-3.8). In some cases, the lower expression temperatures appear to decrease the protein yield, but in no condition does the change in temperature improve solubility of the cytoplasmic domain. Consequently, due to an apparent small increase in yield, 37°C was chosen as the final condition for expression of the cytoplasmic domain.

3.2.2.2. IPTG concentration

Here, T7 promoter expression systems are used, which require IPTG to promote expression of T7 RNA polymerase, which in this case then transcribes the cytoplasmic domain. Similar to reducing the expression temperature, decreasing the concentration of IPTG used to induce expression may lower translation rates and the amount of protein produced. Decreasing the length of time of IPTG induction may also prevent overcrowding of the cell during protein production to promote soluble expression. However, decreasing the concentration and duration of IPTG induction did not affect the solubility of the cytoplasmic domain (Figure 3.5). Additionally, prolonged bacterial growth (30 hours) with no IPTG induction was also attempted as this can produce low levels of ‘leaky expression’ of the protein (Zhang, Z. et al., 2015). However, no IRE1 α appears to be expressed under these conditions. As reduction of IPTG induction does not induce soluble expression, 1mM of IPTG was used for four hours in subsequent expression of the cytoplasmic domain to maximise yield.

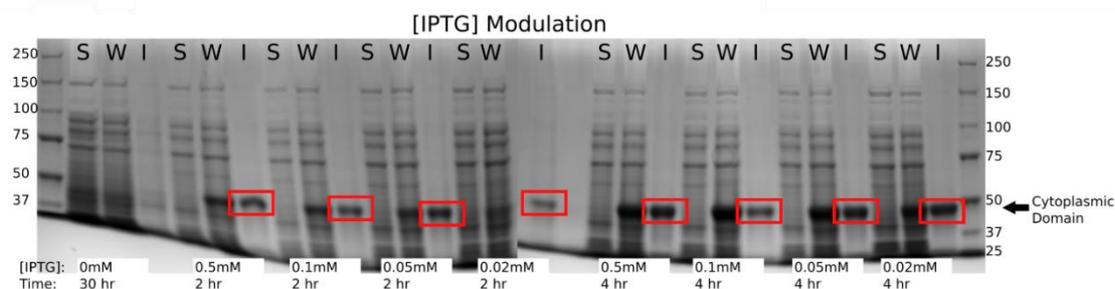


Figure 3.5

SDS-PAGE gel of solubility tests carried out for different *E. coli* expression conditions of the IRE1 α (548-977) construct. S = soluble fraction, W = Whole protein and I = Insoluble fraction.

Here, the construct was expressed in BL21 DE3 cells at 37°C for the indicated time and concentration of IPTG, red boxes show insoluble cytoplasmic domain expressed (Section 2.5.1).

3.2.2.3. Small molecule additives

Cells contain a high concentration of macromolecules and loss of water can increase this concentration further, creating a crowded environment that isn't beneficial for protein folding. Proline is a natural osmoprotectant and also discourages interaction between folding intermediates through favourable interactions with protein side chains, similar to the action of NaCl reducing electrostatic free energy of side chains, and can therefore aid protein folding (Ignatova and Gierasch, 2006; Burgess, 2009). To increase expressed cytoplasmic domain solubility 20mM proline and 300mM NaCl were added to the expression condition (Figure 3.6), however, this did not yield soluble cytoplasmic domain when expressed in BL21 DE3 cells. Therefore, these additives were not included in the final expression condition.

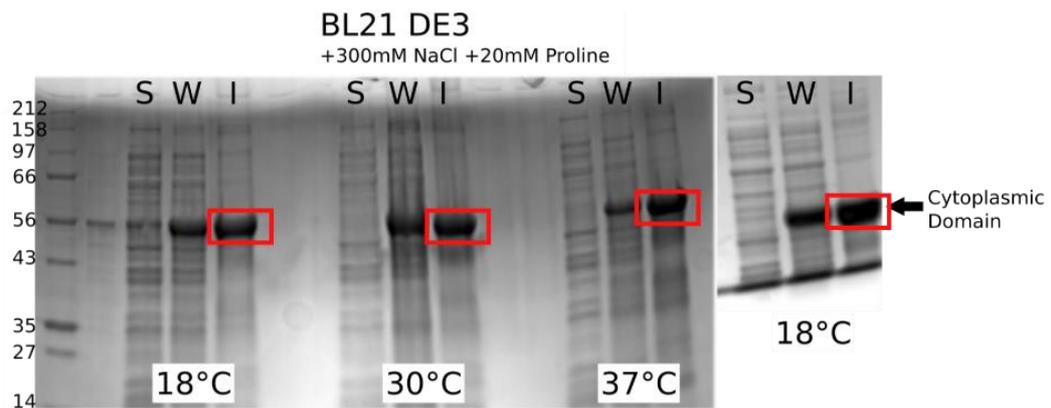


Figure 3.6

SDS-PAGE gel of solubility tests carried out for different *E. coli* expression conditions of the IRE1 α (548-977) construct. S = soluble fraction, W = Whole protein and I = Insoluble fraction. Expression in BL21 DE3 cells at the indicated temperatures for four hours with 1mM IPTG with 300mM NaCl and 20mM proline in the media, red boxes show insoluble cytoplasmic domain expressed (Section 2.5.1).

3.2.2.4. *E. coli* expression strains

Different *E. coli* expression strains were used in an attempt to produce soluble cytoplasmic domain protein (Figure 3.7). The BL21 DE3 strain is optimised for protein expression by being deficient in proteases and having tight control over protein expression through use of the T7 promoter system and is considered the standard T7 expression strain for routine protein expression (Choi et al., 2006). The Rosetta 2 *E. coli* strain is derived from BL21 cells but include

tRNAs for rare *E. coli* codons that are used by eukaryotic cells, so can optimise translation of recombinant proteins. The C3013 strain is also derived from BL21 DE3 and includes higher levels of control over expression and promotes the expression of otherwise toxic recombinant proteins. The Origami expression strain is a derivative of the K-12 strain. The K-12 strain is considered to give a lower yield than BL21 DE3 strains, and therefore may prevent aggregation from increased amounts of the cytoplasmic domain (Blattner et al., 1997; Wu et al., 2010). Here each strain was used to express the cytoplasmic domain construct, although all strains give good levels of recombinant expression of the cytoplasmic domain, none promote its soluble expression (Figure 3.7). As there appear to be insignificant differences between different *E. coli* strains, the BL21 DE3 strain was used subsequently.

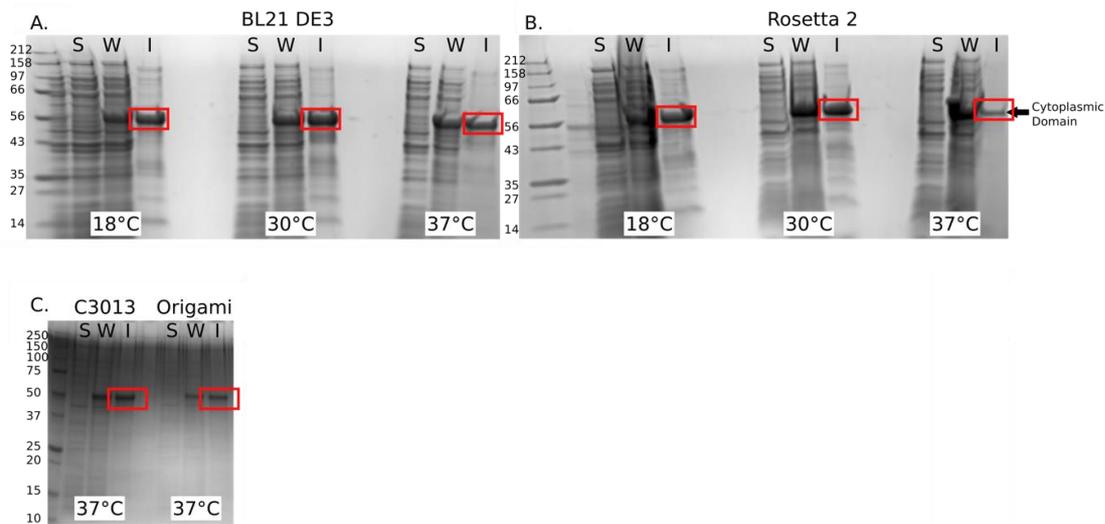


Figure 3.7

SDS-PAGE gel of solubility tests carried out for different *E. coli* expression conditions of the IRE1 α (548-977) construct. S = soluble fraction, W = Whole protein and I = Insoluble fraction. Expression in different *E. coli* expression strains at the indicated temperatures for four hours with 1mM IPTG, red boxes show insoluble cytoplasmic domain expressed (Section 2.5.1).

3.2.2.5. Solubility tags

Fusion of recombinant proteins with soluble proteins can promote correct folding, and there have been numerous successes using this technique (Kapust and Waugh, 1999; Nallamsetty and Waugh, 2006; Hayashi and Kojima, 2008; Sun, C. et al., 2011). Expression of the cytoplasmic domain with the solubility tags MBP and GST was carried out in BL21 DE3 cells (Figure 3.8). However, the solubility tags don't promote soluble expression of the cytoplasmic domain at any temperatures used, therefore the cytoplasmic domain was subsequently expressed without the inclusion of a solubility tag in the construct.

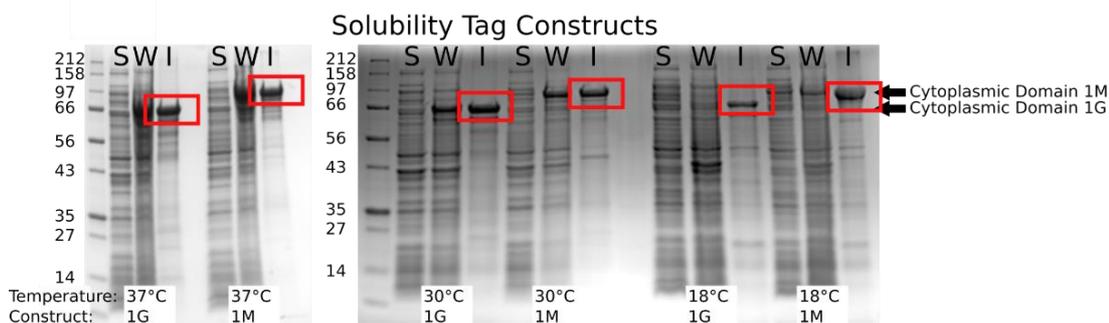


Figure 3.8

SDS-PAGE gel of solubility tests carried out for different E. coli expression conditions of the IRE1 α (548-977) construct. S = soluble fraction, W = Whole protein and I = Insoluble fraction. Expression of different constructs of the cytoplasmic domain with solubility tags, where 1M includes an N-terminal MBP tag and 1G includes a N-terminal GST tag. Expression carried out with the BL21 DE3 expression strain at the indicated temperatures for four hours with 1mM IPTG, red boxes show insoluble cytoplasmic domain expressed (Section 2.5.1)

3.2.3. Selection of an optimal condition

Presented here are multiple alterations to the recombinant expression of IRE1 α 's cytoplasmic domain in *E. coli*. No condition trialled appears to give soluble expression of the domain, this is likely due to the complexity of the protein's kinase and endoribonuclease domain. Although there are many more combinations of conditions and different methods that can be attempted to achieve soluble expression such as co-expression with chaperones and codon optimisation, this can be overly time consuming and may be impossible to achieve.

As differences in temperature and *E. coli* strain do not affect expression yield greatly, a high yield of insoluble cytoplasmic domain can be gained with relatively uncomplicated expression conditions. Therefore, the expression condition used to produce the cytoplasmic domain was using a construct without a solubility tag in the BL21 DE3 *E. coli* expression strain, induced to express with 1mM IPTG for four hours at 37°C (Figure 3.4, Section 2.4.5.1). Rather than attempting to express soluble cytoplasmic domain, the aim became to isolate and purify the cytoplasmic domain from inclusion bodies and optimise conditions for its *in vitro* refolding to the native state.

3.3. Purification of the cytoplasmic domain protein from the insoluble fraction

3.3.1. Preparation and purification of the insoluble fraction

The optimal expression condition was used to express the cytoplasmic domain on a larger scale. The cells with the expressed insoluble protein were pelleted and the insoluble fraction was solubilised in the chaotropic agent guanidinium (Section 2.4.6.1) which allowed the protein to be purified using nickel ion affinity chromatography, utilising the construct's N-terminal 6His-tag (Hochuli et al., 1988). Purification of the cytoplasmic domain gives a repeatable elution of protein with low imidazole concentration (5mM) and elution of a much smaller amount of protein with 500mM imidazole (Figure 3.9).

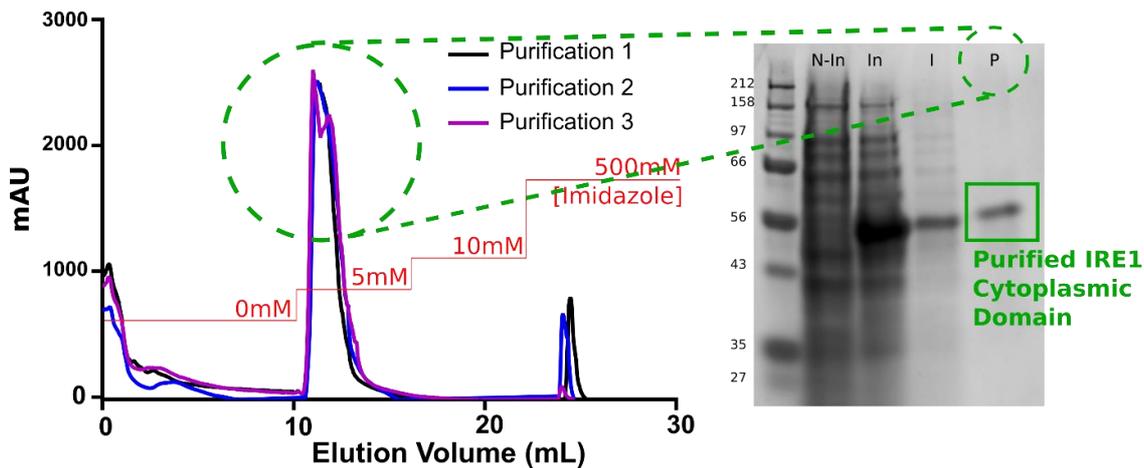


Figure 3.9

Chromatogram for multiple IMAC (Nickel resin) purifications of the cytoplasmic domain in the presence of 6M guanidinium. The concentration of imidazole used is shown in red. The protein consistently elutes with 5mM imidazole, suggesting a weak interaction with the column. Also shown is an SDS-PAGE gel for the cytoplasmic domain. 'N-In' is for non-induced culture, 'In' is for cultures induced to express the cytoplasmic domain construct with IPTG, 'I' is the insoluble fraction of this culture after preparation and 'P' is the product after IMAC purification of the culture from the 5mM imidazole peak, as labelled in green. The purified cytoplasmic domain gel band appears to have shifted, however this is an effect of the gel warping. Purification method in Section 2.5.2.1.

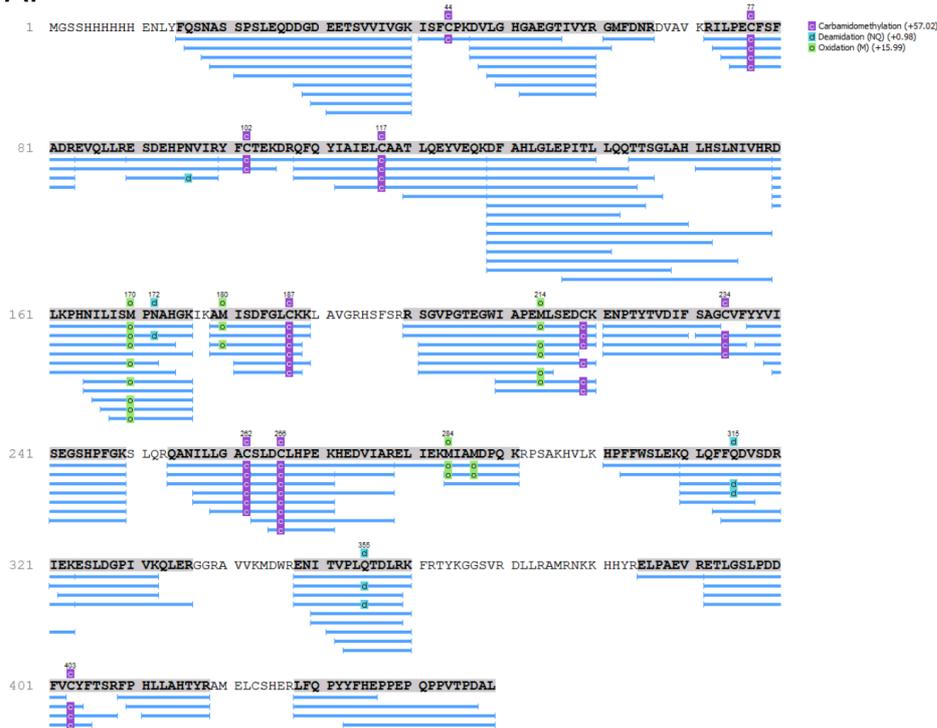
5mM of imidazole is a low concentration to elute a His-tagged protein, for this reason an SDS-PAGE gel was run to determine the purity of this elution peak (Figure 3.9), it depicts the proteins expressed before and after IPTG induction, after isolation of the insoluble fraction and the final product after IMAC purification. After induction a protein band appears at close to 50kDa, where the recombinant cytoplasmic domain is expected. It is apparent that the cytoplasmic domain is the dominant species present in the insoluble fraction and the band for the cytoplasmic domain becomes more dominant after the IMAC purification. This in agreement with previous literature suggesting that purified insoluble proteins from inclusion bodies give high yields and improved purities (Fahnert et al., 2004; Greenshields et al., 2008). It is therefore clear that despite the low

concentration of imidazole required for elution of the cytoplasmic domain from the HisTrap column, the protein is of a high purity and appears on the gel at near the size expected although further verification of the protein species present is required.

3.3.2. Mass spectrometry to validate the size and sequence of the cytoplasmic domain

Mass spectrometry was used to provide the mass and sequence of the purified species to verify that it was the cytoplasmic domain construct. The gel band containing the cytoplasmic domain was prepared for trypsin digest analysis of the fragments by mass spectrometry to determine the sequence of the isolated protein (Figure 3.10A). This suggests the protein has at least 80% identity to the cytoplasmic domain construct. The protein was also sent for molecular mass determination through mass spectrometry (Figure 3.10B), with the observed mass found to be 51292.83 ± 1.31 Da. This is near the expected mass of the protein with a N-terminal methionine cleavage (51294.41Da). N-terminal methionine cleavage is common in *E. coli* and as the methionine cleaved precedes the 6His-tag region, there is likely to be no influence on stability or activity of the protein (Hirel et al., 1989). Mass spectrometry analysis therefore validates that the desired hIRE1 α cytoplasmic domain construct has been successfully cloned into expression vectors, its *E. coli* expression optimised and the protein isolated and purified from inclusion bodies.

A.



B.

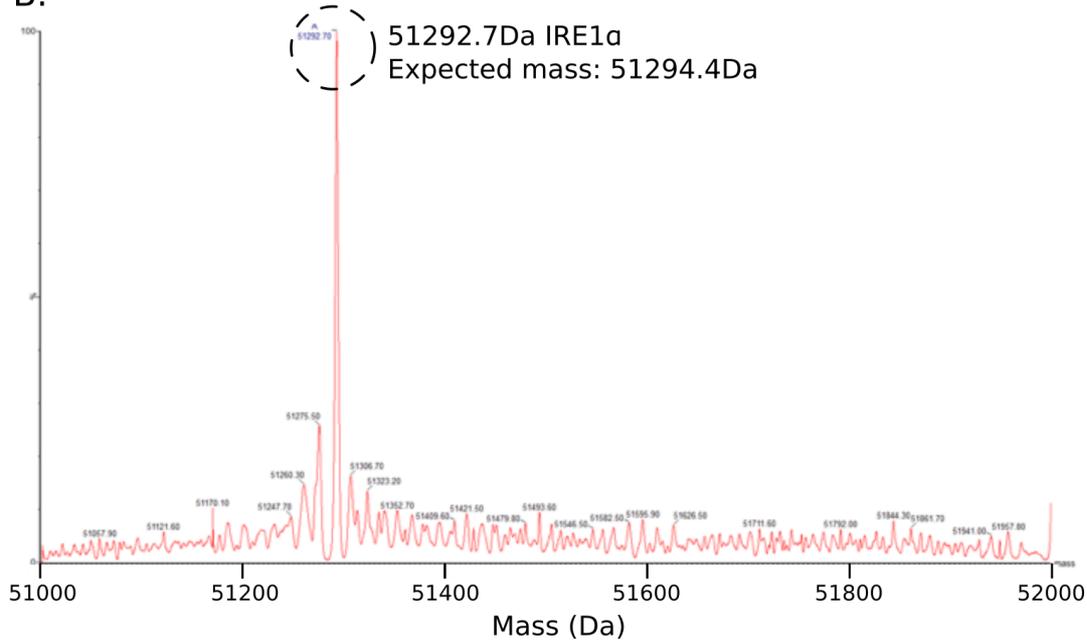


Figure 3.10

Confirmation of the cytoplasmic domain construct expressed in *E. coli*. A: Mass spectrometry sequence analysis after trypsin digest (Section 2.6.1.1). 80% coverage of the cytoplasmic domain construct was observed, suggesting the desired construct was purified. B: Mass spectrometry molecular mass analysis of the cytoplasmic domain (Section 2.6.1.2) suggests a mass of 51292.7Da, the expected mass with a methionine cleavage is 51294.4Da.

3.4. Refolding reaction optimisation

After purification of the cytoplasmic domain, conditions for refolding the protein to its native state could be optimised. There are a large range of additives, conditions and techniques that aid the refolding of proteins already known (Burgess, 2009; Yamaguchi and Miyazaki, 2014), these are utilised in the optimisation of cytoplasmic domain refolding.

3.4.1. Refolding techniques in arginine buffer

The refolding buffer used is critical to the refolding process, its role is to prevent interactions of folding intermediates from forming aggregates and to promote correct folding to the native state. Although there are examples in the literature of some conditions and additives being beneficial to the refolding reaction, every protein will react to these differently. However, there are multiple examples of arginine aiding refolding reactions, therefore its effect on the refolding of the cytoplasmic domain was tested (Baynes et al., 2005; Arakawa et al., 2007). For the refolding reactions presented here refolding reactions unfolded cytoplasmic domain in guanidinium buffer was added drop wise to the initial refolding buffer (50mM Hepes pH 7.5, 150mM NaCl, 5mM TCEP, 10% glycerol; Section 2.8.1.1) at 4°C and stirred for one hour. Refolding reactions were carried out with different concentrations of arginine present and increasing concentrations of cytoplasmic domain. After refolding, the samples were centrifuged to remove aggregates from solution and the protein concentration measured using a Bradford assay (Figure 3.11). When refolded in the absence of arginine the protein concentration was consistently below the range of measurement, with aggregation observed in the reaction. It is clear that increasing arginine concentration in the refolding reaction increases the soluble yield of protein, in agreement with the literature (Dong, X.Y. et al., 2004; Baynes et al., 2005; Arakawa et al., 2007).

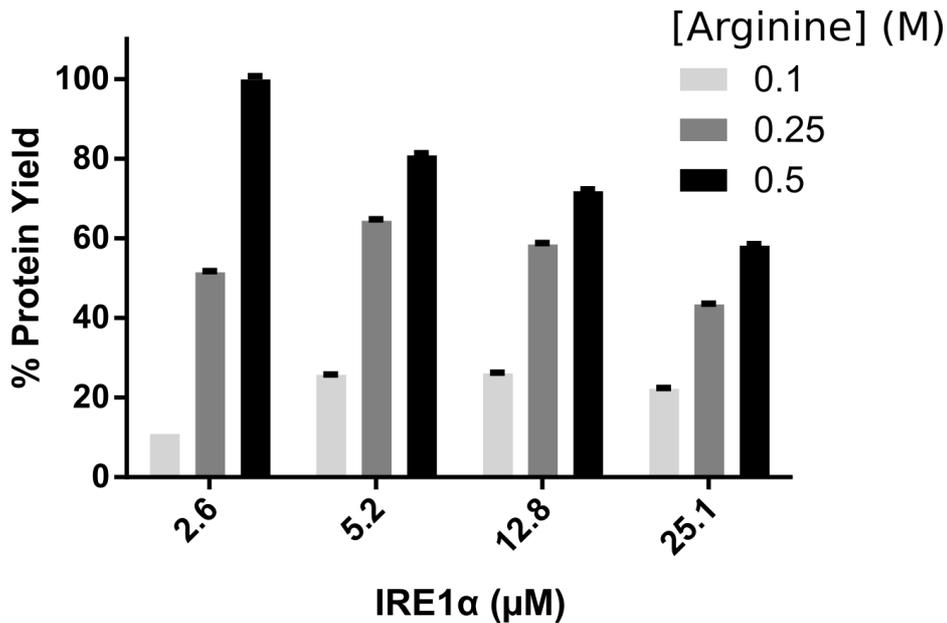


Figure 3.11

The cytoplasmic domain was refolded at different concentrations of protein and arginine. Shown is a graphical representation of protein yield from dilution refolding reactions (Section 2.8.1.1) of the cytoplasmic domain. Increasing arginine concentrations increase the yield at different refolding concentrations, with higher yields achieved at lower protein folding concentrations. Concentration measured by use of Bradford assays (Section 2.5.4).

3.4.2. Buffer optimisation for stability

Inclusion of arginine appears to provide beneficial effects for the yield of soluble protein from the refolding reaction of the cytoplasmic domain, to investigate conditions to improve the stability of the cytoplasmic domain after refolding an Optim 1000 was used. The Optim analyses thermal unfolding and aggregation of the protein across a range of temperatures using different buffer conditions through fluorescence and light scattering. Here the effect of salt, pH and glycerol were measured without arginine, as shown in Figure 3.12. The barycentric mean (BCM) reports on the unfolding of the protein and light scattering at 266 and 473nm report on the aggregation of the protein. The effect of pH appears to be modest, although the lowest pH values trialled cause aggregation, there appears to be an advantage at pH8.0 and pH8.5 in preventing aggregation (Figure 3.12A-C). The concentration of NaCl also appears to have a modest effect, with high salt concentrations preventing unfolding, but lower salt concentrations preventing aggregation. Near physiological salt does not promote aggregation in the same way that >500mM NaCl does (Figure 3.12E-F), but still prevents unfolding more than lower salt concentrations (Figure 3.12D). In all cases it appears that increasing the concentration of glycerol improves the stability of the protein (Figure 3.12G-I), however, high glycerol concentrations are undesirable for use with gel filtration

columns (for future buffer exchange) due to the increased viscosity. The refolding buffer was thus altered for all subsequent experiments using the data from the Optim, the pH was increased from 7.5 to 8.0 and the salt concentration was increased from 150mM to 300mM.

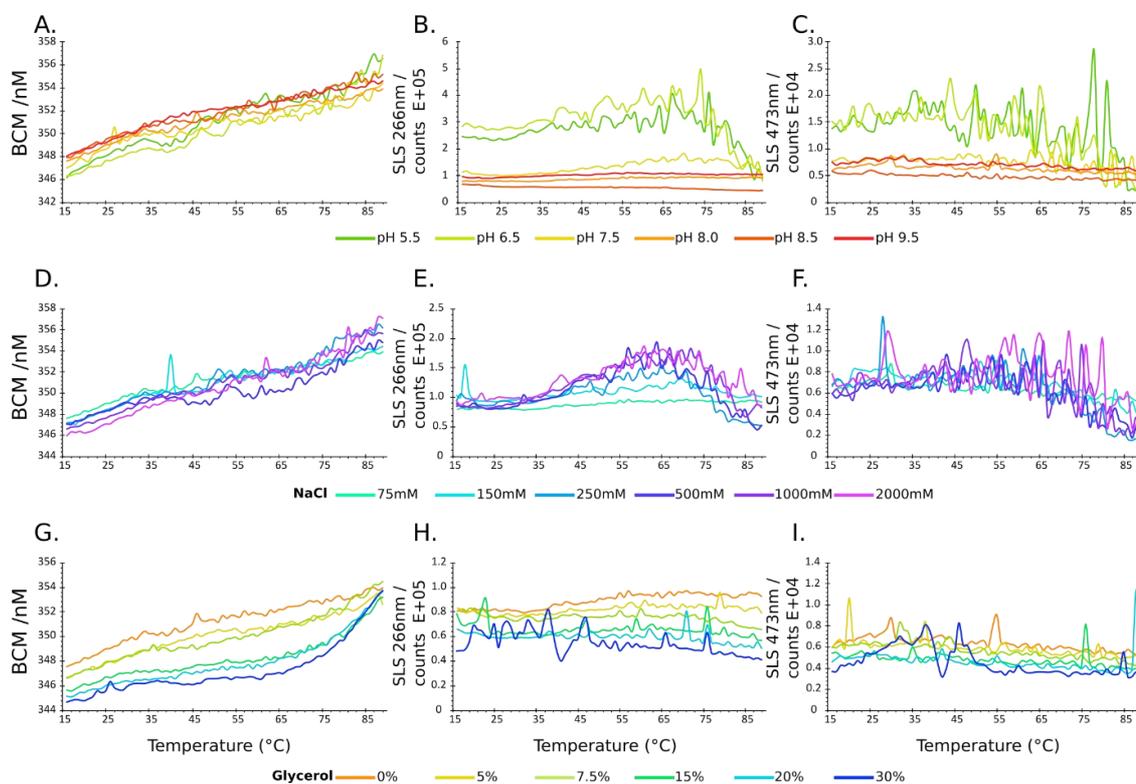


Figure 3.12

Data obtained from analysis of different buffer conditions using an Optim 1000 (Section 2.8.2). BCM (first column) reports on the unfolding of the protein and the light scattering (second two columns) on the aggregation of the protein at increasing temperatures. A-C: Changes in pH's influence on the sample. D-F: The effect of changes to salt concentration on the sample. G-I: The effect of glycerol on the sample.

3.4.3. Optimisation of refolding technique and subsequent removal of arginine

Although arginine's presence improves the soluble yield of the refolding reaction, it is not desirable to have arginine in solution with the protein when assessing activity as it suppresses protein interactions (Arakawa et al., 2007) and its presence interfered with collection of circular dichroism spectra. Arginine also proved difficult to remove from the protein solution after refolding through dialysis. Therefore, the refolding reaction was attempted in the endoribonuclease assay buffer (20mM HEPES pH8.0, 50mM KOAc, 0.5mM MgCl₂, 10% glycerol, 3mM TCEP, Section 2.8.6) to remove the requirement of arginine, and additional steps of buffer exchange after refolding. Different refolding techniques were used and compared to the arginine refolding buffer at a protein concentration of 2.5μM. Figure 3.13 clearly shows that

refolding in each technique gives significantly lower yields in the absence of arginine. Also, that dialysis and dilution refolding give the highest protein yields. Although dialysis refolding gives a higher yield than dilution refolding, it significantly prolongs the protocol and the cytoplasmic domain may lose activity in this time, therefore dilution refolding was preferred.

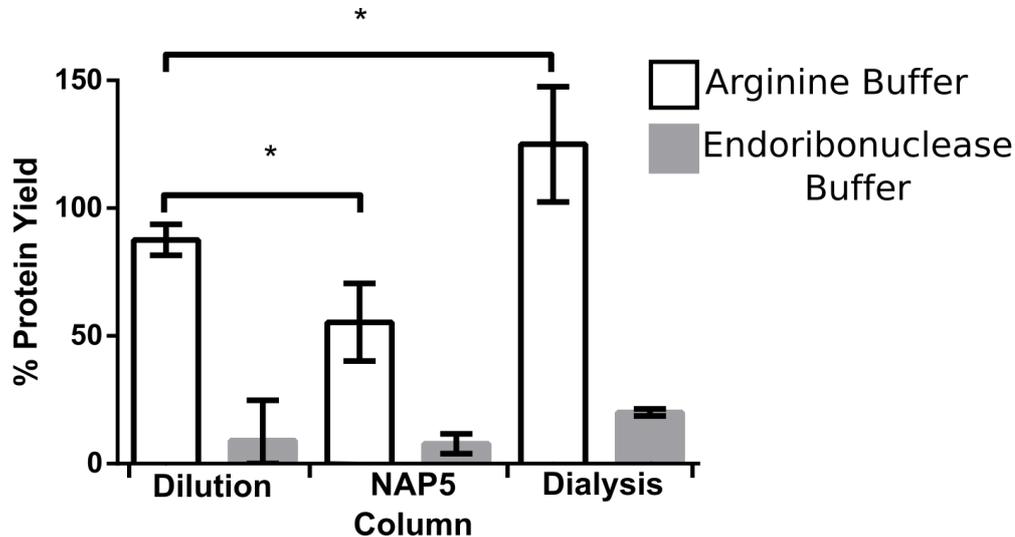


Figure 3.13

The cytoplasmic domain isolated from inclusion bodies after *E. coli* expression was refolded using different techniques and buffer conditions at a concentration of $2.5\mu\text{M}$. After the refolding process (Section 2.8.1) the concentration of soluble protein was measured, and the yield calculated using Bradford assays (Section 2.5.4, $n=3$). The refolding buffer containing arginine gave significantly higher yields in all methods (Dilution refolding $p=0.0013^{**}$, NAP5 refolding $p=0.0062^{**}$ and dialysis refolding $p=0.0013^{**}$). For the arginine buffer dialysis refolding has a significantly higher yield ($p=0.049^*$) than dilution refolding. And dilution refolding gave a significantly higher yield than NAP5 column refolding ($p=0.0266^*$). ($*<0.05$, $^{**}<0.01$, $^{***}<0.001$).

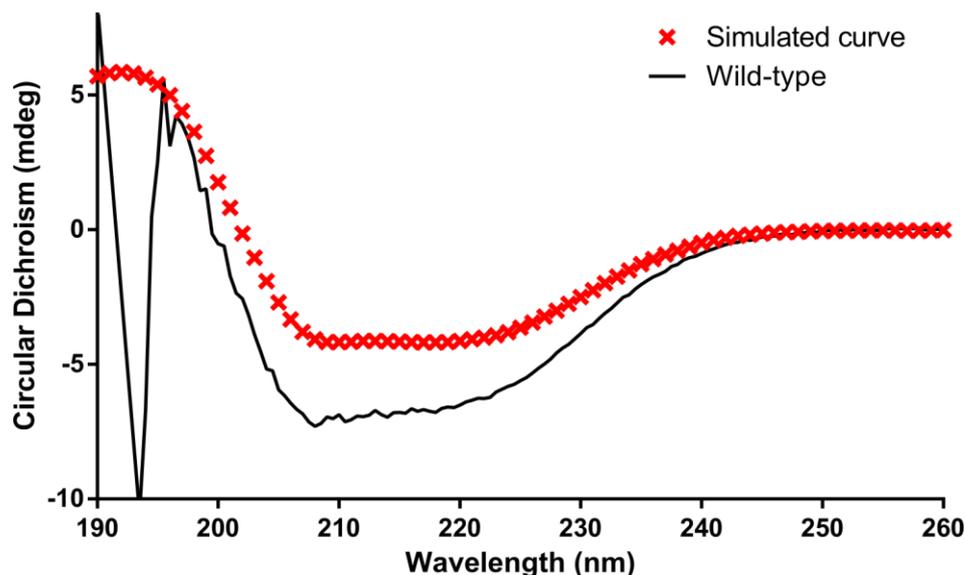


Figure 3.14

Circular dichroism data for the refolded cytoplasmic domain. Experimental data suggests that the refolded cytoplasmic domain contains regions of secondary structure (measured at $4\mu\text{M}$, Section 2.6.3, $n=2$). This is compared to a simulated curve for a solved cytoplasmic domain structure (PDB: 4U6R) using PDB2CD (Harrington et al., 2015; Mavridis and Janes, 2017).

Because arginine is required for high soluble protein yields but is difficult to remove from solution, a NAP5 gel filtration column was next trialled to buffer exchange the refolding reaction to remove arginine (Section 2.5.6.2). The advantage of this method being that it is less time consuming than dialysis and appears to completely remove the remaining arginine; the disadvantage being that the protein concentration is halved upon elution. After exchange to a suitable buffer using a NAP5 column and analysis by circular dichroism, the arginine interference with the spectrum previously described was no longer present. The circular dichroism spectrum is shown in Figure 3.14, although there is some error, the spectrum's profile is indicative of a folded protein, suggesting that the refolding reaction allows the *E. coli* produced cytoplasmic domain to fold and adopt secondary structural elements.

3.4.4. Validation of refolding reaction

To further validate the circular dichroism result, the protein was injected onto a size exclusion column in refolding buffer (Figure 3.15). Size exclusion chromatography analysis reports on both the size and shape of a protein, through use of a calibration curve the cytoplasmic domain eluted at where a 54.2kDa protein would be expected to, which is a similar size to that of the construct calculated through mass spectrometry analysis (52292Da) and suggests a monomeric state. Although 12% of the protein in the chromatogram resides in this monomeric peak, there is an

abundance of protein eluting in the void volume of the column, suggesting the formation of large species. IRE1 α 's cytoplasmic domain is known to form large oligomers in solution at higher concentrations (Korennykh, A. and Walter, 2012), so it's possible that this is the case, although these species may also be aggregates of the protein. The amount of these large species present appeared to be sample-dependent, further investigation is therefore required into their cause to optimise the protocol.

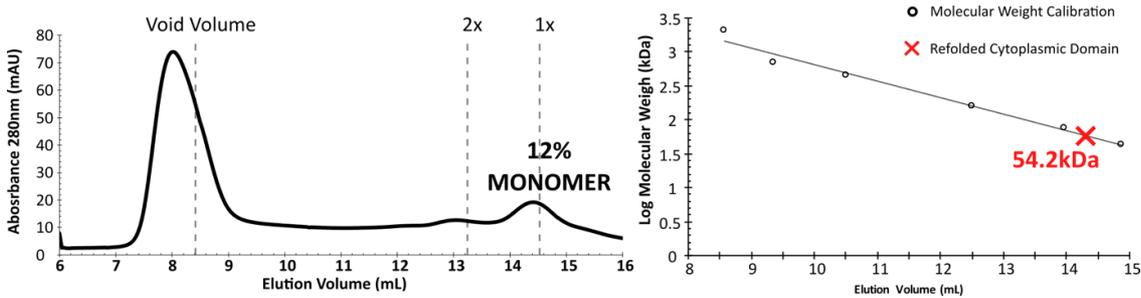


Figure 3.15

Size exclusion chromatography of the refolded cytoplasmic domain. Left: The refolded cytoplasmic domain at injected at 30 μ M (Section 2.6.2). 12% of the protein appears where monomeric protein is expected and there is a significant portion of protein eluting in the void volume. Right: The calibration curve constructed from standard weight markers. The refolded cytoplasmic domain monomeric peak's elution volume suggests a protein mass of 54.2kDa (red cross) where the protein's actual mass is 52.3kDa. The elution of the large species is above 2000kDa and therefore elutes in the void volume.

To verify the refolding reaction further, the cytoplasmic domain was expressed using a 15 N labelling strategy (Section 2.4.5.3) and refolded using the optimised method. The protein was analysed by NMR using a HN TROSY experiment and compared to a sample that was denatured in 6M urea (Figure 3.16). The spectrum for the denatured cytoplasmic domain shows low dispersion and high intensity of peaks, indicative of an unfolded protein (Okazaki et al., 2018; Gupta and Bhattacharjya, 2014). The spectrum for the refolded cytoplasmic domain on the other hand has many peaks disappear, this is expected for a large, 50kDa folded protein (Konrat, 2014).

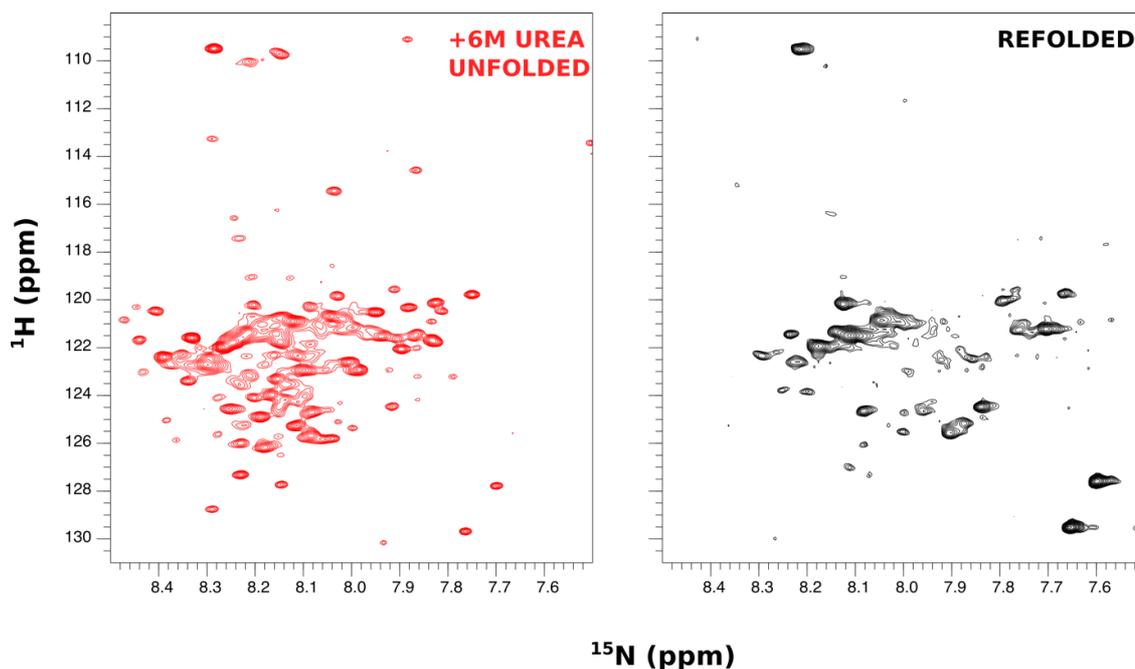


Figure 3.16

NMR HN TROSY spectra for ^{15}N labelled cytoplasmic domain. In red is the cytoplasmic domain in a 6M urea solution, therefore unfolded. In black is the refolded cytoplasmic domain at $60\mu\text{M}$ (Section 2.7.1.1). Many peaks disappearing in the refolded sample suggest it's refolded.

Therefore, it appears as though the refolding reaction of IRE1 α 's cytoplasmic domain after production in *E. coli* gives folded protein as confirmed by circular dichroism, size exclusion chromatography and NMR. The size exclusion chromatogram suggests that at the concentration used in refolding buffer, ~12% of the protein is monomeric and a large portion forms larger species that elute in the void volume.

3.5. Measuring the activity of the refolded cytoplasmic domain

With the refolding conditions optimised, the activities of the cytoplasmic domain were validated. These activities include the kinase domain's ability to convert ATP to ADP, its ability to phosphorylate adjacent monomers and the protein's endoribonuclease activity.

3.5.1. Species present in activity assay buffers

The emergence of large oligomeric species from the size exclusion chromatogram (Figure 3.15) may indicate that the protein is forming oligomeric species, or that the protein has low stability and therefore aggregates. To this end we wanted to first characterise the emergence of these species when the protein was buffer exchanged into the activity assay buffers. To do this the

protein was exchanged into each buffer using NAP5 columns (Section 2.5.6.2) and the hydrodynamic radius of species present was determined by dynamic light scattering (Figure 3.17). Although the proportions of the oligomeric to monomeric protein cannot be determined by this method, we observed the oligomeric species in all buffers used, including the refolding buffer, all at $2\mu\text{M}$. Therefore, these large oligomers will be the major species present in all subsequent activity assays and may influence the apparent activity of the protein.

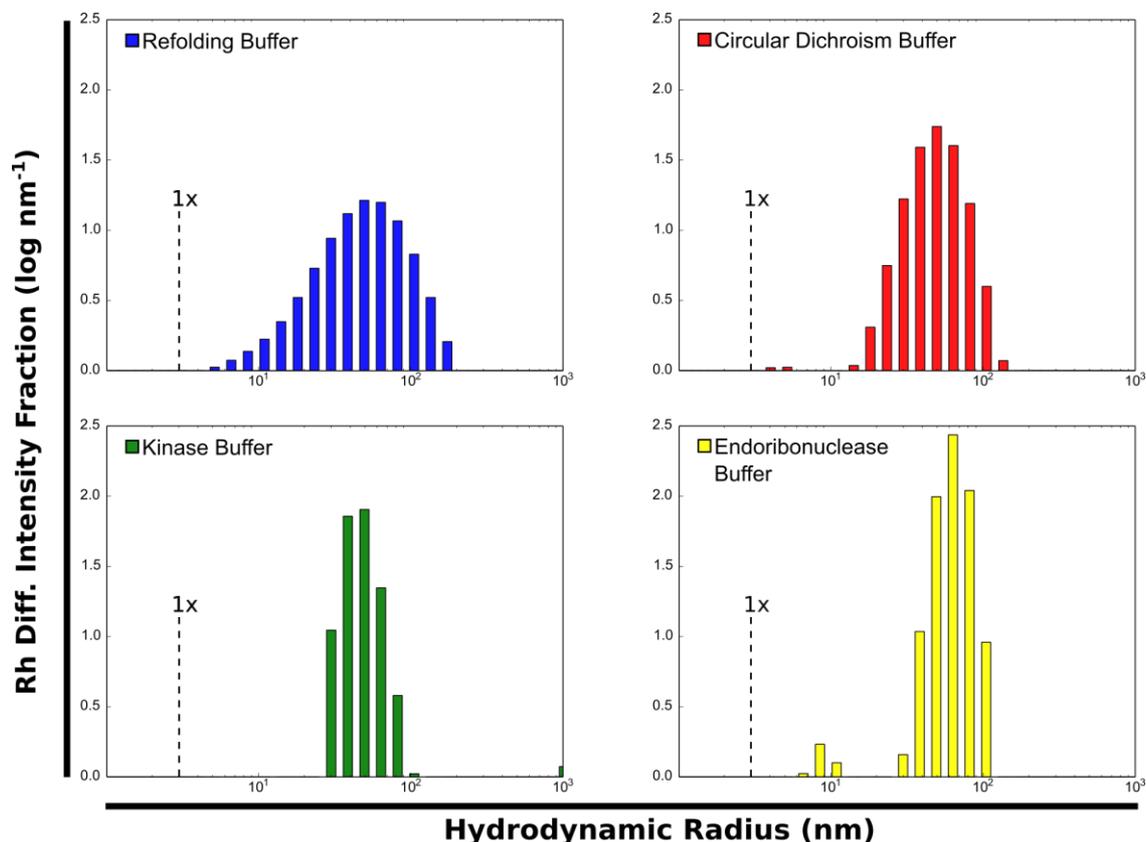


Figure 3.17

Dynamic light scattering analysis of the refolded cytoplasmic domain at $2\mu\text{M}$ when exchanged into different buffers using a NAP5 column (Section 2.8.3). This suggests that large species form in all buffers used. Shown is the predicted hydrodynamic radius of the monomer.

3.5.2. The refolded cytoplasmic domain has kinase activity

The kinase activity of the protein can be compared to that of protein produced in a eukaryotic system using the ADP-Glo kinase assay from Promega (Section 2.8.4), where an increased amount of ADP produced by kinase or ATPase activity causes an increase in luminescence. The kinase assay was carried out with 0, 30 and 60-minute timepoints and the results of the assay are shown in Figure 3.18. When comparing the 60-minute time point for each protein, the cytoplasmic domain produced in *E. coli* has 23.6% the activity of the protein produced by eukaryotic expression systems. This lower activity compared to eukaryotically produced protein may be due

to the protein being inactive and the apparent kinase activity coming from impurities in the sample or because the majority of the protein forms aggregated and inactive species, as observed in size exclusion chromatography and dynamic light scattering experiments (Figures 3.15 and 3.17).

To determine the cause of the lower kinase activity the I642G mutant was created, which lacks kinase activity (Han et al., 2009). The protein was expressed in *E. coli*, validated, purified and refolded using the protocol optimised for the *wild-type* protein. The mass spectrometry analysis and circular dichroism spectrum of refolded I642G are compared to *wild-type* protein in Figure 3.18 and show no significant differences in proportion of secondary structural elements.

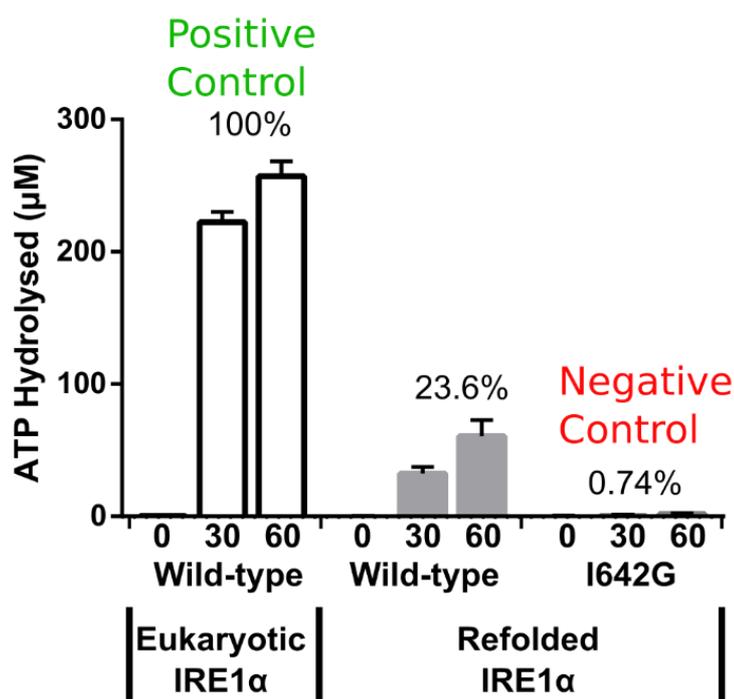


Figure 3.18

The kinase activity of the refolded cytoplasmic domain was compared to the kinase dead I642G mutant and cytoplasmic domain produced eukaryotically using the Promega ADPglo assay (Section 2.8.4, $n=3$). Shown are the activities of the proteins as a percent of eukaryotic cytoplasmic domain activity. The refolded wild-type protein has 23.6% the activity of the eukaryotic protein after 60 minutes, and the I642G mutant has 0.74%.

The role of the I642G mutant is to provide a negative kinase activity control, the mutant has been expressed and refolded in the same protocol as the *wild-type* construct. Therefore, as the mutant has no kinase activity, but *wild-type* protein does, we can conclude that the *wild-type* protein has kinase activity which is not due to impurities in its preparation and the lower kinase activity observed is likely due to the large species present in the sample being inactive.

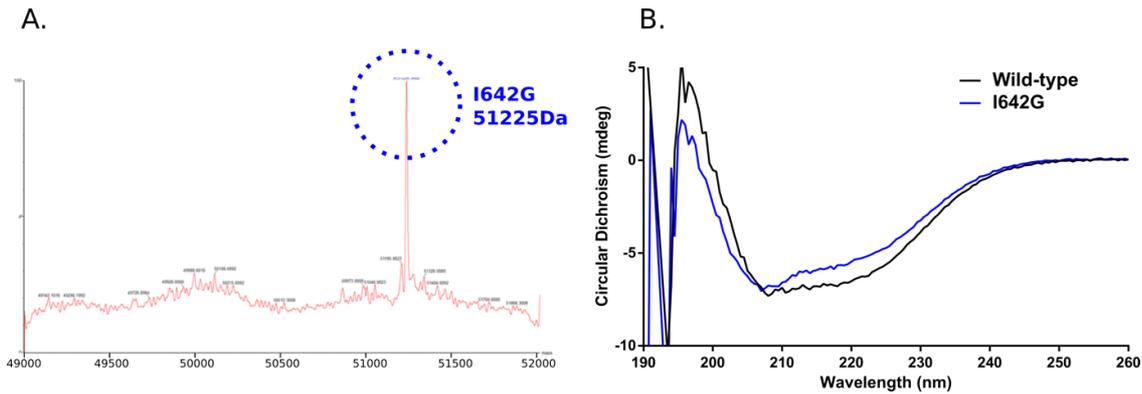


Figure 3.19

Validation for the I642G mutant. A: Mass spectrometry analysis of the mutant suggests a mass of 51225Da, where the expected mass is 51238Da (Section 2.6.1.2). The sequencing result suggests that the protein has the correct sequence, this difference in mass has been observed in samples of different proteins and may be instrument-dependent. However, the protein's lack of kinase activity after purification in the same protocol as wild-type protein suggests the kinase activity observed in Figure 3.17 is not due to impurities. B: Comparison of circular dichroism spectra (Section 2.6.3, $n=2$) for the wild-type and I642G refolded proteins suggests both contain elements of secondary structure.

Kinase domains consist of two lobes that must interact in a precise conformation to achieve activity (Section 1.5.1). Due to the complexity of kinase domains there are seldom examples of successfully refolded kinase activity in the literature (Hibino et al., 1994; Nguyen Sle et al., 2014). Therefore the ~25% kinase activity of the refolded cytoplasmic domain is a positive result. At the 30 μ M injected concentration, 12% of the protein is monomeric (Figure 3.15) and at 500nM for the kinase assays the proportion of monomeric protein is likely to be higher, therefore if the large species are inactive aggregates this would agree with the kinase assay results of ~25% activity. However, further investigation into the formation of larger species is required.

3.5.3. Autophosphorylation of the protein

As evidenced in Figure 3.18 the refolded cytoplasmic domain can hydrolyse ATP, however, its phosphoryl-transfer activity must be tested, as this is important for regulation of endoribonuclease activity (Itzhak et al., 2014; Prischi et al., 2014). To promote phosphorylation, the refolded cytoplasmic domain was incubated with ATP and MgCl₂ (Section 2.8.5) and submitted for molecular mass analysis by mass spectrometry (Figure 3.20). Phosphorylation of three activation loop serine residues is expected to occur, which would add 80Da each. The reaction was carried out with 50 μ M and 25 μ M protein, however, no phosphorylation was observed, as presented in

Figure 3.20A-C. It seemed possible that the 6His tag present in the construct may inhibit phosphorylation, so the TEV site was cleaved by TEV protease before incubation (Section 2.8.7). However, removal of the 6His N-terminal region did not enable phosphoryl-transfer either, as shown in Figure 3.20D-E.

To ascertain whether the kinase region is unable to coordinate the phosphoryl-group for transfer or the activation loop is not able to be coordinated correctly, the refolding reaction was carried out with 1% IRE1 α cytoplasmic domain produced in a eukaryotic system in the presence of ATP before additional incubation at room temperature with ATP. However, no phosphorylation of the *E. coli* produced protein was observed by mass spectrometry (Figure 3.20F). This suggests that eukaryotic protein cannot phosphorylate the protein's activation loop and therefore an incorrect conformation of the activation loop may be adopted in *E. coli* refolded protein. The activation loop is suggested to be flexible before phosphorylation from X-ray crystallography studies (Concha et al., 2015; Joshi et al., 2015) and may become trapped in a conformation that inhibits its orientation for correct phosphoryl-transfer.

To validate that the *E. coli* protein wasn't already phosphorylated during expression, it was incubated with phosphatases before being sent for mass spectrometry analysis (Figure 3.20G-H). Again, the same mass of 51292Da was observed, indicating the protein wasn't already phosphorylated. These results suggest that although the refolded kinase domain can hydrolyse ATP, it may not be able to coordinate the phosphoryl-transfer to the activation loop of an associated monomer. This may be due to the protein not being able to form the face-to-face dimer (Section 1.5.3) (Ali et al., 2011) or the activation loop not adopting the correct orientation. It is possible that as these reactions are carried out with relative high concentrations of the cytoplasmic domain, the large and likely inactive oligomers become the dominant species, any active protein left that is able to trans-autophosphorylate would fall within the noise of the mass spectrometry spectrum. Therefore, ideally more conditions would be trialled for the autophosphorylation reaction.

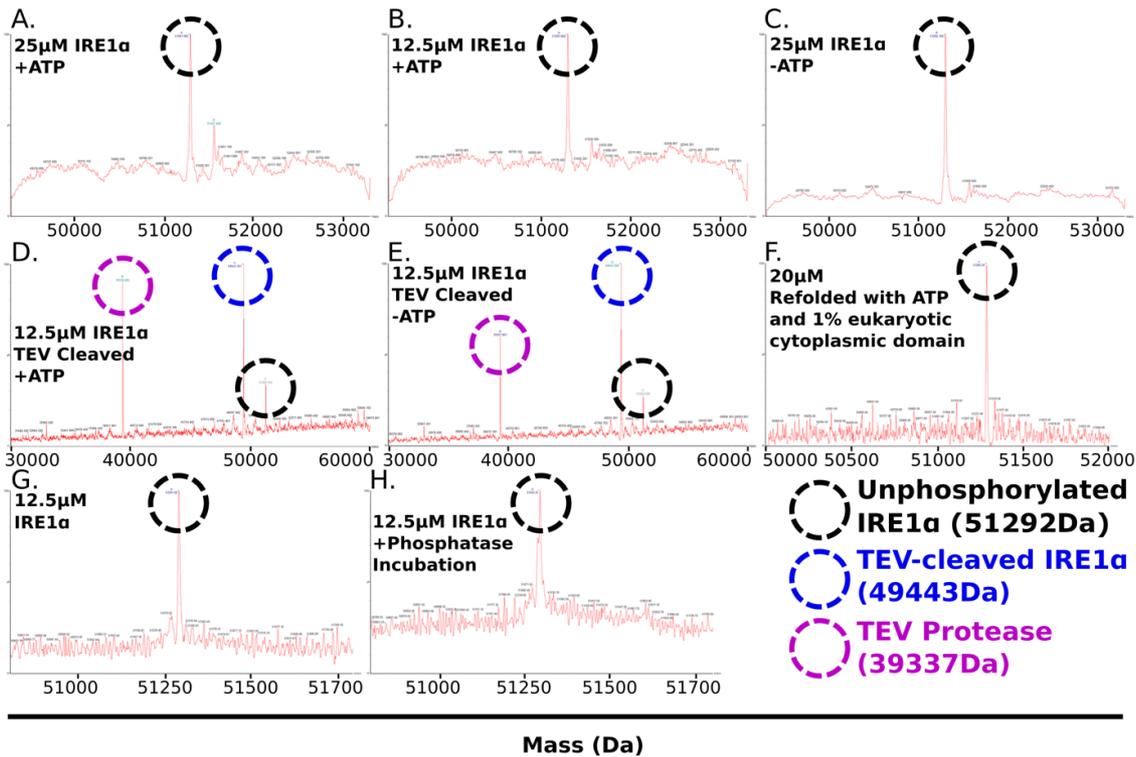


Figure 3.20

Mass spectrometry molecular mass determination results from attempts to promote autophosphorylation of the refolded cytoplasmic domain. A-C: Different concentrations of the refolded cytoplasmic domain incubated at room temperature with 5mM ATP and 5mM MgCl₂ and without before being analysed by mass spectrometry (Sections 2.6.1.2 and 2.8.5, n=2 with only one result shown for each). D-E: TEV-cleavage of the refolded cytoplasmic domain to remove the N-terminal 6His-tag (Section 2.8.7) and incubation with 5mM ATP and 5mM MgCl₂ and without. F: The cytoplasmic domain refolded with 1% of eukaryotically expressed cytoplasmic domain with 5mM ATP and 5mM MgCl₂ and incubation at room temperature. G-H: The refolded cytoplasmic domain with and without incubation with phosphatase inhibitors.

3.5.4. Analysis of the protein's endoribonuclease activity

Endoribonuclease activity is crucial to the cellular role of IRE1 α as it determines the outcome of ER stress. This activity is regulated by the kinase domain's binding of nucleotides and phosphorylation state (Prischi et al., 2014; Wang, L. et al., 2012). The refolded cytoplasmic domain appears to lack phosphoryl-transfer activity but can bind ATP and hydrolyse it. Here, the endoribonuclease activity is assessed, and how it is influenced by the kinase domain and formation of larger species.

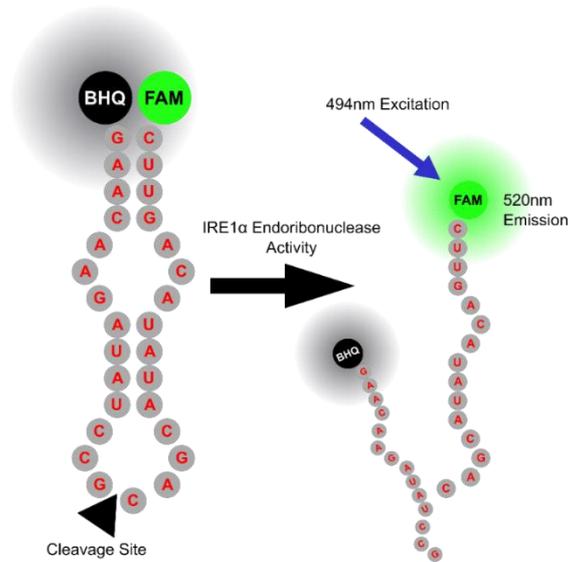


Figure 3.21

Schematic of the XBP-1 mRNA mimic. Before cleavage the black hole quencher (BHQ) group prevents the FAM group from emitting fluorescence. After cleavage the BHQ group is no longer proximal to the FAM group, so it fluoresces after excitation at 494nm. This XBP-1 mRNA mimic is used in the FRET depression assays to measure endoribonuclease activity of IRE1 α cytoplasmic domain (Section 2.8.6).

For *in vitro* detection of endoribonuclease activity, a mimic of the cytoplasmic domain's target mRNA, *XBP-1* (Itzhak et al., 2014) is used. The mimic contains a fluorescent group and a quenching group for use in a fluorescence resonance energy transfer (FRET) depression assay. Before cleavage the fluorescent group is quenched but after cleavage the fluorescent signal increases (Figure 3.21).

3.5.4.1. The refolded cytoplasmic domain cleaves the XBP-1 mRNA mimic

To initially determine whether the refolded cytoplasmic domain has endoribonuclease activity, the mRNA mimic was incubated with 100nM refolded *E. coli* or eukaryotically produced IRE1 α cytoplasmic domain with ATP, ADP or in the absence of both and the fluorescence increase from the mRNA substrate being cleaved was monitored. Figure 3.22 shows that addition of ADP or ATP increase endoribonuclease activity to a similar extent in both the refolded and eukaryotic IRE1 α . In all conditions the refolded protein has similar endoribonuclease activity to the eukaryotic protein (ATP: 77.7%, ADP: 87.4%, APO: 89.3%). Therefore, this suggests that the *E. coli* produced cytoplasmic domain has endoribonuclease activity similar to eukaryotically produced protein and that this activity is modulated by the inclusion of nucleotides. Although the protein appears unable to autophosphorylate to increase its activity, previous study suggest that

binding alone of nucleotides and nucleotide mimics to the kinase domain of the protein can promote a more active endoribonuclease domain conformation (Wang, L. et al., 2012; Han et al., 2009). The low concentration of protein used here (100nM) may inhibit the formation of the large oligomers observed previously for the refolded protein (Figure 3.15 and 3.17) and therefore leave a higher proportion of active protein. Further to this, the *E. coli* protein may be able to autophosphorylate under the conditions used, thus giving similar activity as eukaryotically produced protein in the presence of ATP. However, this may also be an effect of the low concentration of protein used limiting the autophosphorylation-dependent increase in endoribonuclease activity of the eukaryotic protein expected (Prischi et al., 2014), when previously studied at the concentrations used here there was only a small increase in mRNA cleavage observed (Itzhak et al., 2014).

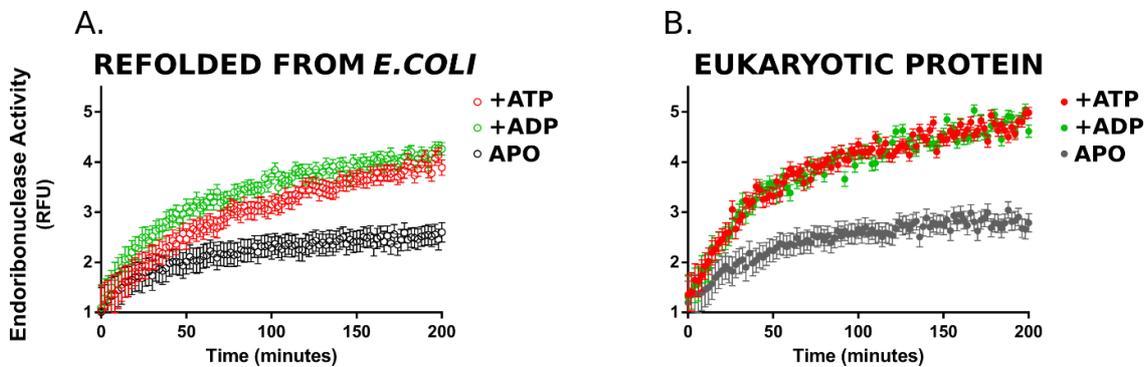


Figure 3.22

The endoribonuclease activity of the refolded cytoplasmic domain at 100nM is measured over time compared to eukaryotically produced cytoplasmic domain in the presence of ATP and ADP and without. 100nM FRET mRNA substrate used (Section 2.8.6.1, $n=3$) and values reported in relative fluorescence units (RFU) to samples containing no protein. A. Refolded *E. coli* expressed cytoplasmic domain. B: Eukaryotic protein. The refolded protein has an increase in endoribonuclease activity upon addition of ADP or ATP and has similar rates of mRNA cleavage as eukaryotically produced protein.

3.5.4.2. The effect of concentration on endoribonuclease activity

The cytoplasmic domain's endoribonuclease activity is drastically increased at higher concentrations due to oligomerisation *in vitro*, with a reported coefficient of 114nM for phosphorylated protein and 369nM for unphosphorylated protein (Itzhak et al., 2014). This suggests that the assays in Figure 3.22 are representative of monomeric or dimeric forms of the protein. Therefore, the refolded protein was next assessed for its ability to oligomerise and increase endoribonuclease activity (Figures 3.23 and 3.24). Increasing concentrations of refolded

cytoplasmic domain were incubated with the same mRNA substrate shown in Figure 3.21. The fluorescence was measured, and the amount of mRNA cleaved was calculated from a standard curve using RNase A (2.8.6.2). Figure 3.23 suggests that the effect of increasing concentration is not linear, indicating that the refolded cytoplasmic domain is capable of dimerisation/oligomerisation in the conditions of this assay.

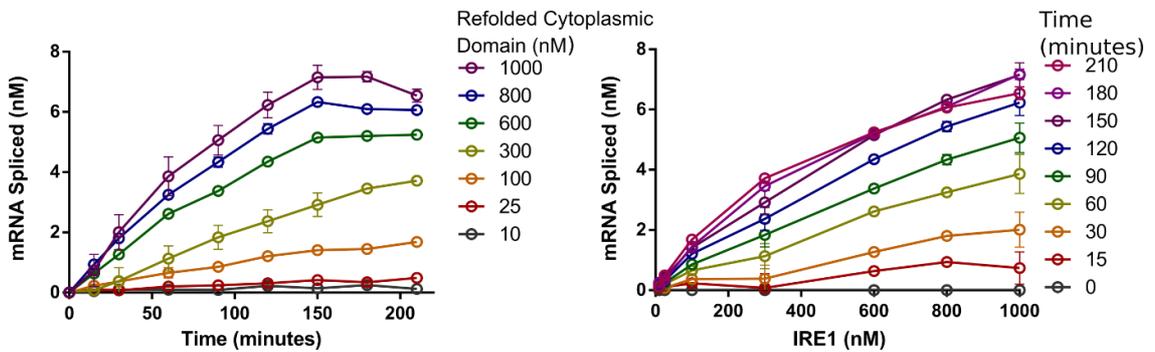


Figure 3.23

The endoribonuclease activity assay was carried out with refolded cytoplasmic domain at a range of concentrations over multiple time points with 25nM FRET mRNA substrate. This data suggests that the relationship between concentration of the refolded cytoplasmic domain and endoribonuclease activity is not linear, suggesting dimerisation or oligomerisation of the protein (Section 2.8.6.2, $n=3$). As a negative control the K907A mutant (Tirasophon et al., 2000) was created and refolded in the same way as the wild-type protein. However, the constructs refolding reaction could not be validated and thus was not used, therefore there is no negative control for endoribonuclease activity.

Upon plotting the data as a function of mRNA spliced against the concentration of IRE1 α present the refolded IRE1 α appears to have an IC₅₀ value of 517nM where the literature value for unphosphorylated protein is 369nM (Figure 3.24) (Itzhak et al., 2014). In the same study the IC₅₀ value of cytoplasmic domain after phosphorylation was 114nM, upon incubation with ATP the splicing activity of the refolded cytoplasmic domain increases but a plateau could not be achieved as the refolded protein is not stable above 1 μ M in these assays, therefore no IC₅₀ value could be calculated (Itzhak et al., 2014). Although these results lack the K907A mutant (mutant created but refolding assay not validated) as a negative control, it is unlikely that an impurity from purification would be able to cleave the *XBP-1* mimic in an ATP-dependent manner, suggesting that the refolded protein can dimerise/oligomerise and responds to ATP for an increased endoribonuclease activity.

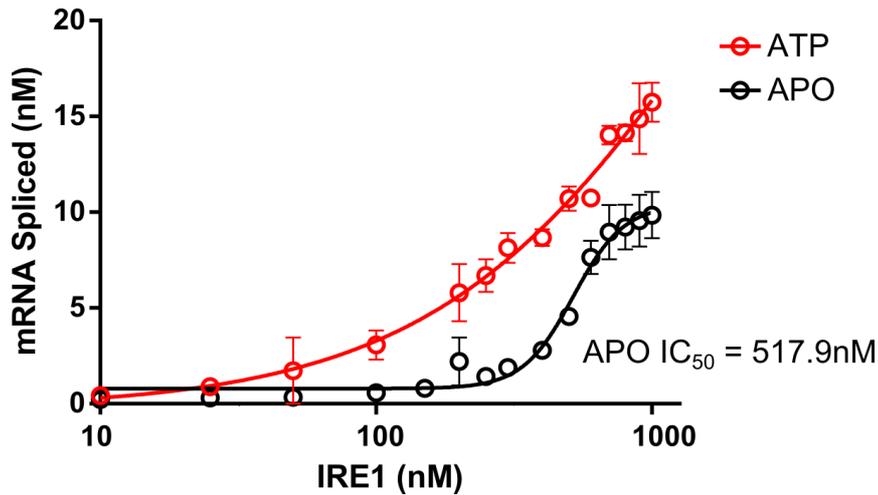


Figure 3.24

Endoribonuclease activity of increasing concentrations of the refolded cytoplasmic domain, with and without ATP and 25nM FRET mRNA substrate. The inclusion of ATP increases the endoribonuclease activity of the protein. The IC_{50} value of the APO state is 517.9nM, this is somewhat similar to that reported in the literature of 369nM (Itzhak et al., 2014). Due to the instability of the protein at higher concentrations a fit wasn't obtained for the assay in the presence of ATP (Details of assay in Section 2.8.6.2, $n=3$).

The protein losing activity at higher concentrations in this assay again suggests that the refolded protein is unstable and liable to form inactive species at higher concentrations, as observed in Figures 3.15 and 3.17 and as suggested by the ~25% kinase activity of the protein. It was therefore hypothesised that the large species are inactive aggregates, and that they would also lack endoribonuclease activity.

3.5.4.3. Characterising the endoribonuclease activity of monomeric and oligomeric species

To characterise the activity of the large species observed, the cytoplasmic domain was incubated with ATP, ADP or neither and analysed by size exclusion chromatography. The monomeric and oligomeric fractions were collected, their concentrations measured, and endoribonuclease assays carried out using them (Figure 3.25). The endoribonuclease experiments used 100nM of protein only, to prevent further loss in activity from extensive sample preparation or potential aggregation from higher concentrations.

It is apparent that the oligomeric fractions do not have endoribonuclease activity (Figure 3.25), whereas, in the presence of ADP and ATP the monomeric fractions splice the mRNA target.

Interestingly the APO monomeric fraction doesn't appear to have endoribonuclease activity, this may be due to the protein losing the small amount of endoribonuclease activity it has in the APO state due to the extent of sample preparation for this assay. The active portion of the protein appears to have lower activity than previously observed in Figure 3.22 and this is also likely due to the lengthy sample preparation. These results suggest that large species are aggregates of the refolded cytoplasmic domain and appear to be the cause of lower than expected kinase activity (Figure 3.18) and the inability to measure endoribonuclease activity at higher protein concentrations (3.24).

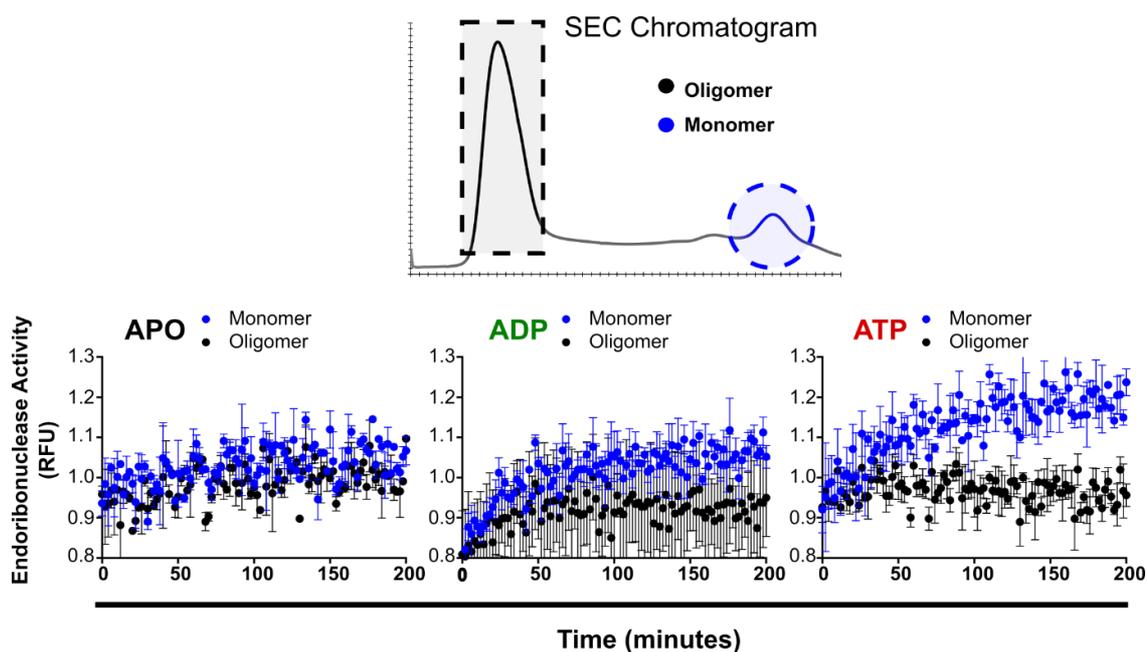


Figure 3.25

The endoribonuclease activity of monomeric and oligomeric species from size exclusion chromatography in the presence of ATP and ADP or neither. The refolded cytoplasmic domain was injected onto a size exclusion column (Section 2.6.2) at a concentration of $30\mu\text{M}$ with no nucleotide added, with 2mM ATP and with 2mM ADP (all in refolding buffer + 5mM MgCl_2). The monomer (blue) and oligomer (black) fractions were collected and analysed by endoribonuclease activity assays with 100nM of protein and 25nM FRET mRNA substrate (Section 2.8.6.1). The void volume fraction appears to be inactive aggregate rather than the protein forming canonical oligomers.

Due to restraints of the endoribonuclease assays caused by protein instability and incomplete characterisation of the K907A mutant as a negative control more work is needed to compare the cytoplasmic domain produced in *E. coli* to eukaryotically expressed protein in all aspects. However, the results presented here suggest that the refolded protein that doesn't form aggregates has endoribonuclease activity similar to that of eukaryotically expressed protein at low

concentrations and can respond to both nucleotide binding to the kinase domain and canonical oligomerisation effects as expected from the literature (Wang, L. et al., 2012; Concha et al., 2015; Prischi et al., 2014).

3.6. Conclusion

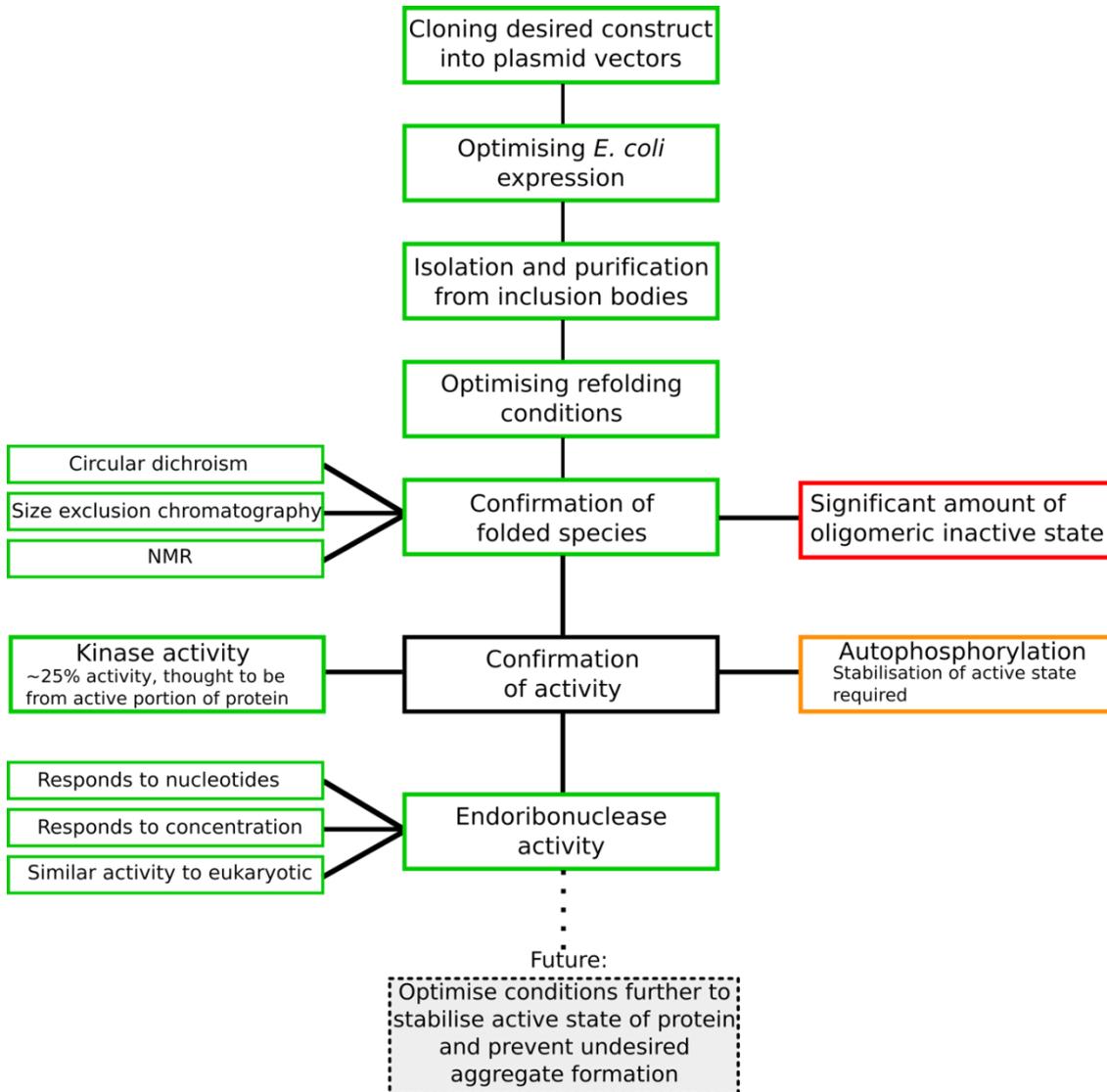


Figure 3.26

A schematic summarising the process, results and future direction of the expression and characterisation of the cytoplasmic domain of IRE1 α in *E. coli*.

The cytoplasmic domain of human IRE1 α has never been stably produced in *E. coli* expression systems before, likely due to the protein's insolubility upon *E. coli* expression and complex kinase extension endoribonuclease conformation making solubilisation or refolding of the protein a daunting task. Here, the IRE1 α cytoplasmic domain construct is prepared and cloned it into three different plasmid vectors. Different growth conditions for soluble expression of the construct in

E. coli were trialled and optimised yield for expression of the protein in inclusion bodies and the subsequent isolation and purification of the construct. The refolding reaction of IRE1 α 's cytoplasmic domain was then optimised to produce folded protein confirmed by multiple techniques.

This protocol required some further optimisation, as the main detractor is the appearance of high order species, the amount of which appear to be dependent on concentration and sample handling. It was first considered that these may be canonical oligomers forming, but when the protein's activity was measured it became apparent that the large species are inactive and therefore likely aggregates suggesting instability of the refolded protein. However, after optimising a series of assays to measure the refolded protein's kinase, phosphoryl-transfer and endoribonuclease assays the protein's native state refolding could be truly validated. Demonstrated here is a protocol to give 12% or above (seemingly dependent on concentration) near native folded protein with kinase activity and exhibiting endoribonuclease activity which increases in the presence of ATP or ADP and upon an increase in concentration. This suggests that the refolded *E. coli* produced protein can form canonical oligomers to increase endoribonuclease activity and also that binding to its kinase domain can cause long range conformational rearrangements to affect endoribonuclease activity, as in eukaryotically produced protein (Lee, K.P. et al., 2008; Itzhak et al., 2014; Wang, L. et al., 2012; Concha et al., 2015). A schematic summary of this chapter is shown in Figure 3.26.

Throughout the study of IRE1 α 's cytoplasmic domain, the difference in conformation of an adaptively acting protein and a protein targeting RIDD substrates has not been understood. By optimising a protocol to produce the cytoplasmic domain construct in an *E. coli* expression system, much larger quantities of protein can be produced which will allow use of high-powered structural techniques that can elucidate the dynamic endoribonuclease domain of the cytoplasmic domain (Concha et al., 2015). One high precision structural technique that particularly benefits from *E. coli* expression of proteins is NMR, with which an initial result for the refolded cytoplasmic domain has been obtained here (Figure 3.16). Increased understanding of the conformations adopted by this domain can aid in the design of drugs to differentiate between the protein's opposing activities. Therefore, with some further optimisation to reduce the emergence of the inactive large oligomers, production of IRE1 α 's cytoplasmic domain in *E. coli* using this protocol will allow for a more precise understand of the protein's activation.

4. Mechanistic understanding of the luminal domain's activation cascade

4.1. Introduction

The endoplasmic reticulum luminal domain of IRE1 α senses endoplasmic reticulum stress and acts as the initiating factor for downstream signal propagation by clustering to increase the local concentration of the cytoplasmic domain (Walter and Ron, 2011). As described previously (Section 1.4.5), there are discrepancies in the field as to the activation mechanism of the luminal domain.

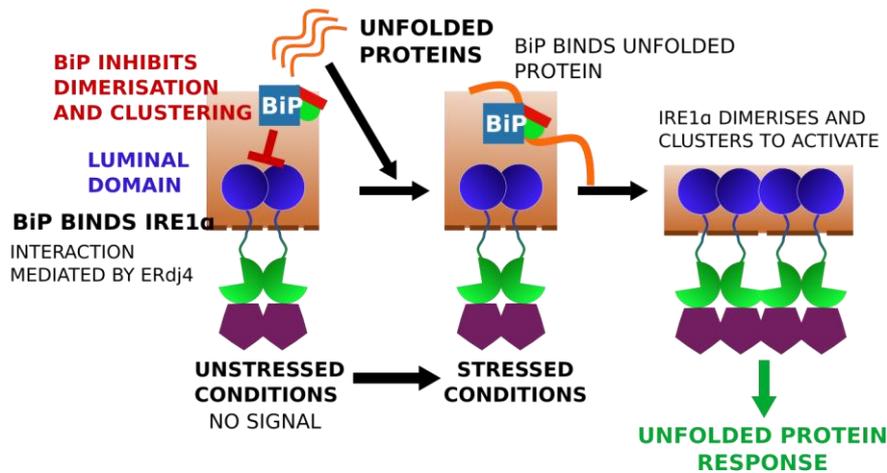
The luminal domain is thought to exist in an equilibrium of monomers and dimers with the potential to form oligomers at higher concentrations *in vitro* (Zhou et al., 2006; Liu et al., 2003; Karagoz et al., 2017). Upon endoplasmic reticulum stress, oligomers are thought to be formed and stabilised by the binding of unfolded proteins to a groove in the luminal domain, promoting formation of IRE1 α clusters (Karagoz et al., 2017). However, an interaction between BiP and IRE1 α 's luminal domain dimers, promoted by ERdj4, represses activation (Amin-Wetzel et al., 2017). There is also evidence of BiP's interaction with the luminal domain via a non-canonical interaction of its nucleotide binding domain, repressing activation (Kopp et al., 2018; Carrara et al., 2015).

It is not yet understood whether the release of BiP repression or direct unfolded protein binding is the principle cause of activation. In addition to this, the activation by unfolded proteins has not been fully characterised and no mechanism to abrogate this signal has been identified. Existing models for activation of the luminal domain are described in Section 1.4 and are summarised in Figure 4.1.

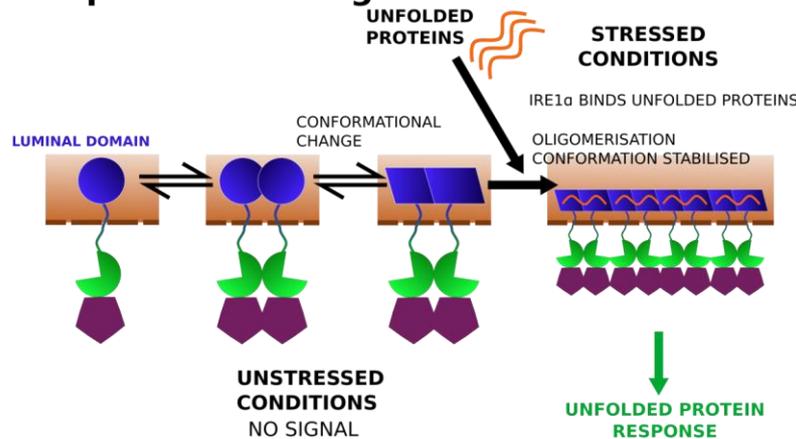
As with the cytoplasmic domain in Chapter 3, the luminal domain can be studied in the absence of the cytoplasmic domain *in vitro* (Liu et al., 2002). Previous *in vitro* study of the luminal domain in the Zhuravleva research group by Sam Dawes and Łukasz Wieteska has provided three conclusions: the human IRE1 α luminal domain has a sub- μ M dimerisation constant, in the presence of high concentrations of the unfolded protein mimic peptide, Δ EspP (Gardner and Walter, 2011), the luminal domain forms large insoluble species and that the inclusion of BiP and ATP de-oligomerise the large insoluble species formed to smaller oligomers (Figure 4.1C).

These findings suggest that the luminal domain's precipitate may be a physiologically relevant active conformation of IRE1 α , that BiP can target and dispel. However, understanding of the precipitated luminal domain's conformation and BiP's interaction are lacking, therefore, quantitative approaches are required. In this chapter the preliminary results from Sam Dawes and Łukasz Wieteska are used as a basis to firstly investigate the luminal domain's conformational landscape in the absence of stress, then quantitative approaches are used to study the effect of Δ EspP on this before characterising BiP's role in resolubilising IRE1 α and peptide complexes.

A. BiP binding model



B. Unfolded protein binding model



C. Previously identified mechanisms

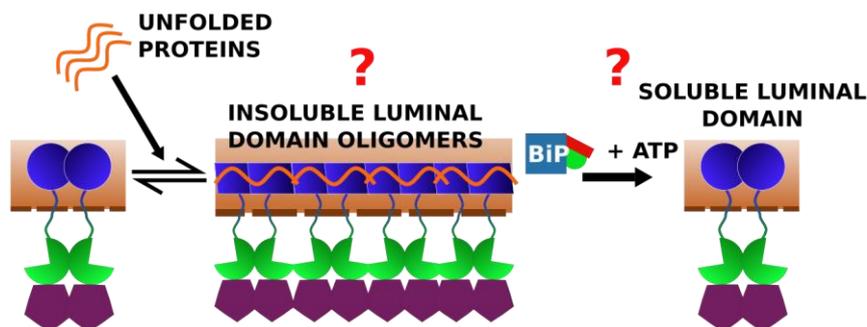


Figure 4.1

Different models for the activation and termination of IRE1 α 's signalling, focusing on the endoplasmic reticulum luminal domain's activation. A. The BiP binding model, where the molecular chaperone BiP binds to and represses activation of IRE1 α , mediated by ERdj4 (Amin-Wetzel et al., 2017). Upon stress, an increase in unfolded proteins present in the endoplasmic reticulum cause BiP to dissociate from IRE1 α to bind to these. IRE1 α 's luminal domain is then able to dimerise and oligomerise to elicit a downstream stress response. B. The unfolded protein binding model. Here BiP release does not activate the luminal domain, instead the luminal

domain directly binds unfolded proteins through its MHC-like binding groove, once bound the protein's active state is stabilised allowing it to oligomerise and elicit a response (Karagoz et al., 2017). C. Previous research carried out by Sam Dawes and Łukasz Wieteska shows that the luminal domain forms insoluble species after incubation with high concentrations of $\Delta EspP$. The inclusion of BiP and ATP with these insoluble species is able to re-solubilise them. However, detail about each of these mechanisms is lacking.

4.2. IRE1 α luminal domain conformations in the absence of binding partners

Initially, the luminal domain's conformational landscape in the absence of interacting partners was explored. To investigate whether the full-length luminal domain formed oligomers in the absence of peptide, increasing concentrations of the luminal domain were injected onto a size exclusion column (Figure 4.2), using the size of the peak eluted, the actual concentration of the protein on the column was calculated (Section 2.6.2.2). As discussed in Section 1.4.1, the luminal domain is highly disordered and therefore the elution volume of the protein is not in agreement with the size of globular protein calibrants used. For this reason a dimerisation perturbed mutant (D123P) (Zhou et al., 2006) was used to dictate the elution volume of monomeric protein and therefore, where dimeric protein will elute was estimated (Figure 4.2A).

In addition to the expected monomeric and dimeric protein, larger species are identified at higher concentrations ($>3\mu\text{M}$), suggesting that oligomerisation occurs without the requirement of unfolded proteins binding. To validate that the protein's equilibrium is not significantly different to that observed previously, microscale thermophoresis study of the protein was carried out (Figure 4.2B). The microscale thermophoresis data suggests that the protein undergoes a conformational change with a sub- μM constant ($0.36\pm 1.2\mu\text{M}$), this in good agreement with previous study of the protein's dimerisation constant ($K_{1/2}=0.2\pm 0.03\mu\text{M}$, work carried out by Sam Dawes). Similar to the previous microscale thermophoresis study, at higher concentrations of luminal domain ($>6\mu\text{M}$) inflections in the data are observed, suggesting additional changes in conformation are occurring. The size exclusion data presented here now suggests that this may be due to oligomerisation of the domain. Size exclusion does not provide the same precision as microscale thermophoresis and therefore further study is required to gain a more precise oligomerisation constant for the luminal domain in the absence of peptide interaction.

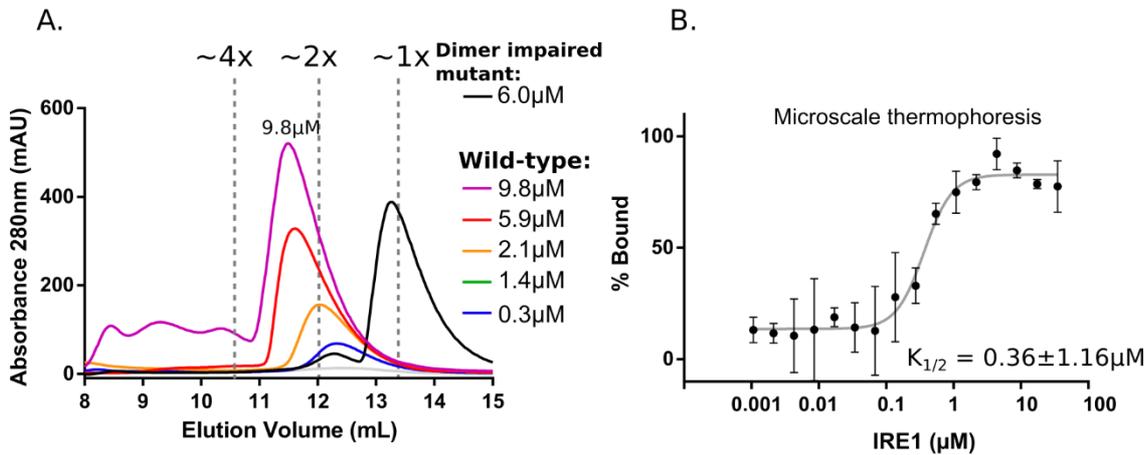


Figure 4.2

With no binding partners IRE1 α 's luminal domain exists in an equilibrium between monomeric and dimeric protein with oligomers forming at increased concentrations. A. Size exclusion chromatogram of wild-type luminal domain at a range of concentrations and 6 μM (calculated concentration, Section 2.6.2) of the dimerisation impaired mutant (D123P) (Zhou et al., 2006). B. Microscale thermophoresis (Section 2.9.5, $n=3$) was used to calculate the association of IRE1 α 's luminal domain from monomers into presumed dimers. An apparent $K_{1/2}$ value of $0.36 \pm 1.16 \mu\text{M}$ was calculated from the fitted curve, in agreement with previously obtained sub-μM values for this constant in the research group (work carried out by Sam Dawes).

Luminal domain oligomerisation in the absence of peptide, observed here with the full-length protein, has previously been described with the core luminal domain protein (lacking the 60 residue-long juxtamembrane linker region) (Karagoz et al., 2017). The core domain was thusly analysed using the same methodology as in Figure 4.2A (Figure 4.3) to validate that the full-length and core luminal domain exist in similar states at the concentrations used. The data suggests that the same effect is observed for the core domain and is therefore in agreement with the Karagoz et al. (2017) publication and that a similar response is observed with the core domain protein as the full-length. Due to the lack of a monomeric control sample for the core luminal domain, it is not possible to accurately calculate the size of the multimers observed or estimate dimerisation and oligomerisation constants. However, using molecular weight calibrants (Section 2.6.2.2) the size of the smallest species eluted was calculated to be 103kDa, using this estimation it is clear that species over double the size of these elute from the column at higher concentrations, suggesting the appearance of oligomeric species in the core luminal domain also.

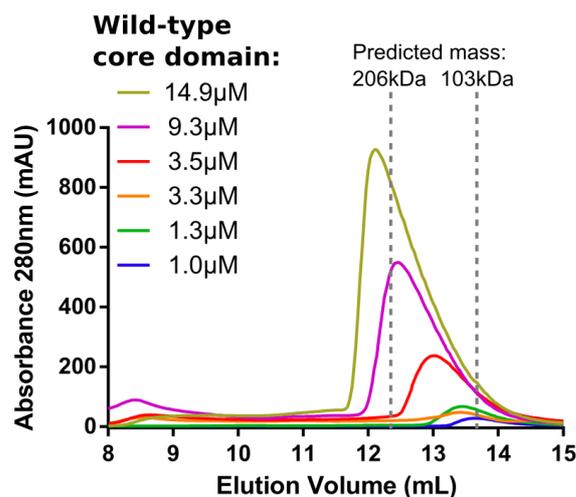


Figure 4.3

Higher concentrations of IRE1 α 's core luminal domain form larger species than dimers, in agreement with previous study (Karagoz et al., 2017). Increasing concentrations of the core luminal domain were analysed on a size exclusion chromatography column (Section 2.6.2). At the highest concentration measured, species greater than two-fold the size of the smallest species observed eluted from the column, suggesting the formation of larger species than dimers.

In this section, IRE1 α 's conformations in the absence of binding partners were investigated. The first finding is that the luminal domain exists as an equilibrium between monomers and dimers at lower concentrations, with a sub- μ M dimerisation constant (in agreement with work carried out by Sam Dawes). This is a much reduced dimerisation constant that observed for yeast protein ($\sim 10\mu$ M) (Credle et al., 2005), suggesting that the processes that determines IRE1 α activation may be vastly different between the two organisms. The increased propensity for human IRE1 α luminal domain dimerisation corroborates with the requirement for ERdj4 to recruit BiP to luminal domain dimers to repress clustering of the domain (Amin-Wetzels et al., 2017). The dimerisation constant for the core luminal domain of IRE1 α has been previously reported to be 2.5μ M by use of analytical ultracentrifugation (Karagoz et al., 2017), this is weaker than the interaction observed here for full-length luminal domain. This may be due to different techniques used or may represent a role of the linker region in promoting dimerisation; ideally the two domains would be compared by microscale thermophoresis to give a more precise comparison.

The second finding presented in this section is that the full-length luminal domain construct is also able to oligomerise in the absence of unfolded protein mimics, with an estimated constant of $>10\mu$ M (Figure 4.2A; different methods of study are required to validate this). This is in agreement with a previous study that suggests that the luminal domain transiently adopts an oligomer-active state and therefore oligomerises at higher concentrations in the absence of peptide (Karagoz et al., 2017). The data presented here therefore suggests that IRE1 α 's luminal domain

exists in a conformational equilibrium in the absence of unfolded protein, with a high affinity to dimerise and higher concentrations required for oligomerisation, as represented in Figure 4.4.

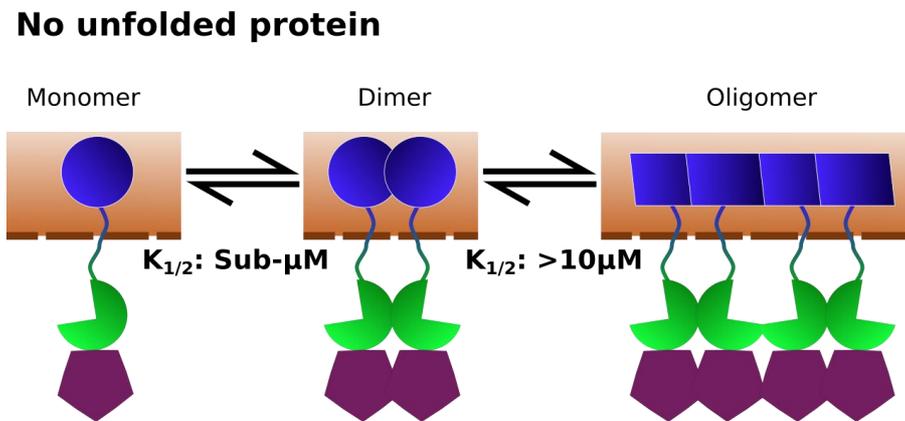


Figure 4.4

Summary of the luminal domain's dynamic equilibrium in the absence of unfolded protein. In the absence of unfolded protein, the luminal domain appears to exist in a concentration-dependent equilibrium between monomeric, dimeric and oligomeric species. Oligomers are formed at higher concentrations due to a presumed transient state adopted by the dimer allowing for oligomers (Karagoz et al., 2017). The apparent dimerisation dissociation constant is sub- μM .

4.3. IRE1 α substrate-induced oligomerisation

Next, the effect of unfolded proteins on the luminal domain's conformational landscape (Figure 4.4) will be investigated. As discussed in Section 1.4, recent evidence suggests that unfolded proteins are able to bind to the human luminal domain's groove region and promote the formation of oligomers through an identified oligomerisation interface (residues 357-361 WLLI) (Karagoz et al., 2017). Further to this, preliminary results (work carried out by Sam Dawes) suggest that interaction with ΔEspP promotes luminal domain oligomerisation and incubation with higher peptide concentrations causes formation of insoluble luminal domain species. Therefore, in the following section the effect of unfolded protein mimics on IRE1 α 's luminal domain's multimeric state and formation of insoluble species will be investigated by use of quantitative methods.

4.3.1. Interaction with peptide induces luminal domain oligomerisation

Initially, to ascertain whether the unfolded protein mimic; ΔEspP (sequence: MKKHKRILALCFLGLLQSSYSAKKKK) (Gardner and Walter, 2011) binds to IRE1 α and initiates oligomerisation, fluorescence polarisation (Section 1.6.1.1) was used (Figure 4.5). In previous studies the peptide used was the fluorescently labelled species, therefore giving

information about peptide binding to the luminal domain (Karagoz et al., 2017; Gardner and Walter, 2011), here information is given about the luminal domain's oligomerisation upon addition of peptide as the luminal domain is the labelled species. An increase in luminal domain size (increase in anisotropy) is observed upon addition of Δ EspP, suggesting that it promotes formation of oligomeric species. As the luminal domain is the fluorescently labelled species and Δ EspP is only 3kDa in size, binding to the 50kDa luminal domain is not expected to elicit a significant increase in the observed polarisation, therefore indicating that this increase in fluorescence polarisation is due to IRE1 α oligomerisation.

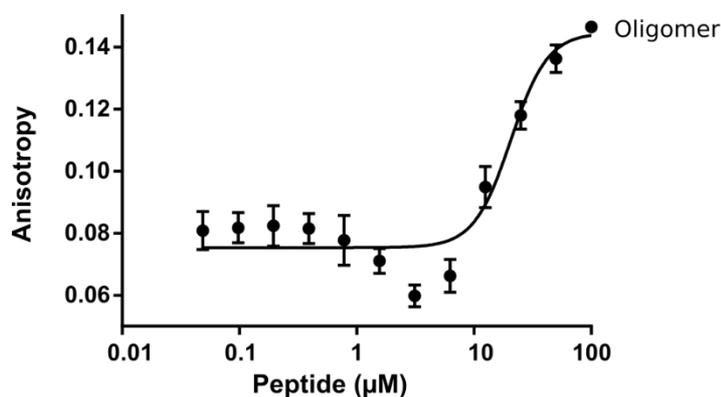


Figure 4.5

Increasing concentrations of Δ EspP cause an increase in the size of IRE1 α luminal domain multimers. Δ EspP peptide sequence: MKKHKRILALCFLGLLQSSYSAAKKKK (Gardner and Walter, 2011). FITC-labelled IRE1 α luminal domain at 50nM was incubated with increasing concentrations of Δ EspP for 30 minutes and the fluorescence polarisation measured (Section 2.9.4, n=3). The apparent constant for this peptide-dependent oligomerisation in wild-type protein is $20.59 \pm 1.13 \mu\text{M}$.

The apparent constant for peptide-dependent oligomerisation is $20.6 \mu\text{M}$. This is weaker than values for peptide binding to the luminal domain observed in the literature ($6.4 \mu\text{M}$) (Carrara et al., 2015), suggesting that the constant observed here involves multiple processes, likely the binding of peptide and the subsequent oligomerisation of the luminal domain.

4.3.2. Quantification of peptide-induced formation of insoluble oligomers

Fluorescence polarisation doesn't allow for quantification of insoluble species formed by the luminal domain. Formation of insoluble species suggests large scale oligomerisation of the luminal domain with increased concentrations of Δ EspP may be occurring, where increasingly large oligomers eventually precipitate. This effect has also been reported for increasing concentrations of the cytoplasmic domain, which forms a rotating filament of dimeric species and

light absorbance measurements suggest high-order oligomerisation (Korennykh, A. and Walter, 2012). Investigation of high-order oligomerisation of the luminal domain can provide more information about a potential mechanism of the luminal domain to facilitate large scale clustering of IRE1 α 's cytoplasmic domain in high stress conditions, as observed previously (Li, H. et al., 2010; Sundaram et al., 2017). Here, turbidity assays are used alongside solubility assays to quantify the formation of larger insoluble species (Borgia et al., 2013) in response to increasing peptide concentrations (representing stress conditions), as described in Section 1.6.3.

4.3.2.1. Δ EspP promotes formation of insoluble species in a concentration-dependent manner

Turbidity assays were first carried out with the *wild-type* luminal domain construct using increasing concentrations of Δ EspP (Figure 4.6A). Increasing concentrations of peptide promote increased OD400 values in the turbidity assay, indicating that larger oligomers are formed in a peptide-dependent manner, this is in agreement with the preliminary data described (work carried out by Sam Dawes). It is also apparent that the luminal domain oligomerises instantly upon addition of peptide and that the measured OD400 values plateau after ~180 minutes, suggesting a time dependence for the formation of larger species.

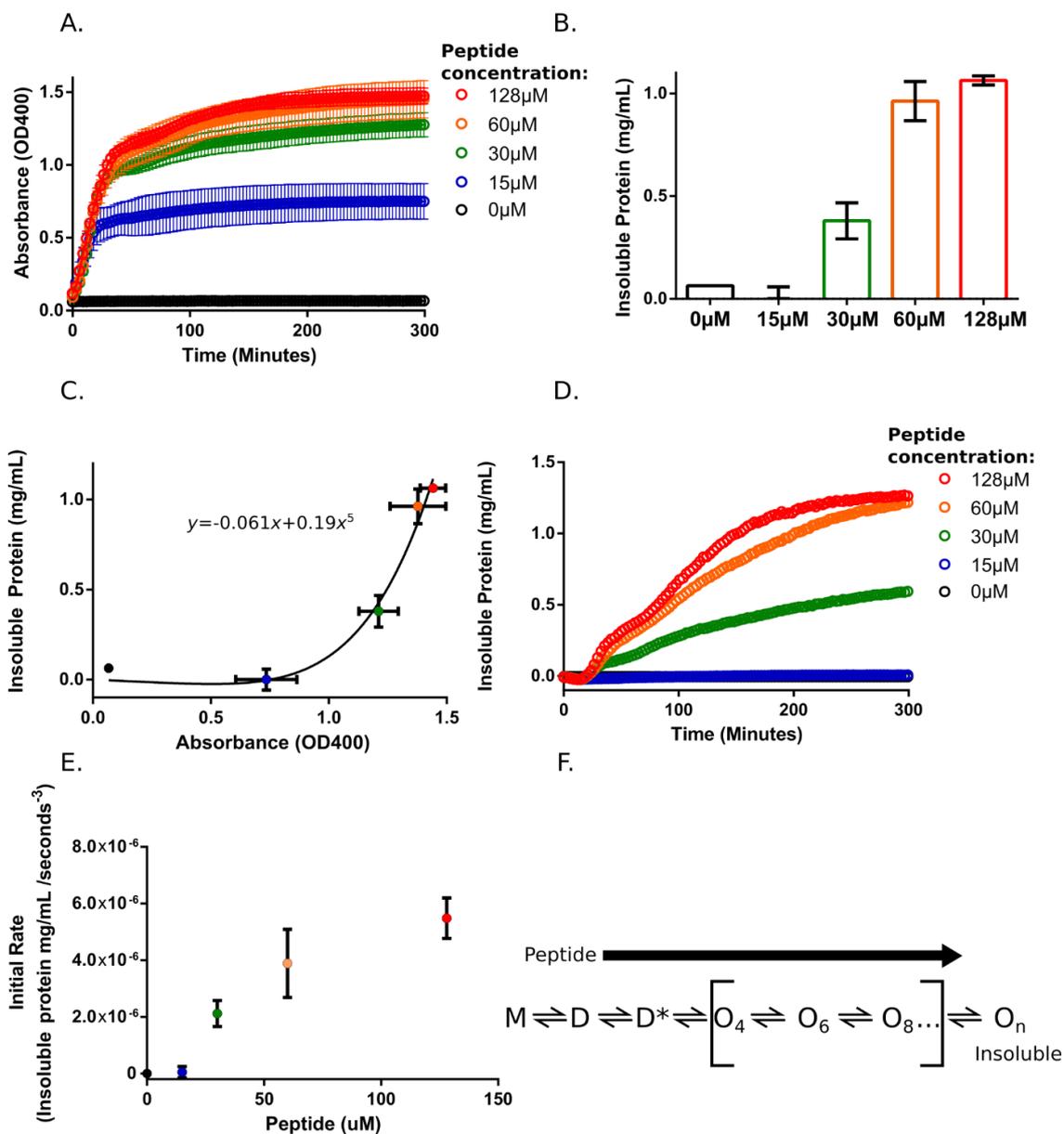


Figure 4.6

Upon addition of peptide, IRE1 α 's luminal domain forms large insoluble oligomers. A. 20 μM of IRE1 α 's luminal domain was incubated with different concentrations of ΔEspP and the OD400 value was measured over time in a turbidity assay (Section 2.9.6.1, $n=3$). B. After 180 minutes the reaction of IRE1 α 's luminal domain with increasing concentrations of ΔEspP has the amount of soluble protein measured and the amount of insoluble protein calculated (Section 2.9.6.2, $n=3$). C. From the results in panel B a calibration curve of the amount of aggregated protein to OD400 absorbance value was produced (equation $y = -0.06083x + 0.1902x^{5.039}$). D. Using the calibration curve in panel C the amount of aggregated protein over time was plotted from the data in panel A. E. The amount of insoluble protein was plotted against time³ (Borgia et al., 2013) to give near linear relationships for each peptide concentration. The gradient of these linear relationships was calculated and plotted against peptide concentration. The rate of large

oligomer formation is dependent on peptide concentration. *F*. The expected mechanism observed by OD400 analysis. *M* is monomeric protein, *D* is dimeric, *D** is the oligomer-active dimeric protein, *O* are oligomers of different sizes observed by OD400, and O_n represents the oligomeric size required for protein precipitation.

Although OD400 measurements cannot report on the size of the species formed, they can report on the amount of insoluble oligomers present and so the turbidity assay conditions were repeated and the amount of soluble and insoluble protein calculated after 180 minutes (Figure 4.6B). A calibration curve was then constructed for the relationship between OD400 measurements and the amount of insoluble protein (Figure 4.6C) (Borgia et al., 2013). The calibration curve was used to understand the peptide concentration and time dependence of insoluble oligomer formation (Figure 4.6D). It is apparent that the amount and rate of insoluble oligomer formation is also dependent on the amount of peptide added to the reaction (Figure 4.6D and E). The mechanism hypothesised to be observed by the turbidity assays is shown in Figure 4.6F, where monomeric protein dimerises, the dimeric protein adopts an oligomer-active state as previously suggested (Karagoz et al., 2017) which forms repeating units of dimeric protein, similar to the oligomers suggested to occur with yeast protein (Credle et al., 2005). Increasing the oligomer unit size eventually causes protein precipitation and as observed here, the presence of peptide pushes equilibrium towards the formation of larger species until they precipitate. To demonstrate the effect of peptide concentration on formation of large insoluble oligomers, the amount of insoluble protein was plotted against the Δ EspP concentration used, the apparent constant was calculated to be 30.5 μ M, although the data is only plotted using five data points (Figure 4.7).

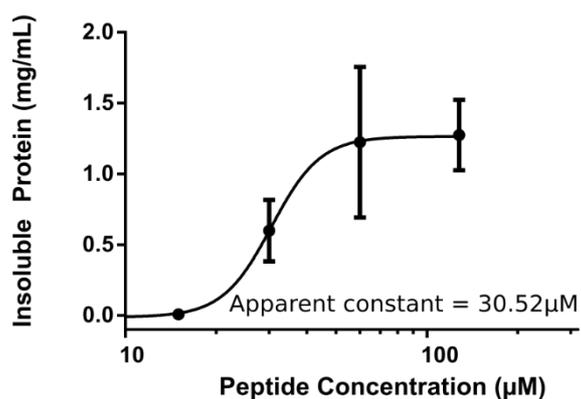


Figure 4.7

Amount insoluble IRE1 α luminal domain after incubation with different amounts of Δ EspP. The end point OD400 values from turbidity assays were measured and the calibration curve in Figure 4.6C was used to calculate the amount of insoluble protein (Section 2.9.6) (Borgia et al., 2013). The amount of insoluble protein was plotted against the concentration of Δ EspP. The constant of the interaction was calculated to be 30.52 \pm 1.12 μ M.

Here, turbidity assays are introduced as a technique capable of further characterising formation of luminal domain oligomers upon addition of peptide. Using turbidity assays the emergence of larger insoluble oligomers and their dependency on peptide concentration is characterised as well as the time taken for the reaction to plateau. The data thus far doesn't conflict with the model shown in Figure 4.6F, that the inclusion of peptide influences IRE1 α 's luminal domain's resting conformational landscape of monomers, dimers and oligomers to promote the formation of larger repeating units of protein, further evidence is supplied in the following sections.

4.3.2.2. The effect of peptide binding affinity on luminal domain oligomerisation

To determine whether the formation of oligomers is dependent on the affinity for the peptide added, the MPZ1 peptide (sequence: LIRYCWLRRQAALQRRISAME) was used. The MPZ1 peptide has previously been shown to bind to the luminal domain with lower affinity than Δ EspP (MPZ1: $24 \pm 4.7 \mu\text{M}$ Δ EspP: $6.4 \mu\text{M}$) (Carrara et al., 2015; Karagoz et al., 2017). The lower affinity of interaction for the peptide produces a reduced affinity for the formation of oligomers and insoluble oligomers of IRE1 α (Figure 4.8), where only the highest peptide concentration used produced insoluble protein. This demonstrates that the MPZ1 peptide is able to shift the luminal domain's equilibrium to promote oligomerisation and formation of large insoluble oligomers and further suggests that peptide binding to the luminal domain's groove region and therefore peptide affinity influences the luminal domain's conformational landscape.

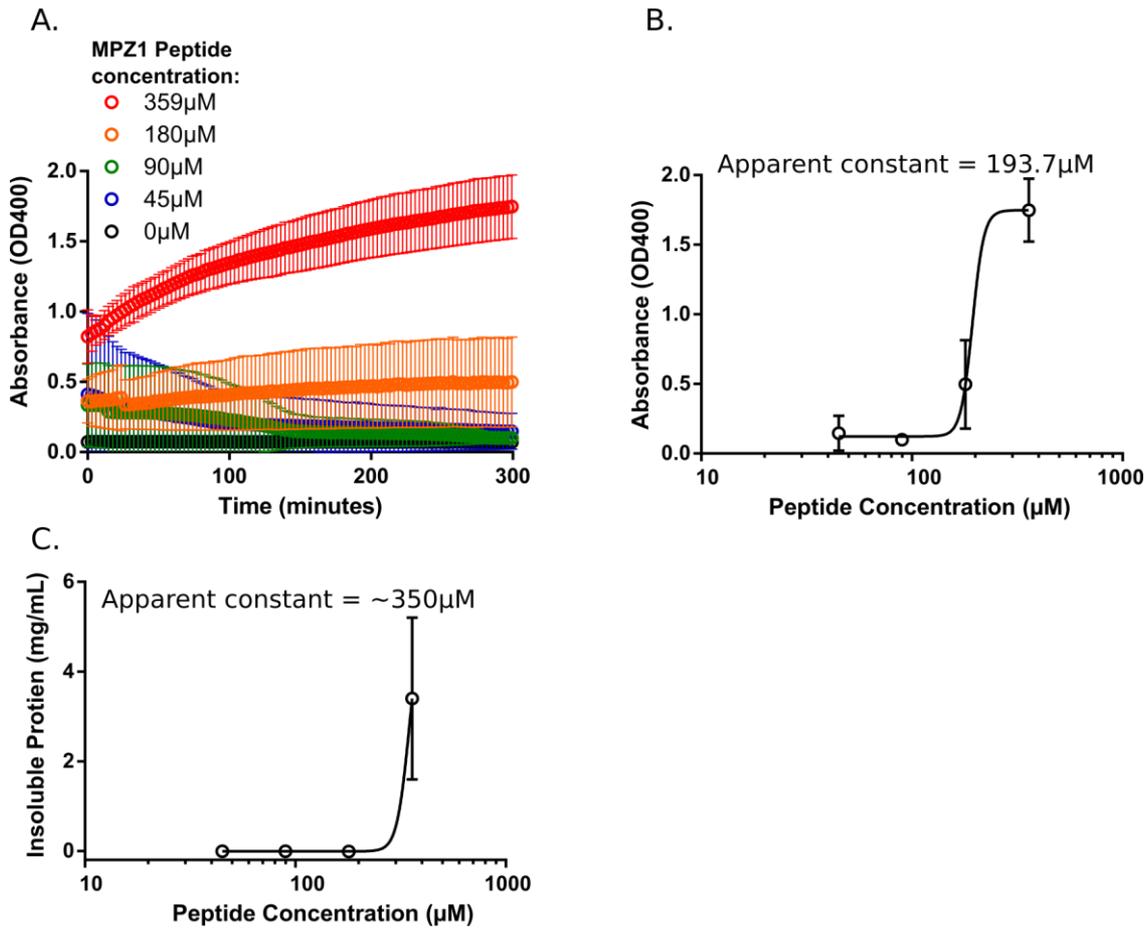


Figure 4.8

MPZ1 peptide promotes IRE1 α 's luminal domain to oligomerise. A. MPZ1 peptide causes an increase in the OD400 value in turbidity assays (Section 2.9.6.1, $n=3$) over time. MPZ1 peptide sequence: LIRYCWLRRQAALQRRISAME. The effect appears to be dependent on peptide concentration, however, the effect is weaker than with Δ EspP (Figure 4.6A). B. Plotting the OD400 turbidity measurement against the peptide concentration used demonstrates that this interaction is weaker than Δ EspP and gives a lower apparent constant of 193.7 μ M. C. The OD400 values were corrected to give the amount of insoluble protein as in Figure 4.6D. Only the highest concentration of MPZ1 peptide incubated with the luminal domain promoted formation of insoluble protein. Due to only one data point giving insoluble protein formation the calculated constant is inaccurate.

4.3.2.3. Luminal domain interface mutants affect peptide-induced oligomerisation

To provide further evidence of the luminal domain's conformational landscape being influenced by peptide binding, dimerisation and oligomerisation deficient mutants were used (Figure 4.9). Both mutations show reduced activity of the protein in endoplasmic reticulum stress models as they shift the protein's equilibrium to disfavour formation of dimers (D123P) (Zhou et al., 2006)

and oligomers (WLLI-GSSG) (Karagoz et al., 2017). The dimerisation deficient mutant perturbs the β -sheet floor at the dimerisation interface of the luminal domain (Zhou et al., 2006) and the oligomerisation mutant was identified using the known oligomerisation interface of the yeast protein (Credle et al., 2005) and mutating it in the human protein (Karagoz et al., 2017).

The oligomerisation interface mutant is able to bind peptides with a similar affinity as *wild-type* protein, suggesting that tetramers and small oligomers are not required for peptide binding, as observed in Karagoz et al. (2017). The authors suggest that the WLLI-GSSG mutant does not oligomerise at 5 μ M with equimolar amounts of peptide added (MPZ1-N peptide 5.4 μ M affinity) (Karagoz et al., 2017), this suggests that ~40% of IRE1 α is bound to peptide, which would not induce significant oligomerisation of the luminal domain detectable by turbidity assays (lowest condition that induces MPZ1-induced oligomerisation is estimated as ~85% of luminal domain is bound, Figure 4.8A). The two interface mutants used here can be considered to disfavour formation of luminal domain oligomers. Therefore, if inclusion of peptide modulates the luminal domain's conformational landscape's equilibrium to promote the formation of oligomers, these mutants are expected to oppose this effect but not prevent it.

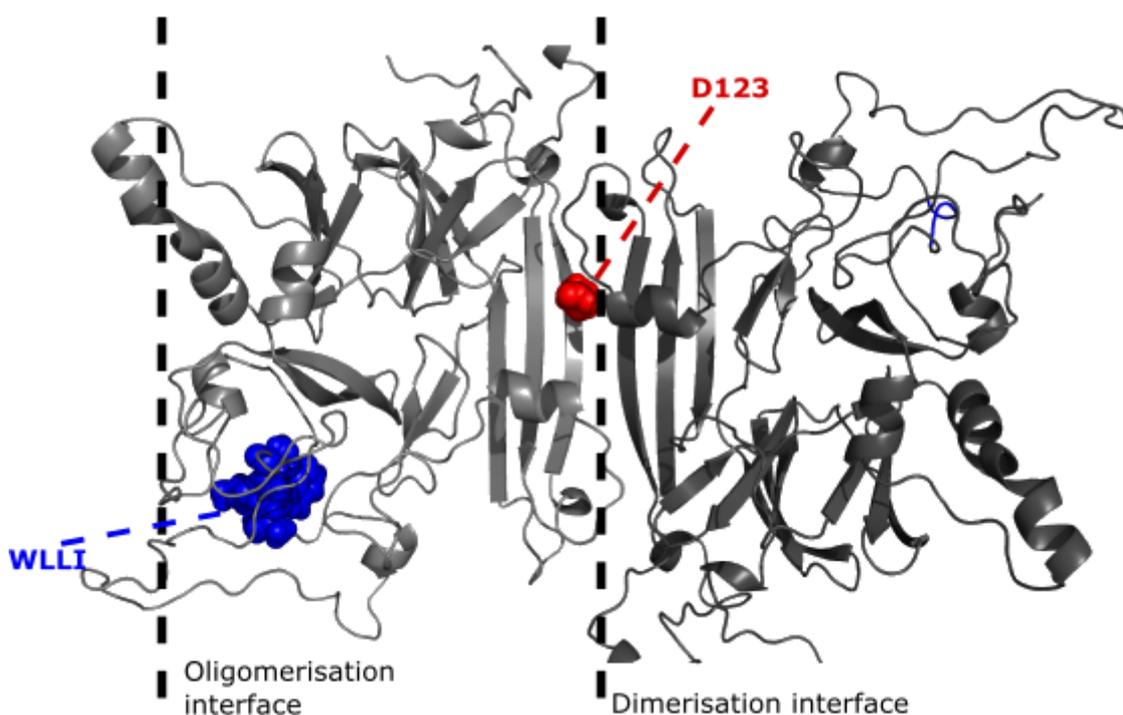


Figure 4.9

The locations of the D123 residue and 357-361 WLLI motif that are mutated to create dimerisation and oligomerisation impaired mutants respectively are shown (Zhou et al., 2006; Karagoz et al., 2017). Shown also are the dimerisation and oligomerisation interfaces of the

luminal domain on a structural model based on the luminal domain's crystal structure (PDB: 2HZ6) (Zhou et al., 2006; Yang et al., 2015). The mutants were created as described in Section 2.3.8. Figure prepared using PyMOL version 1.7.

Using turbidity assays with the oligomerisation disfavoured mutants as in Figure 4.6, it is apparent that interaction with peptide still induces oligomerisation of the mutants, but higher concentrations of peptide are required for both mutants compared to *wild-type* protein to induce large scale oligomerisation detectable by the technique (Figure 4.10). The apparent constant for formation of insoluble species is reduced significantly for both mutants, assuming that the mutants do not affect the affinity for Δ EspP binding to the luminal domain (Figure 4.11C). Interestingly, the formation of larger insoluble oligomers appears to be inhibited more by the dimerisation deficient D123P mutant. The protein has an increased propensity to form dimers (sub- μ M constant) than oligomers ($>10\mu$ M constant) in the absence of unfolded protein (Figure 4.4), and therefore disfavoured dimerisation may represent a larger shift in the protein's equilibrium, therefore inhibiting peptide-induced oligomerisation to a greater extent. Interestingly, this suggests that the monomeric protein is able to bind unfolded proteins to promote its dimerisation and oligomerisation, however further investigation of this mechanism is required.

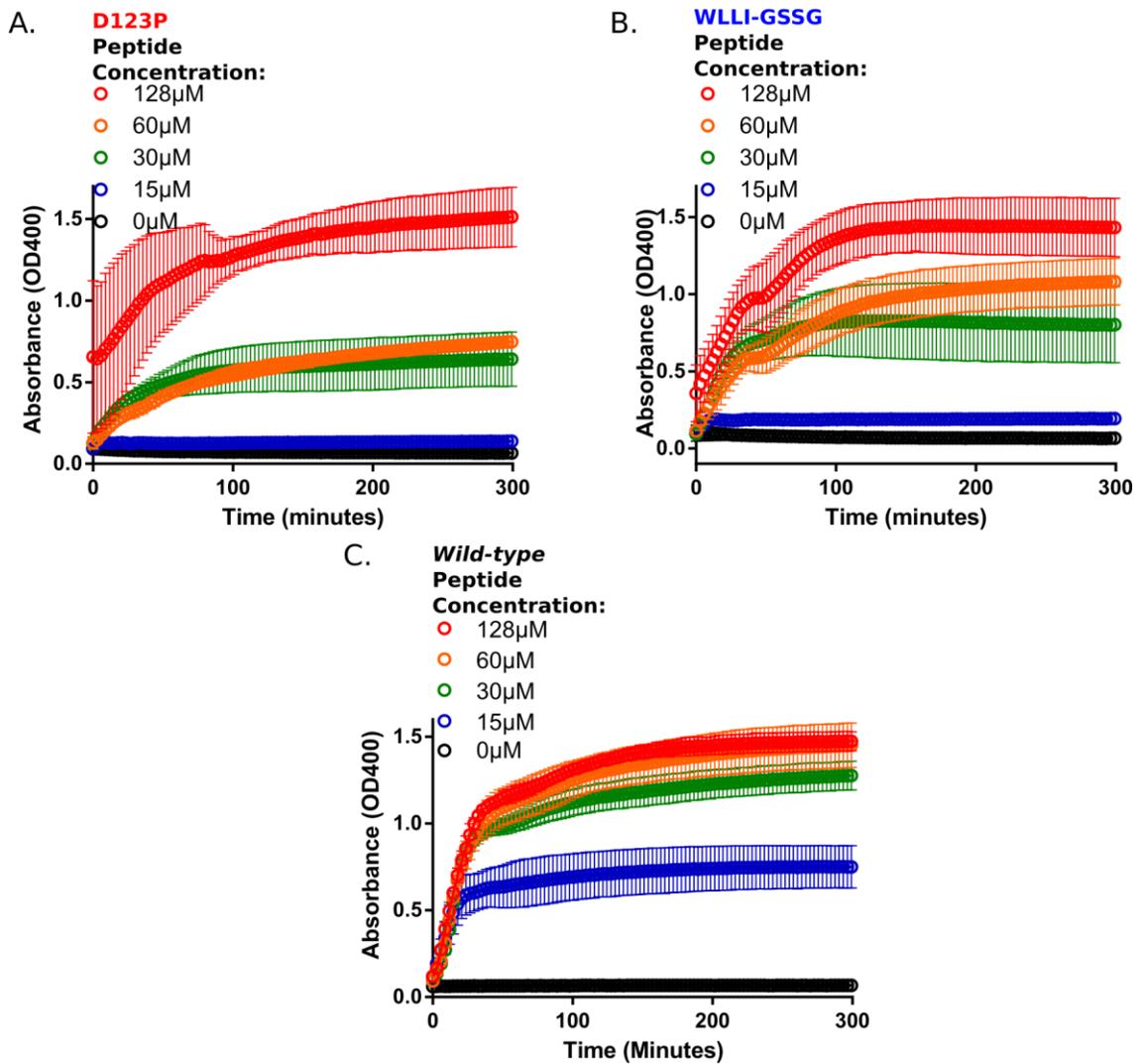


Figure 4.10

The dimerisation- and oligomerisation-deficient mutants (D123P and WLLI-GSSG respectively) of IRE1 α 's luminal domain also form large oligomers over time upon the inclusion of Δ EspP as measured in turbidity assays. A. D123P forms oligomers detectable by turbidity assays (Section 2.9.6.1, $n=3$) in a manner that is dependent on the concentration of Δ EspP added. B. The WLLI-GSSG mutant of IRE1 α forms oligomers also in a similarly peptide-dependent manner ($n=3$). C. For comparison, the wild-type protein's turbidity assay measurement is shown.

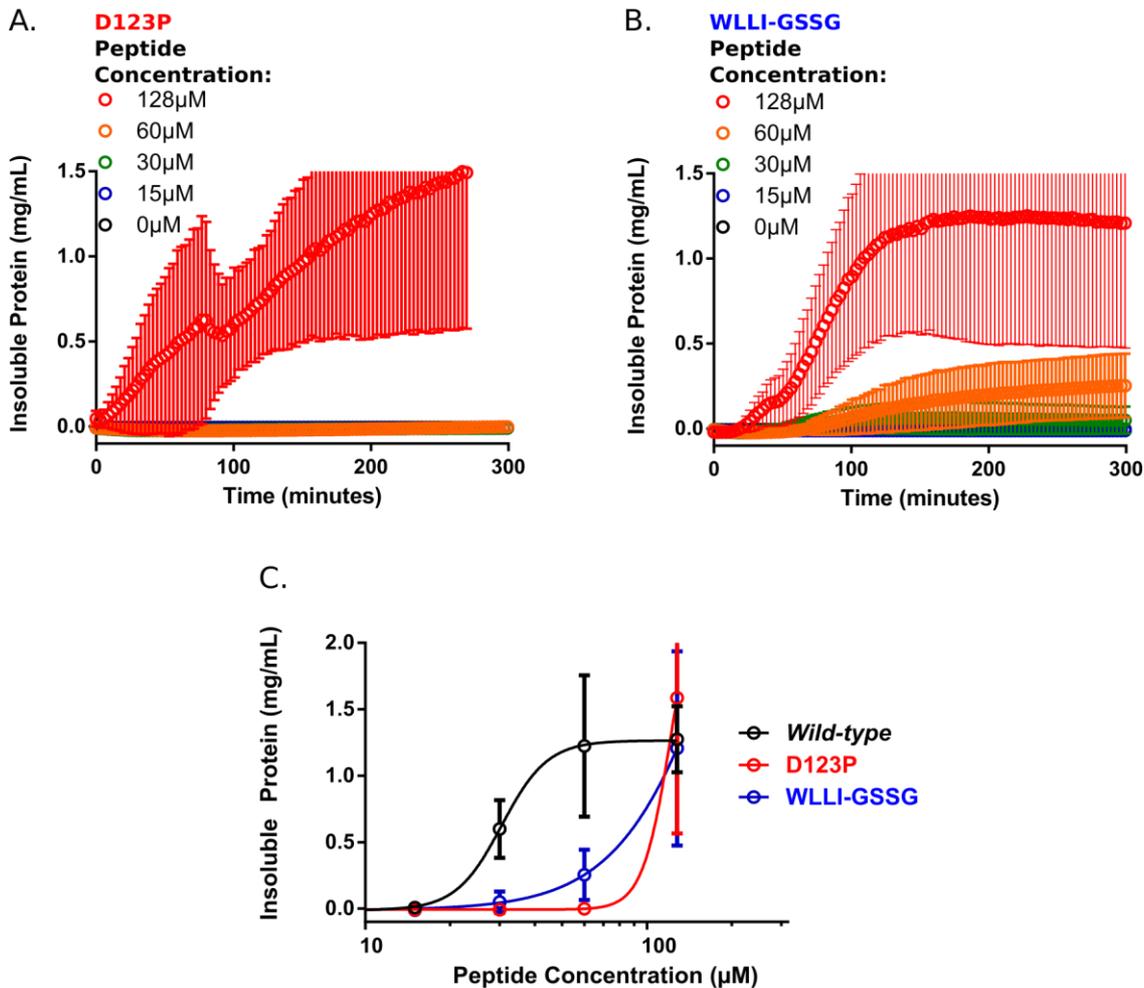


Figure 4.11

The dimerisation perturbed and oligomerisation perturbed mutants (D123P and WLLI-GSSG respectively) of IRE1 α 's luminal domain also form large insoluble oligomers over time upon the inclusion of Δ EspP as measured in turbidity assays. A. D123P forms insoluble oligomers with larger concentrations of Δ EspP required (Section 2.9.6.1 with insoluble protein correction from Figure 4.6C, $n=3$). B. The WLLI-GSSG mutant of IRE1 α forms insoluble oligomers with the inclusion of higher concentrations of Δ EspP required ($n=3$). C. Upon comparison of wild-type protein with the two mutants, the mutants shift the protein's equilibrium so that larger concentrations of peptide are required to promote insoluble oligomer formation. It is notable that the calibration curve for OD400 values to insoluble protein (Borgia et al., 2013) was constructed using wild-type protein and there is a high degree of error after conversion of the mutants' values to insoluble protein.

This result provides further evidence that interaction with unfolded protein mimic peptides influences the luminal domain's conformational landscape and shifts equilibrium towards the formation of oligomeric species (Figure 4.12). The effects on the luminal domain's conformational landscape observed *in vitro* hint at how its equilibrium may be shifted in cellular

responses also, where the amount of unfolded proteins and BiP present influences the proteins conformational landscape to push equilibrium towards larger or smaller species, promoting different downstream responses. Additionally, there may be a time barrier to formation of larger oligomeric species in high stress conditions.

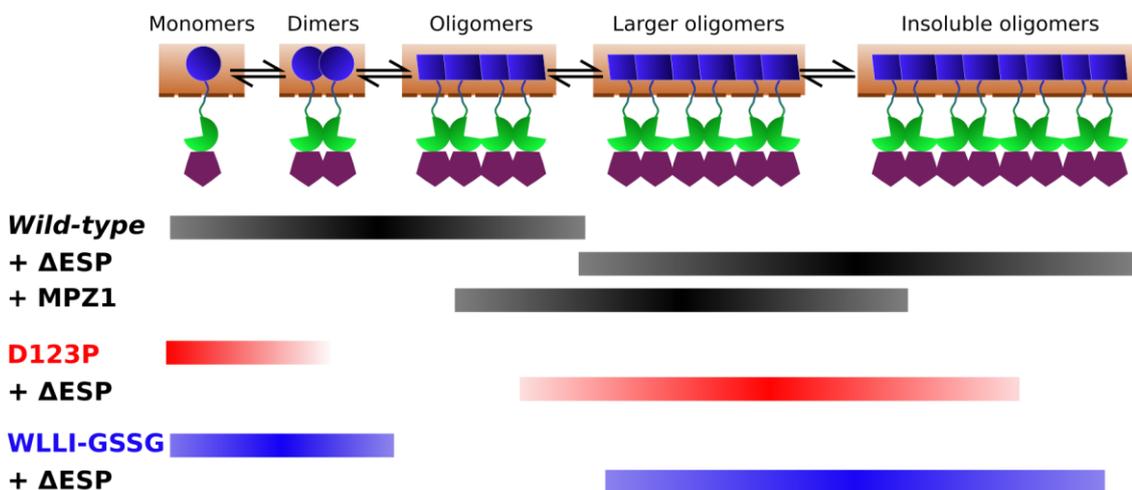


Figure 4.12

Summary for the effects of interface mutations and different peptides on the luminal domain's conformational landscape. Shown are representations of the states that the luminal domain mostly occupies in each condition (shown as coloured bars) as determined by the data presented in this section, Section 4.2 and previous literature for the interface mutants (Zhou et al., 2006; Karagoz et al., 2017).

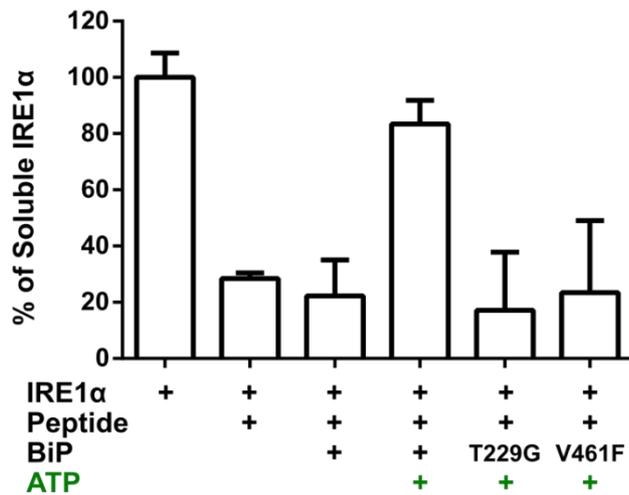
4.4. BiP's interaction with oligomerised IRE1 α

With the increased understanding of luminal domain peptide-induced insolubility from the previous section, BiP's mechanism to de-oligomerise luminal domain oligomers (preliminary data by Sam Dawes and Łukasz Wieteska) can be quantitatively and more precisely investigated. The following section will therefore characterise the BiP and ATP-dependent termination the peptide-induced insoluble oligomers of IRE1 α 's luminal domain.

Insoluble luminal domain oligomers were formed before BiP was incubated with them, firstly in the absence of ATP and solubility tests were carried out (Figure 4.13). Unexpectedly, BiP stably binds to the insoluble luminal domain oligomers in the absence of ATP, observed by BiP's appearance in the insoluble fraction upon incubation. After incubation with ATP the insoluble luminal domain and BiP species are observed in increasing quantities in the soluble fraction over time. Additionally, the BiP mutants T229G (unable to hydrolyse ATP) and V461F (substrate

binding deficient) were used, both mutants were unable to de-oligomerise and re-solubilise the luminal domain oligomers. These results suggest that BiP stably binds to the luminal domain oligomers before ATP hydrolysis induces release of BiP from the oligomers which also promotes a change in luminal domain conformation to induce de-oligomerisation, this suggests that this mechanism may be a typical chaperone-substrate interaction (Schmid et al., 1994).

A.



B.

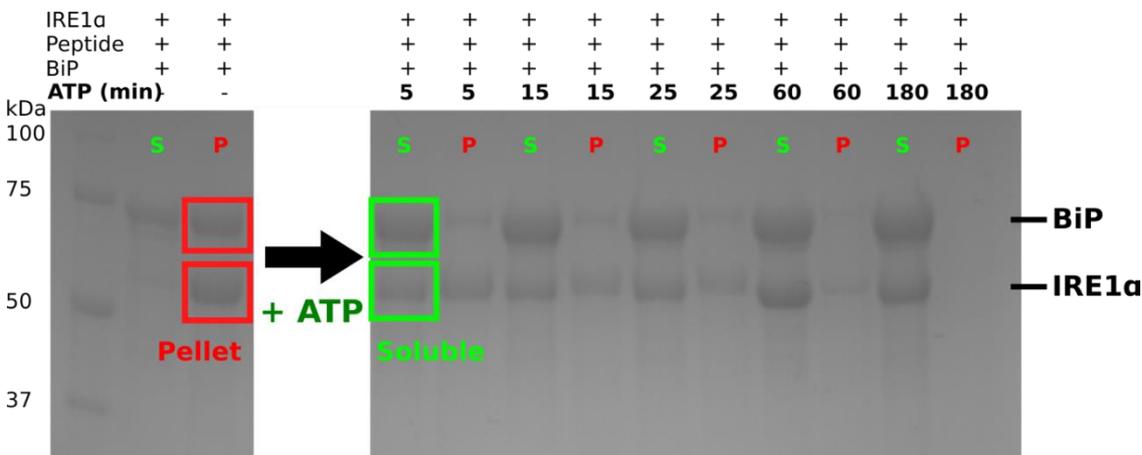


Figure 4.13

BiP's canonical ATP-dependent activity re-solubilises insoluble IRE1α luminal domain oligomers. A. The soluble concentration of IRE1α's luminal domain was measured after incubation in different conditions (Section 2.9.7, n=3). The conditions used are as follows: 20μM IRE1α alone, after incubation with 169μM ΔEspP for three hours, after this incubation with peptide and an incubation with BiP for three hours and a sample with these steps but with a further incubation with 40mM ATP for three hours. The final condition was also carried out with the T229G and V461F BiP mutants that prevent BiP's canonical activity. This shows that BiP uses its canonical ATP-dependent activity to re-solubilise insoluble IRE1α oligomers. B. SDS-

PAGE gel (Section 2.5.5) of the reaction taking place in panel A, gel samples were taken at different time points of the ATP incubation after IRE1 α 's luminal domain had been incubated with peptide and BiP for three hours each. Over time, in the presence of ATP, IRE1 α is brought from the pellet (insoluble fraction) to the soluble fraction.

To validate that the mechanism observed is indeed a transient ATP-dependent chaperone interaction (therefore expected to proceed with sub- μ M concentrations of BiP), the ratio of luminal domain to BiP was adjusted and ratios of ~1:1 (5:6), 20:1 and 100:1 were analysed (Figure 4.14). In all cases the inclusion of BiP with an excess of ATP was able to re-solubilise the insoluble peptide-induced IRE1 α oligomers. This further validates that the mechanism is a chaperone-like transient interaction of BiP in terminating active luminal domain oligomers and potentially the IRE1 α signal *in vivo*.

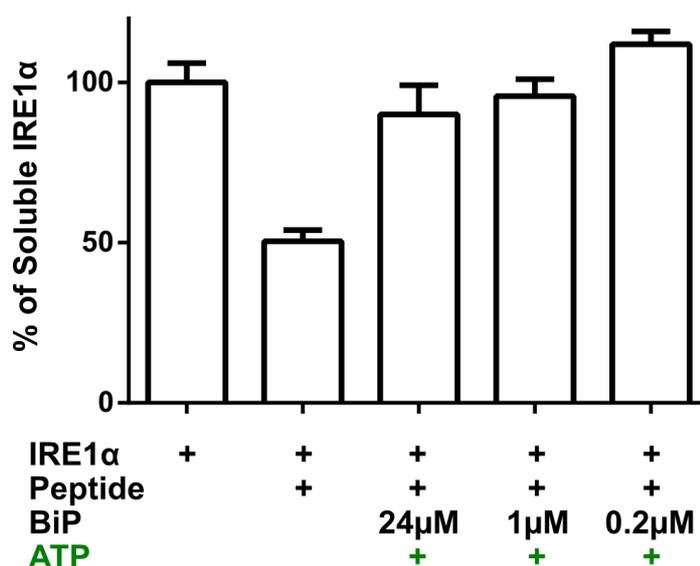


Figure 4.14

BiP re-solubilises insoluble IRE1 α oligomers using its chaperone activity. Lower concentrations of BiP are able to rescue the insoluble IRE1 α oligomers after incubation with 40mM ATP for three hours. The same conditions as in Figure 4.13 were used (Section 2.9.7).

To ensure that BiP's chaperone activity to re-solubilise the luminal domain isn't a general chaperone activity that BiP would have for an aggregating protein, the luminal domain was first heated to 60°C for one hour to promote aggregation and the sample was allowed to cool before BiP and ATP were added (Figure 4.15). As expected, in the absence of co-chaperones (Craig et al., 2006; Zuiderweg et al., 2013), BiP was not able to re-solubilise the heat denatured luminal domain after a three hour incubation. Therefore, this suggests that BiP's canonical activity to

solubilise the luminal domain is specific to the insoluble luminal domain oligomers formed with inclusion of peptide.

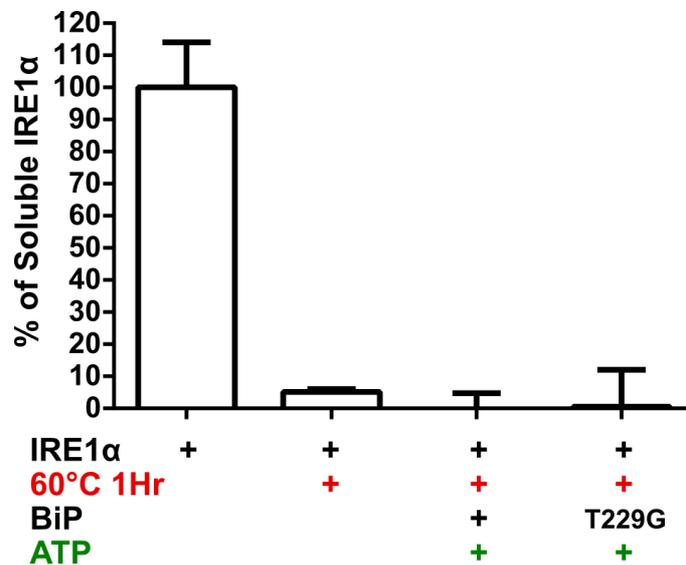


Figure 4.15

BiP is unable to rescue heat aggregated IRE1α luminal domain. IRE1α's luminal domain was aggregated and therefore insoluble after incubation at 60°C for one hour. After this the sample was left to cool to room temperature before BiP was added with 40mM ATP for three hours. BiP was unable to resolubilise the aggregated IRE1α luminal domain (n=3).

From the evidence presented here, it is clear that BiP is able to bind to and solubilise insoluble IRE1α luminal domain oligomers through its canonical ATP-dependent chaperone activity. It was first considered that this process may be a promiscuous chaperone activity, however, heat-aggregated luminal domain is not solubilised by the same mechanism. It was also considered that it may be an effect of BiP binding to peptide as to prevent it from interacting with the luminal domain, and therefore destabilising the formed oligomers. However, there is an excess of peptide used compared to BiP and IRE1α, and there is also evidence that BiP is unable to bind to the ΔEspP used in these assays by NMR (work carried out by Sam Dawes) and isothermal calorimetry studies (Carrara et al., 2015). Methods can now be used to gain an understanding about the conformation of the luminal domain insoluble oligomeric species and further details about BiP's stable interaction with them.

4.5. Conformational characterisation of the insoluble oligomers and BiP interaction

Formation of insoluble oligomers means that the sample is no longer amenable to study by structural techniques such as X-ray crystallography and solution NMR. However, electron microscopy and solid-state NMR excel at studying large insoluble particles (Iadanza et al., 2018; Sun, S. et al., 2012) and are therefore used to characterise the conformation of the insoluble luminal domain peptide-induced oligomers and BiP's interaction with these.

4.5.1. Insoluble luminal domain species form fibrillar assemblies

Negative-stain electron microscopy was used to observe the insoluble luminal domain oligomers and interestingly, long fibrillar structures were identified (Figure 4.16). In samples of IRE1 α in the absence of peptide no such structures were encountered after extensive observation of the sample grid. This suggests that the fibril structures observed are oligomers of the luminal domain which are the insoluble species observed.

The oligomerisation interface identified in the human protein (Karagoz et al., 2017) and the solved oligomeric structure of the yeast luminal domain (Credle et al., 2005) suggest that the fibrils observed may be repeating units of the luminal domain dimer. Indeed, when the thickness of the fibrils formed is compared to the size expected based on the crystal structure of the human protein, the fibril's diameters are in good agreement with that of the dimers observed in the X-ray crystal structure model (Figure 4.17) (Zhou et al., 2006; Yang et al., 2015). The small degrees of curvature in the fibrils are likely due to the intrinsically disordered region (loop region 2, Section 1.4.1) surrounding the proposed oligomerisation interface (Karagoz et al., 2017). Figure 4.16C shows a group of fibrils, possibly suggesting an interaction between fibrils, similar to that observed in the yeast crystal structure (Credle et al., 2005), however, this requires validation, possibly through mutational studies.

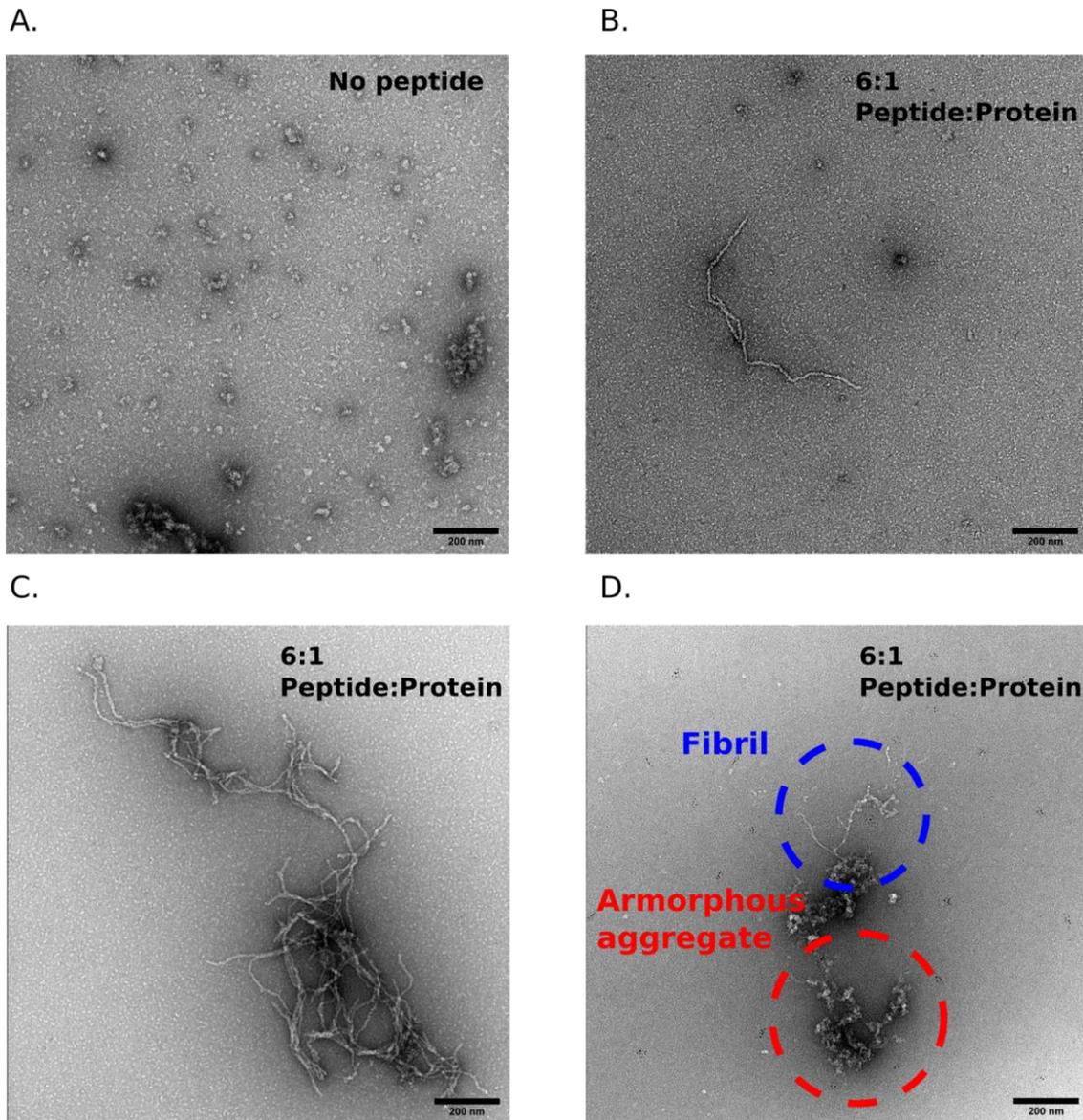


Figure 4.16

TEM images show IRE1 α 's luminal domain forming fibrils upon inclusion of Δ EspP. Protein at 20 μ M was incubated with and without 128 μ M Δ EspP for 90 minutes before grids were prepared with 1:10 diluted sample (Section 2.9.8). The size bar in all images is 200nm. A. No fibrils were observed in samples in the absence peptide, multiple locations on the grid were observed but no fibrils were identified. B. When incubated with Δ EspP at a molar ratio of 6:1, fibrils are observed, two fibrils seemingly interacting with each other are shown here. C&D. Further examples of fibrils observed for the luminal domain upon incubation with peptide. Observed is the difference between the fibrils and amorphous aggregate present in the sample.

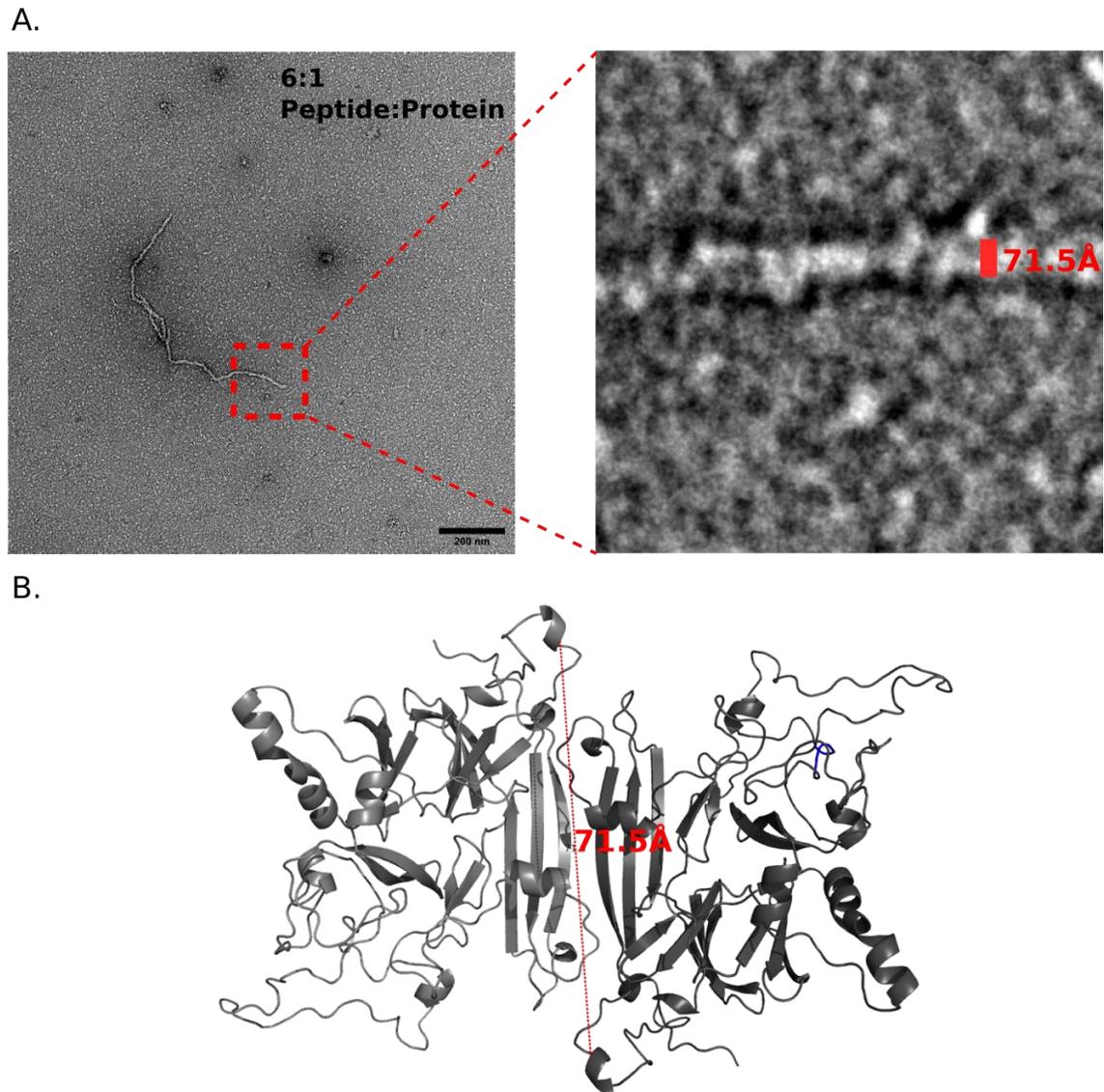


Figure 4.17

TEM images show IRE1 α 's luminal domain forms fibrils upon inclusion of Δ EspP of a similar size to that expected if formation is through the suggested oligomerisation interface. A. The fibril shown in Figure 4.16B is expanded with a size bar representing 71.5Å. This is the expected width of the fibril if it forms through the luminal domain's identified oligomerisation interface (Karagoz et al., 2017). The fibril has good agreement with this size. B. Crystal structure of the luminal domain (PDB: 2HZ6) (Zhou et al., 2006) with unresolved regions modelled (Yang et al., 2015). A measurement has been made from the top to the bottom of the dimer model, this is the expected width of the formed fibril, shown in panel A. Figure prepared using PyMOL version 1.7.

4.5.2. The fibrils consist of folded protein and BiP binding affects the luminal domain's disordered regions

To further investigate the conformation of the insoluble fibrils and BiP's interaction with them, solid-state NMR was used (Figure 4.18), this technique allows study of insoluble proteins and gives residue specific information about the changes in environment or dynamic processes that the protein experiences (Sun, S. et al., 2012; Iadanza et al., 2018). Samples of insoluble oligomers of the isotopically labelled luminal domain were analysed by Prof. József Lewandowski (University of Warwick).

The solid-state NMR data suggests that the insoluble luminal domain oligomers in the presence of BiP consist of folded protein rather than misfolded aggregates (Figure 4.18A). The spectrum in Figure 4.18A shows well dispersed peaks in the ^1H dimension, where each peak represents one HN group in the protein. The large ^1H dispersion of the peaks suggests that regions of secondary structure are present (Okazaki et al., 2018; Gupta and Bhattacharjya, 2014). Therefore, it's apparent that the luminal domain is in a folded state and is further proof that the insoluble oligomers are not amorphous aggregate.

To investigate BiP's interaction with the folded, insoluble luminal domain oligomers, HC spectra were compared in the presence and absence of BiP (Figure 4.18B). In these spectra each peak reports on the chemical environment of a HC group in the protein's disordered regions. Interestingly, the changes in the peak positions upon inclusion of BiP suggest that BiP binds to or affects specific disordered regions of the luminal domain protein. However, further study by the techniques is required to determine the precise binding sites of BiP.

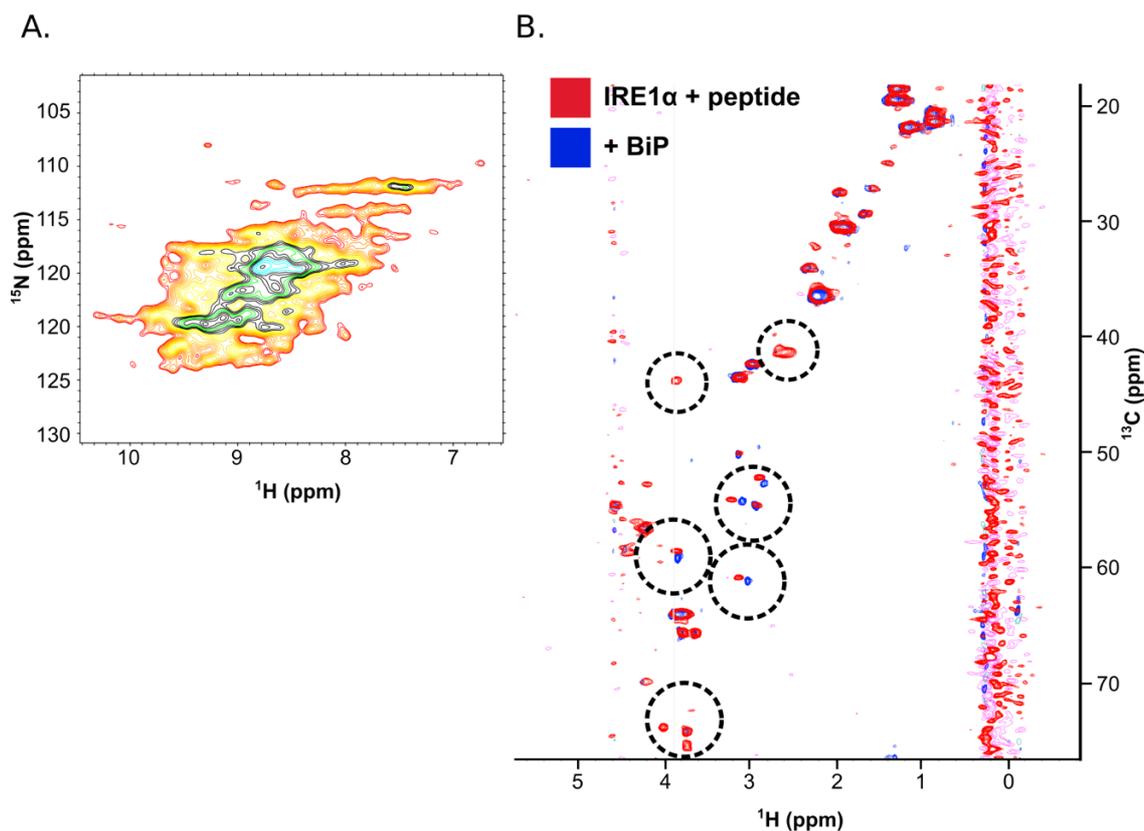


Figure 4.18

Solid state NMR results for the insoluble luminal domain peptide-bound oligomers in the presence and absence of BiP. Samples prepared as in Section 2.7.2 and analysed by solid-state NMR by József Lewandowski (University of Warwick). A. Solid-state NMR spectrum for the insoluble luminal domain peptide-bound oligomers in the presence of BiP suggest that the luminal domain is folded. A HN cross polarisation experiment shows that the insoluble luminal domain oligomers have regions of structure (large ^1H dispersion). Each peak reports on the chemical environment of one HN group in the protein. B. A HC-heteronuclear single quantum correlation experiment, where each peak represents a proton bonded to a carbon atom in the backbone of disordered luminal domain residues. Highlighted are regions where there are chemical shift perturbations with the inclusion of BiP, suggesting these residues of the luminal domain's disordered regions are involved in BiP's interaction and therefore have altered chemical environments.

4.6. Conclusion

In this chapter, data supporting the hypothesis that IRE1 α 's luminal domain exists primarily as an equilibrium between monomers and dimers in solution (Figure 4.2B) is presented, with a similar dimerisation constant to that reported for the protein lacking the linker region (Karagoz et al., 2017) and in agreement with values gained previously in the Zhuravleva research group. It has been suggested that the luminal domain is able to transiently adopt a conformation that

promotes oligomerisation and therefore at higher concentrations forms oligomeric protein (Karagoz et al., 2017), the results presented here corroborate this finding and suggest that the same oligomers are adopted by the full-length luminal domain that also includes the domain's juxtamembrane linker region (Figures 4.2A and 4.3).

It is known that unfolded proteins and peptides interact with the luminal domain to promote oligomerisation (Karagoz et al., 2017) and that higher concentrations of Δ EspP promote insoluble luminal domain species to form (work carried out by Sam Dawes). Here, methods to characterise the formation of these oligomers have been developed and give information about the apparent constant values for peptide-induced oligomerisation and formation of insoluble oligomers, providing evidence that formation of larger species is dependent on peptide concentration. The results here also suggest that inclusion of peptide affects IRE1 α 's conformational landscape identified in the absence of peptide and shifts equilibrium to the formation of oligomers. This is evidenced by dimerisation and oligomerisation inhibiting mutations that shift the equilibrium towards smaller species, requiring higher concentrations of peptide to induce oligomerisation and formation of insoluble species (Figure 4.11).

In this chapter the formation of luminal domain fibrils is observed upon incubation with high concentrations of Δ EspP (Figure 4.16). The fibrils appear to be repeating units of the dimeric protein, likely forming through the human protein's identified oligomerisation interface (Karagoz et al., 2017), in a similar way as was observed in the yeast IRE1 luminal domain oligomeric structure (Credle et al., 2005). The insoluble fibrils are also confirmed to contain folded luminal domain protein through analysis by solid-state NMR (Figure 4.18). Similar long fibrillar structures may form when the protein is activated in high stress conditions in the cell and be responsible for the clustering of IRE1 α observed previously (Li, H. et al., 2010). This may represent IRE1 α forming a signalling hub similar to that created from large scale oligomerisation of the MAVS protein (Hou et al., 2011). Interestingly, the timescale for fibril formation here is similar to that reported for maximal clustering of IRE1 α (Li, H. et al., 2010), although the *in vitro* conditions are unlikely to accurately mimic cellular stress conditions, the result does suggest a time barrier for luminal domain clustering for a maximal signalling response.

Previously BiP's ATP-dependent activity to resolubilise peptide-induced IRE1 α oligomers was observed (work carried out by Sam Dawes and Łukasz Wieteska). Here the effect was further characterised to give a novel mechanism for termination of the luminal domain's activated form. BiP is able to bind to the insoluble luminal domain fibril-like oligomers, which appears to be through specific binding sites in the luminal domain's disordered regions as identified by solid-state NMR (Figure 4.18B). Using the optimised methods for study of luminal domain insoluble

oligomerisation, data is presented here that clearly shows BiP's ATP-dependent and chaperone-like activity to resolubilise the luminal domain oligomers (Figures 4.13 and 4.14).

The results presented here allow for improvement of the previous models with more information about how the luminal domain is activated and terminated, as shown in Figure 4.19. The mechanism of the identified BiP interaction appears to be by binding to luminal domain oligomers and terminating the luminal domain activated signal without the requirement of ERdj4. Previous models suggest that BiP represses IRE1 α by binding to luminal domain dimers in an interaction promoted by ERdj4 (Amin-Wetzel et al., 2017). However, it is not yet clear whether ERdj4 is also able to mediate the process of BiP binding to and resolubilising luminal domain oligomers and therefore requires future investigation. Interestingly, in previous studies, BiP has coprecipitated with activated IRE1 α and prevention of BiP's interaction with IRE1 α has slowed the disassembly of activated luminal domain oligomers (Pincus et al., 2010; Liu et al., 2002). These findings suggest that the mechanism explored here may have been previously identified in cellular models but was not further characterised.

BiP binding to the folded IRE1 α luminal domain represents an interaction uncommon in Hsp70 chaperones, with only a small number of examples of BiP binding to a disordered region of a folded protein to stabilise a conformation. Some examples include HSC70's interaction with disordered motifs in clathrin triskelions (to promote de-coating), the glucocorticoid receptor and the σ^{32} transcription factor (Clerico et al., 2015). In the case of IRE1 α , the luminal domain has many regions of disorder, particularly in the long loop region 2 (Section 1.4.1) at its oligomerisation interface, it is possible that BiP is able to bind one or multiple disordered sites here to destabilise the active form of the protein, promoting a change in conformation to abrogate oligomerisation, peptide binding or both.

The negative stain images of fibrillar IRE1 α and the fact that BiP is able to bind to these as well as the solid-state NMR data for each condition represent pilot structural studies of the two mechanisms. The data represents the possibility of further structural studies by solid-state NMR and cryo-electron microscopy to elucidate information about peptide binding to the human luminal domain protein, the oligomerisation interface and the conformations of the oligomerised luminal domain and BiP's interaction with these oligomers. The assays used to observe the different mechanisms involved in the luminal domain's activation and termination also provide a way to observe how influencing factors such as the identified cancer-associated mutations affect the protein's dynamic equilibrium.

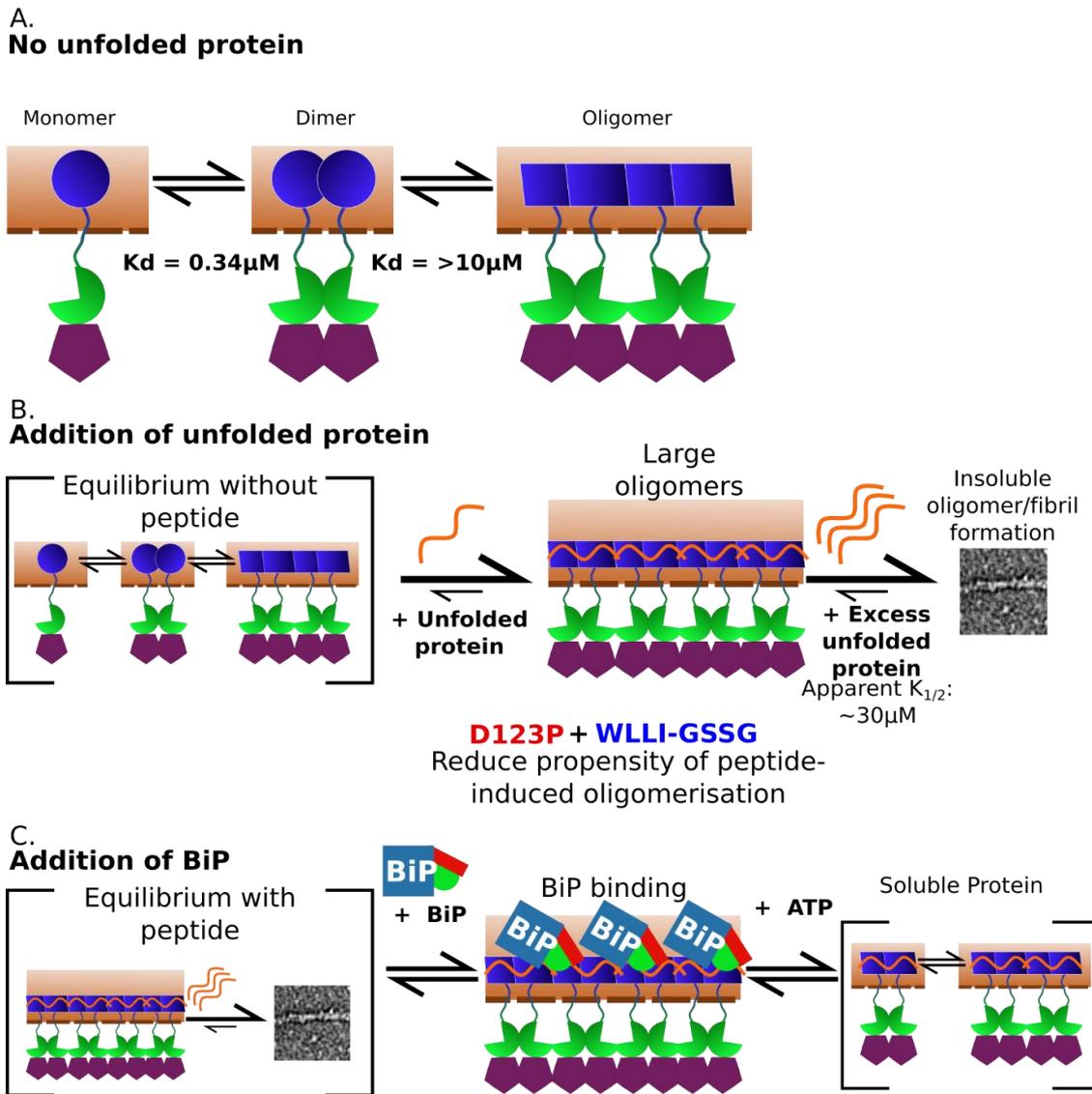


Figure 4.19

The results of this chapter add new features to the understanding of the luminal domain's activation. A. As previously reported (Karagoz et al., 2017), the data presented here suggests that larger species of the luminal domain are able to form without the presence of unfolded protein. Additionally, it has been shown that the oligomers forming from the transient state adopted by the luminal domain are able to form with the presence of the juxtamembrane linker region also. B. Upon inclusion of peptide the luminal domain readily forms oligomers, this has been identified previously (Karagoz et al., 2017; Gardner and Walter, 2011) and has been validated by the results presented here. Oligomers form over time with *in vitro* activation plateauing after three hours. The emergence of large insoluble oligomers has been identified and these characterised to be fibrils of the luminal domain that consist of folded luminal domain multimers. The formation of oligomers is in a dynamic equilibrium with other states of the protein, and when the equilibrium is perturbed by dimerisation- and oligomerisation-deficient mutants there is a lower affinity for oligomerisation upon addition of peptide. C. Here a novel mechanism for termination of unfolded

protein-bound luminal domain oligomers has been identified and characterised. BiP is able to stably bind to the large oligomers through specific binding sites in the protein's disordered regions, and upon addition of ATP is able to solubilise the large oligomers and bring IRE1 α back into solution through a chaperone-like activity.

5. Characterisation of the luminal domain's intrinsically disordered regions using solution NMR

5.1. Introduction

In the previous chapter the conformational landscape of IRE1 α 's luminal domain was explored using various biophysical assays and how mutations and inclusion of binding partners can affect states adopted by the protein. From previous study of the luminal domain there is some understanding about the structured regions of the protein that are functionally important. The solved crystal structure and mutational studies revealed the dimerisation interface (Zhou et al., 2006), the luminal domain's oligomerisation interface has been inferred through study of the yeast (Credle et al., 2005) and human proteins, regions involved in promoting oligomerisation have been identified through use of NMR (Karagoz et al., 2017) and the unfolded protein binding site has been determined (Credle et al., 2005; Kono et al., 2017).

The protein's intrinsically disordered regions have much less known about them even though the luminal domain has a high proportion of disordered regions (Section 1.4.1). These disordered regions include the domain's two loop regions and the 60-residue linker region that tethers the globular portion of the protein to the transmembrane helix (Figure 5.1). It's possible that this linker acts to give the luminal domain flexibility from the membrane, allowing it to dimerise and oligomerise and interact with its various binding partners (Sepulveda et al., 2018; Amin-Wetzel et al., 2017; Eletto et al., 2014; Sundaram et al., 2017). Numerous studies of the luminal domain's interactions *in vitro* have omitted the linker region from the construct used and through these studies the linker region has been deemed unnecessary for BiP repression (Amin-Wetzel et al., 2017; Carrara et al., 2015), unfolded protein-induced oligomerisation (Karagoz et al., 2017; Credle et al., 2005) and dimerisation (Zhou et al., 2006). This is in agreement with Section 4.2, which suggests that the core luminal domain (lacking the linker region) exists in an equilibrium similar to full-length protein as determined by size exclusion chromatography.

No structural information has been obtained for the intrinsically disordered regions of the luminal domain, however, there is also evidence for the linker region's role in IRE1 α activation and repression. The most N-terminal segment of the linker region (residues 390-408) was determined to have a role in BiP repression, where omission of these residues reduced BiP binding to the protein and prompted increased *XBP-1* splicing activity of IRE1 α (Oikawa et al., 2009). A mutation to the central region of the linker region (A414T) has been partially characterised in a

cell model to cause an increase in IRE1 α oligomerisation and activity (Lhomond et al., 2018). In addition to this, the most C-terminal region of the linker region has been identified to have a role in transmitting lipid bilayer stress signals in the yeast and human proteins to activate the luminal domain (Halbleib et al., 2017; Kono et al., 2017) and is suggested to be the binding site for the repressive Sec61 translocon protein (Sundaram et al., 2017).

Interestingly, secondary structure predictions of the linker region suggest a high α -helical propensity in two of these three segments of the linker region, the most N-terminal and the most C-terminal segments as well as a small region in between the two (Figure 5.2) (Shen, Y. et al., 2014; Thevenet et al., 2012; Drozdetskiy et al., 2015). The evidence for the importance of the luminal linker region therefore suggests that three different sections of the linker region may have a role in activation and therefore warrant further study. Study of the linker region's interactions in the multimeric states of the luminal domain that were isolated through mutagenesis and peptide binding in Chapter 4 may elucidate potential allosteric interactions and roles of the different sections of the linker region.

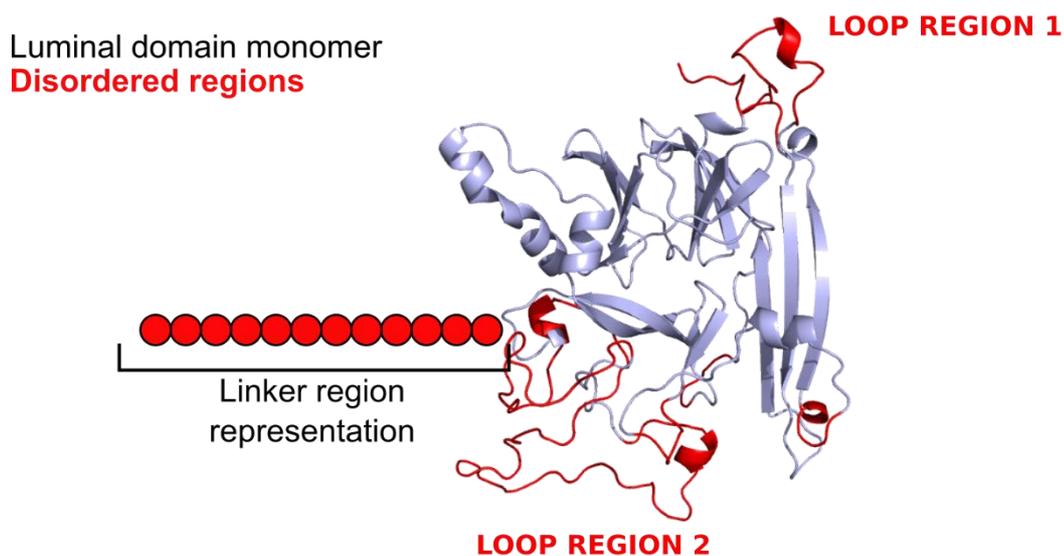


Figure 5.1

The crystal structure model of the luminal domain monomer (PDB: 2HZ6) (Zhou et al., 2006) with remodelled loop regions (Yang et al., 2015). Regions of disorder are shown in red including a cartoon representation of the luminal domain's linker region. Figure prepared using PyMOL version 1.7.

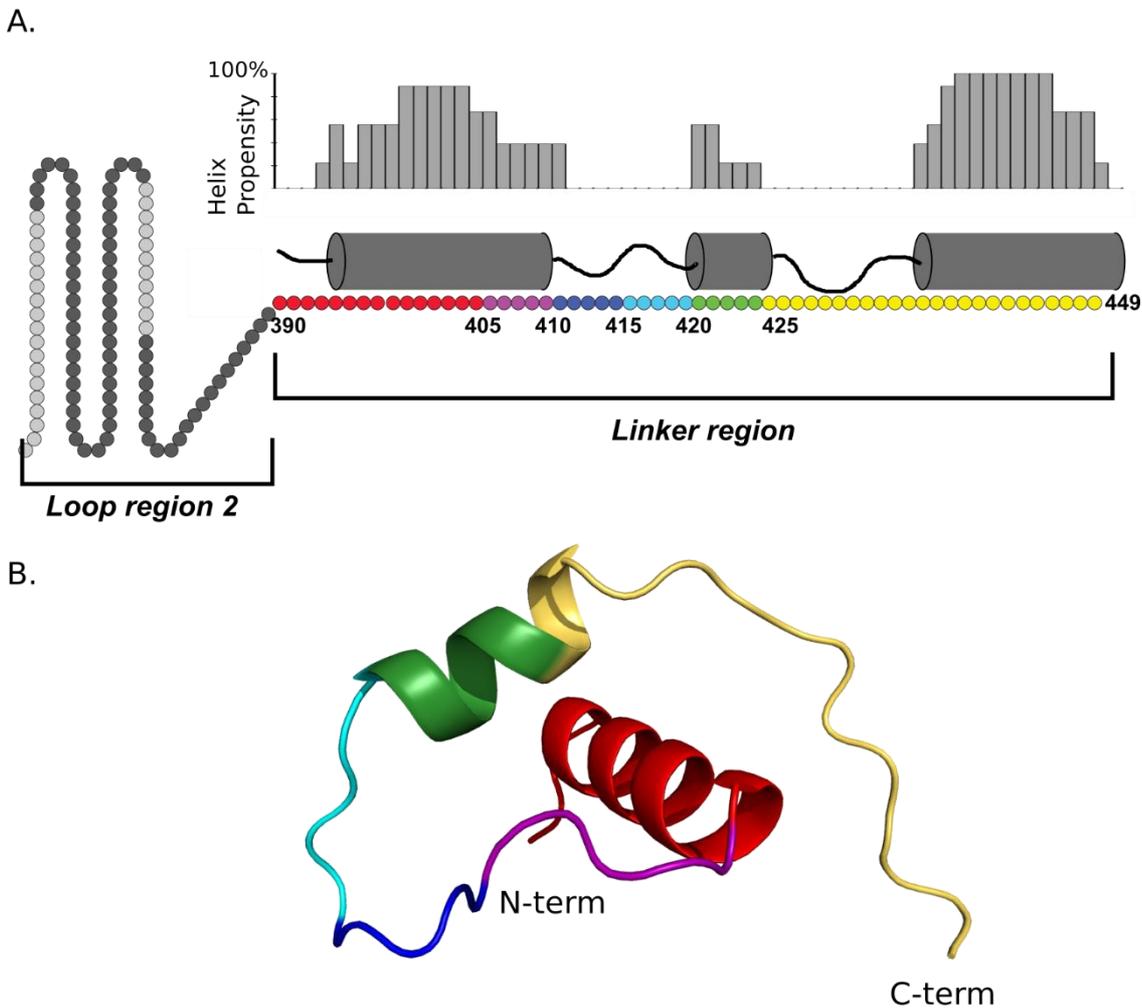


Figure 5.2

The linker region has multiple regions with a high predicted propensity for helix formation. A. The helix propensity of the linker region. Helix propensity of the linker region was calculated using three different programmes; *jpred4* (Drozdetskiy et al., 2015), *PEP2D* and *PEP-FOLD* (Shen, Y. et al., 2014; Thevenet et al., 2012), where each result was given equal weighting to calculate an overall helix propensity prediction. B. A structural model of the linker region's secondary structure from *PEP-FOLD* (excluding C-terminal proximal end where high α -helical propensity is predicted). Figure prepared using PyMOL version 1.7.

X-ray crystallography study of the luminal domain has given an understanding of its structured regions, but the technique is unable to give information about the protein's disordered regions. NMR on the other hand provides an excellent technique for the study of intrinsically disordered regions, methods for which have been well established (Kosol et al., 2013; Brutscher et al., 2015) and therefore NMR will be used to study the luminal domain's disordered regions to uncover their potential roles in activation and regulation.

5.2. Initial characterisation of the luminal domain by solution NMR (repeated work of Sam Dawes)

Previous research carried out the Zhuravleva group by Sam Dawes compared the full-length and core luminal domains' HN TROSY spectra (NMR methodology described in Sections 1.6.1.3 and 2.7.1.2) (Pervushin et al., 1997). The two experiments are overlaid (experiments repeated here and shown in Figure 5.3) and suggest that a significant amount of the luminal domain's residues are not observed by NMR, that all residues observed are suggested to be from disordered regions of the protein (peaks have a $^1\text{H}^{\text{N}}$ chemical shift dispersion between 7.5 and 8.5 ppm) (Okazaki et al., 2018) and that the majority of the HN backbone peaks observed report on linker region residues. Additionally, analysis of mutants in each of the core domain's intrinsically disordered loop regions suggested that the core domain peaks observed are from the longer loop region 2 (Figure 5.1; Work carried out by Sam Dawes, personal communication). The lack of structured regions observed in the luminal domain's spectra is likely due to the protein exhibiting multiple conformations in solution therefore causing broadening of peaks in the spectrum, this has been observed previously in NMR study of the protein (Karagoz et al., 2017). This suggests that the residues observed by NMR compliment those observed in the solved crystal structure (Zhou et al., 2006), and so the technique can be used to complete our understanding of the luminal domain's conformations and the role of the protein's intrinsically disordered regions.

Therefore, in this chapter a detailed solution NMR characterisation of the visible intrinsically disordered regions of the luminal domain (particularly the linker region) will be performed and their roles in the different conformations of the luminal domain's conformational landscape will be investigated.

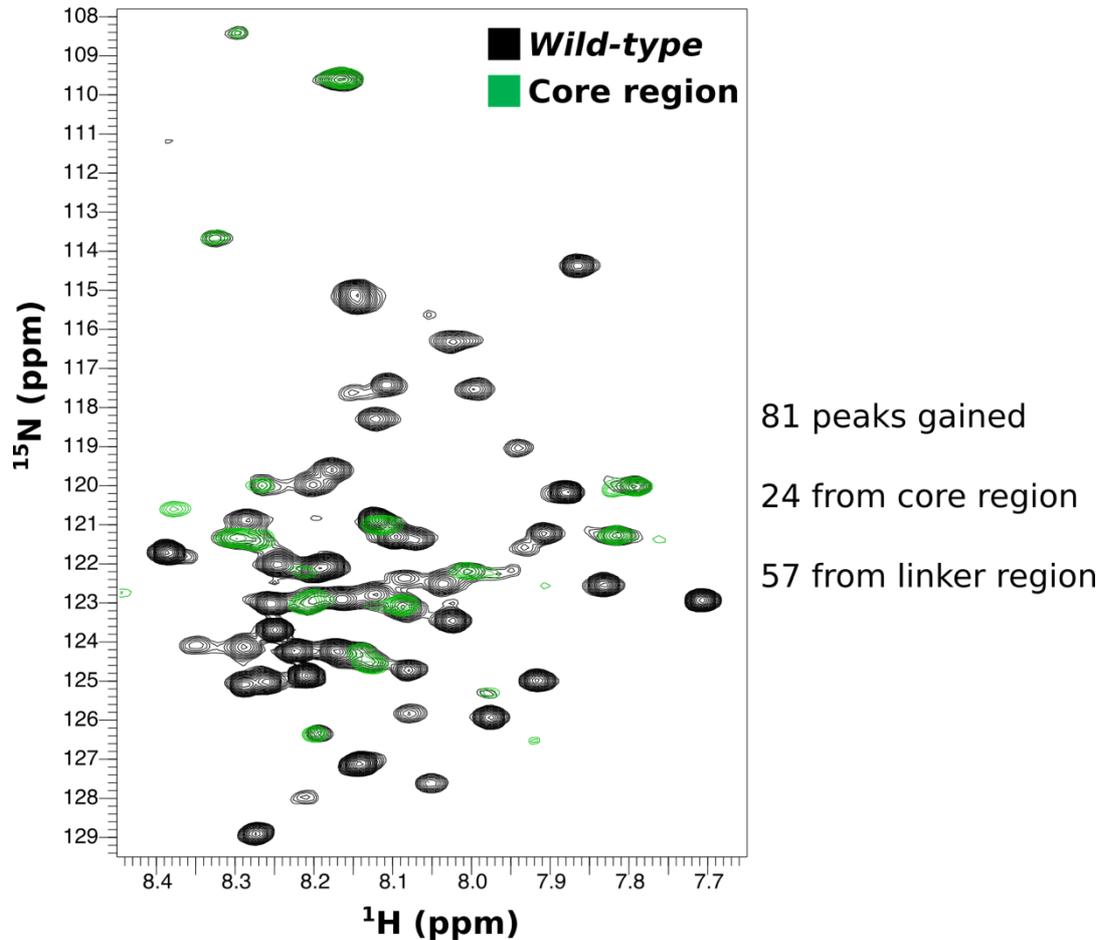


Figure 5.3

The HN amide TROSY spectrum for wild-type luminal domain is overlaid with the spectrum of the core region that lacks the linker region (both at 50 μ M). It is clear to see which peaks are from residues of the core region of the protein (suggested to be from the longer loop region 2) and that the majority of peaks observed are from linker region residues (Section 2.7.1.2).

5.3. Partial assignments of IRE1 α 's luminal domain spectrum

In order to gain more information from the luminal domain's NMR spectrum, peaks can be assigned to their corresponding residues. Usually this would be carried out using a set of triple resonance NMR experiments. However, due to the quality of the luminal domain's NMR spectrum this is not possible. Instead, mutagenesis can be used to mutate or introduce truncations to the protein, peaks that disappear in the subsequent spectrum represent the removed residues, thus allowing assignments to be carried out. This method of assignment is a commonly used alternative to traditional NMR assignment (Siivari et al., 1995; Wieteska et al., 2017).

In order to assign the luminal domain's spectrum multiple truncations were made to the C-terminal of the linker region. The truncation mutants were expressed with ^{15}N -labelling and analysed by NMR HN TROSY experiments (Figure 5.4). Using the multiple truncations of the linker region, the truncated segments were assigned. Figure 5.4A shows the truncated constructs sequentially, one example of assignment can be observed in the panel with the 24-410 and 24-405 constructs displayed. The 24-405 construct is shown in purple and has 5 less residues than the construct shown in black, the five peaks disappearing in the shorter construct therefore indicate that these peaks belong to the five-residue region that is removed. In this way the whole spectrum was assigned and is shown in Figure 5.4B. Using these assignments, further study of the protein by NMR can give more insight into the role of the different segments of the linker region in the protein's adopted conformations.

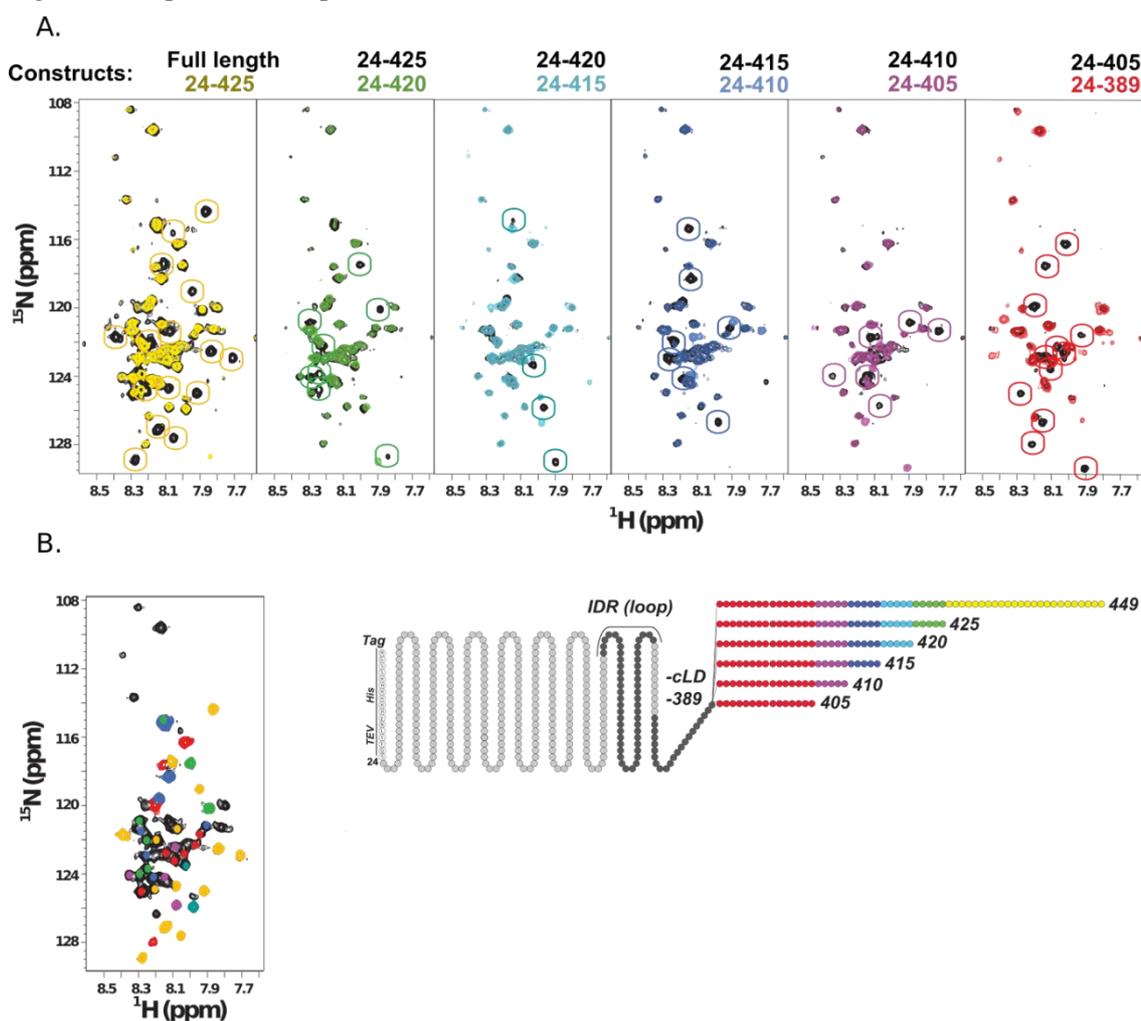


Figure 5.4

Truncation mutants were used to assign different areas of the linker region. A. Shown are spectra for each truncation mutant (Section 2.3.8 for mutagenesis) used overlaid onto the next longest construct of the luminal domain in black (Section 2.7.1.3). Circled are the chemical shift perturbations from which assignments of the deleted segments of the linker region were made. B.

Summary of which peaks belong to which area in the linker region on the HN TROSY spectrum, and a schematic of the truncation mutants used.

5.4. $^1\text{H}^{\text{N}}$ Chemical shift temperature coefficients suggests that no secondary structure is observed

The linker region is predicted to adopt secondary structure (Figure 5.2), however, ^1H dispersion in the HN TROSY experiment suggests that all residues visible are disordered as all peaks are observed in the 7.5 -8.5 $^1\text{H}^{\text{N}}$ ppm range (Figure 5.3). This method of determination is thought to be prone to false negatives (Okazaki et al., 2018). Therefore, to validate the disorder of observed peaks more accurate parameters were used, one such parameter is the amide ($^1\text{H}^{\text{N}}$) chemical shift temperature coefficient (Tomlinson and Williamson, 2012; Cierpicki and Otlewski, 2001; Okazaki et al., 2018), as described in Section 1.6.1.3. In this analysis HN TROSY spectra are first obtained at different temperatures. Each peak's $^1\text{H}^{\text{N}}$ chemical shift (ppm) is highly dependent on the experimental temperature in a linear function due to the presence of hydrogen bonds from the residue's backbone, which are indicative of secondary structural elements (Cierpicki and Otlewski, 2001; Tomlinson and Williamson, 2012; Okazaki et al., 2018). Therefore, a threshold value for this relationship can be used to define intrinsically disordered regions of the protein. The less stringent -4.6 ppb/K threshold value was found to report on the presence of hydrogen bonds from the residue's backbone (Cierpicki and Otlewski, 2001), however, chemical shift temperature coefficients have also been suggested to be weak reporters on hydrogen bond length and strength (Tomlinson and Williamson, 2012). Recently, the more stringent threshold value of -3.6 ppb/K has been suggested to account for structured regions that are still partially solvent exposed and therefore increase the accuracy of intrinsically disordered region determination (Okazaki et al., 2018).

The luminal domain was therefore analysed by HN TROSY experiments at different temperatures and the $^1\text{H}^{\text{N}}$ chemical shift temperature coefficient values calculated to determine intrinsically disordered regions using the -3.6 K/ppm threshold value (Figure 5.5). Using this methodology, all peaks analysed appear to be from disordered regions of the protein, corroborating the chemical shift range (7.5-8.5 ppm) that the peaks appear in. Use of the partial assignments suggest a good coverage of the luminal domain's disordered regions in this analysis, although the analysis does exclude 11 loop region 2 peaks (Table 5.1). This is due to the core domain peaks having lower intensity values, the lower peak intensities of these peaks indicate that these residues may be involved in dynamic processes and possibly the adoption of secondary structure and further investigation into these regions is therefore required.

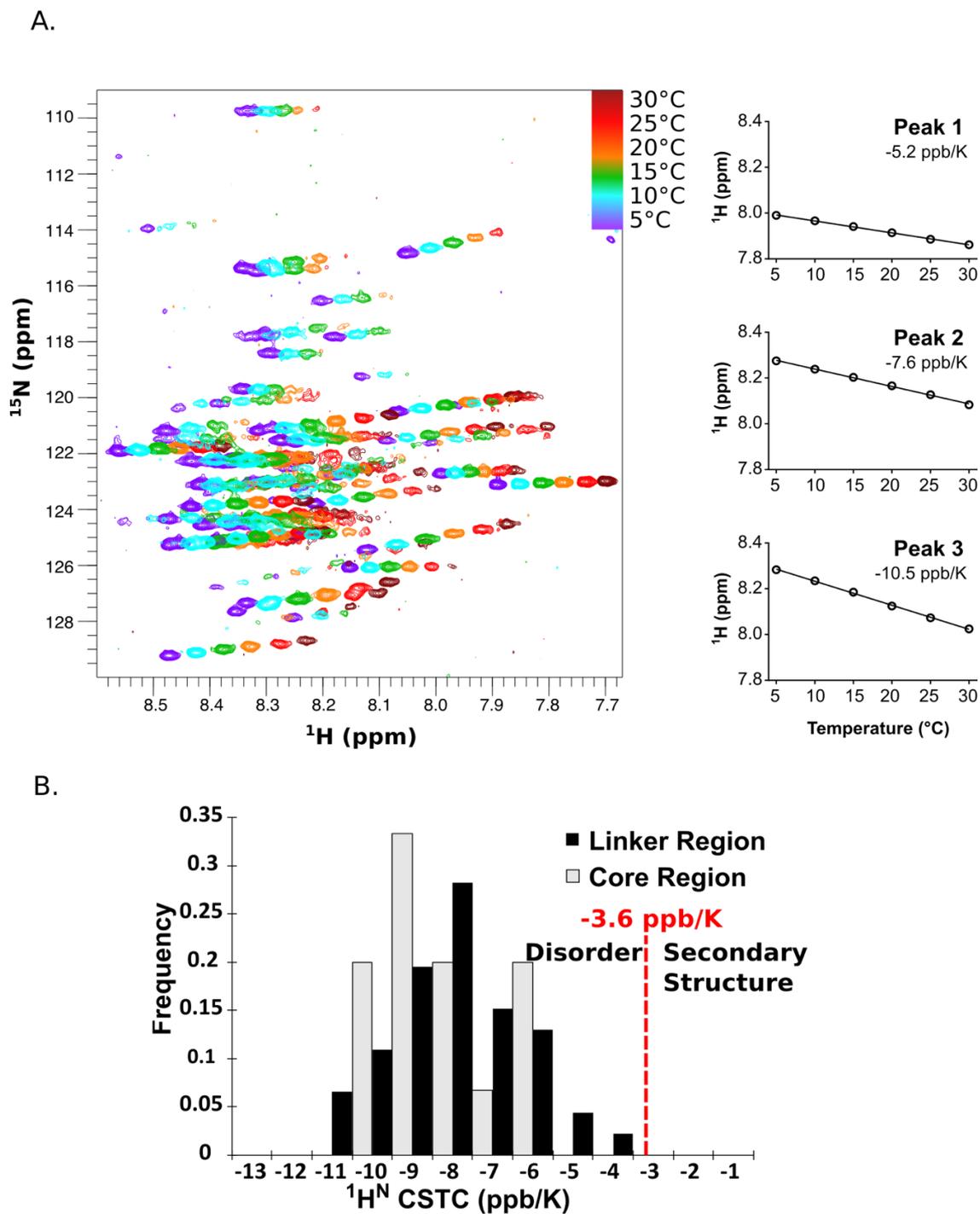


Figure 5.5

Chemical shift temperature coefficient (CSTC) experiments of the luminal domain suggest all peaks analysed are from disordered regions of the protein. A. HN TROSY experiment of the wild-type luminal domain at different temperatures (Section 2.7.1.4). The peaks shift across the x-axis (^1H domain) as temperature is increased. Shown are the plots for three selected peaks where the change in ^1H position is plotted against temperature; the linear relationship is calculated to give a value of ppb/K which is used to determine whether the residue is from a region of disorder. B.

The chemical shift temperature coefficient values for each peak are plotted for the core domain and linker region peaks. All peaks occur within the threshold of -3.6ppb/K , suggesting that the peaks are derived from regions of disorder (Okazaki et al., 2018). A full list of values is shown in appendices Figure 9.6 and Table 9.1.

Table 5.1

The coverage of different segments of the luminal domain's linker region and core region peaks by the $^1\text{H}^{\text{N}}$ chemical shift temperature coefficient analysis with the average values for each segment and standard deviations.

Region (residues)	Total peaks	Peaks analysed	% coverage	Average $^1\text{H}^{\text{N}}$ CSTC (ppb/K)	
Loop region 2	24	13	54	-7.83	± 1.57
390 - 405	12	10	83	-7.40	± 1.98
406 - 410	5	5	100	-8.47	± 1.67
411 - 415	5	5	100	-7.07	± 1.55
416 - 420	3	3	100	-8.89	± 1.66
421 - 425	8	7	88	-8.23	± 1.27
426 - 449	20	16	80	-8.27	± 1.43

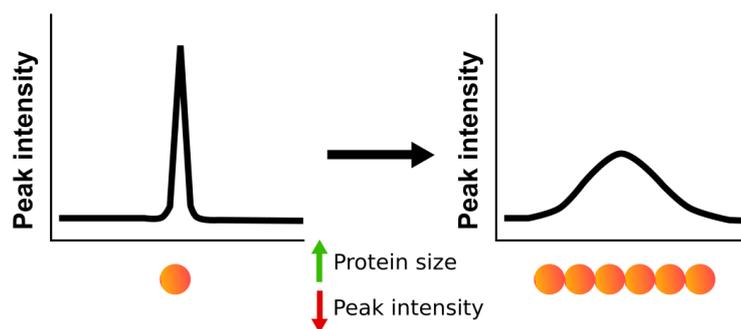
5.5. Different segments of the linker region experience differential dynamic processes

The temperature coefficient analysis of the luminal domain's spectrum was unable to cover all the peaks identified due to the method of experimentation used, the peaks that were not analysed were those with a lower intensity. NMR peak intensity can report on the dynamic processes experienced by a residue, described briefly here are three molecular factors that can influence a peak's observed intensity (Figure 5.6).

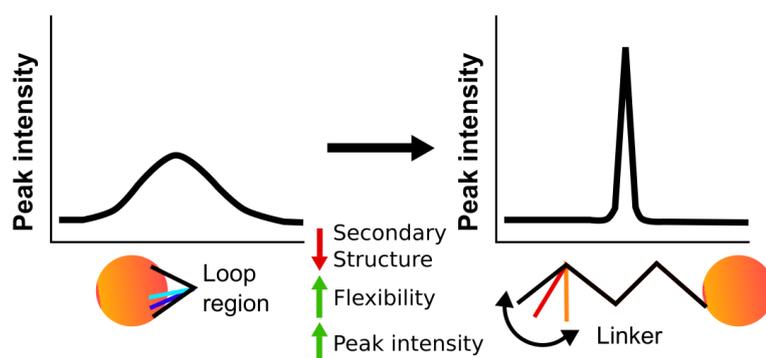
Proteins tumble in solution, with the time taken to rotate dependent on the size of the protein and termed the rotational correlation time (τ_c). Larger species tumble more slowly in solution than smaller species, and therefore have larger τ_c values which cause increased relaxation in the NMR active nuclei; this translates as line broadening leading to a decreased signal intensity in the detected spectrum (Figure 5.6A) (Lee, D. et al., 2006). Therefore, larger molecules (such as the luminal domain oligomers) are expected to have broader peaks and therefore lower signal intensities from the global effect of protein tumbling in solution.

Where the tumbling of the protein can affect peak intensity globally, local changes can influence peak line widths and thus intensity also. Flexible regions of the protein experience increased ns dynamics which cause a sharper peaks and an increase in peak intensity (Figure 5.6B) (Konrat, 2014). In the case of the luminal domain, the linker region's flexibility effectively counteracts the effects of protein tumbling, therefore causing increased peak intensities. It is therefore expected that more C-terminal and distal segments of the linker region from the globular core domain will exhibit increased peak intensities. Therefore, changes in the size of small multimers are not expected to drastically affect the signal intensities of flexible linker region residues, only more rigid regions such as loop region 2.

A. τ_c : Global rotational correlation time



B. Local ns dynamics of flexible regions



C. μ s-ms dynamics (local or global)

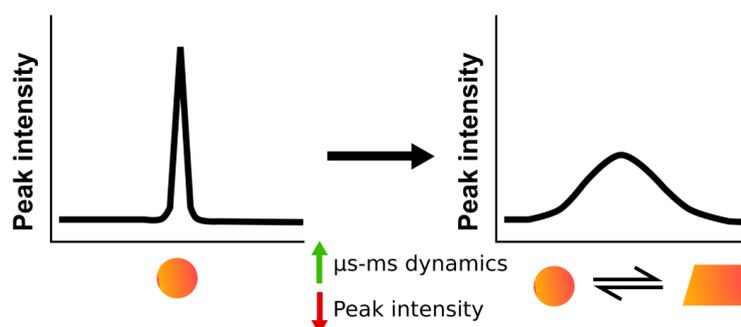


Figure 5.6

Representation of the effects of molecular factors on line width of recorded NMR peaks and therefore their effects on peak intensity. A. An increased size of the protein causes increased rotational correlation times (τ_c) causing broadening of peaks through relaxation and therefore

decreased peak intensities. B. Increased flexibility means increased local ns timescale dynamics which increase the peak intensity. C. Increased μ s-ms dynamics from domain rearrangements can be local or global protein effects, these can cause decreased peak intensities of the residues experiencing the effect.

Another contribution to peak line width and therefore intensity comes from μ s-ms dynamic processes, these can consist of global or more local rearrangements in protein conformation (Figure 5.6C). The contribution of global μ s-ms dynamics of the luminal domain between multiple states has previously been suggested as the causative factor for the lack of peaks observed in the protein's NMR spectrum (Karagoz et al., 2017). Therefore, increased μ s-ms dynamics and transient adoption of an NMR invisible state of the residue will cause a drop in observed peak intensity due to line broadening. It's worthy to note that this is a simplified summary of these processes, which have been extensively reviewed previously (Kleckner and Foster, 2011; Mittermaier and Kay, 2009; Lee, D. et al., 2006).

The peak intensities for different segments of the linker region in the *wild-type* protein spectrum were therefore compared to understand differential dynamic processes occurring (Figure 5.7). In Figure 5.7A it is clear that the different segments of the linker region analysed have different peak intensity profiles, which is further visualised in Figure 5.7B. As expected, there is a trend for increased average peak intensities for segments of residues that are more distal from the luminal domain's core domain, likely owing to increased flexibility and local ns dynamics (Konrat, 2014). An overall trend of the data presented is that regions with secondary structure predicted have more heterogeneity in their residues' peak intensities, this suggests that some peaks in the region have increased ns local dynamics and others experience increased μ s-ms dynamics or adopt secondary structure, in agreement with the structural prediction.

Therefore, the ^1H dispersion and temperature coefficient data suggests that the luminal domain's NMR spectrum consists of peaks from residues in disordered regions of the protein. However, analysis of the peak intensities for different segments of the protein shows that regions of predicted secondary structure have increased heterogeneity in their peak intensities to imply that these regions may transiently adopt the predicted secondary structural elements (Shen, Y. et al., 2014; Thevenet et al., 2012; Drozdetskiy et al., 2015) or at least have differential dynamic processes experienced in different areas of the linker region. This may represent functionally important interactions and conformations adopted by the linker region that may affect the protein's conformational landscape, therefore, using this information and the assignments of the luminal domain, the role of these disordered regions will be investigated.

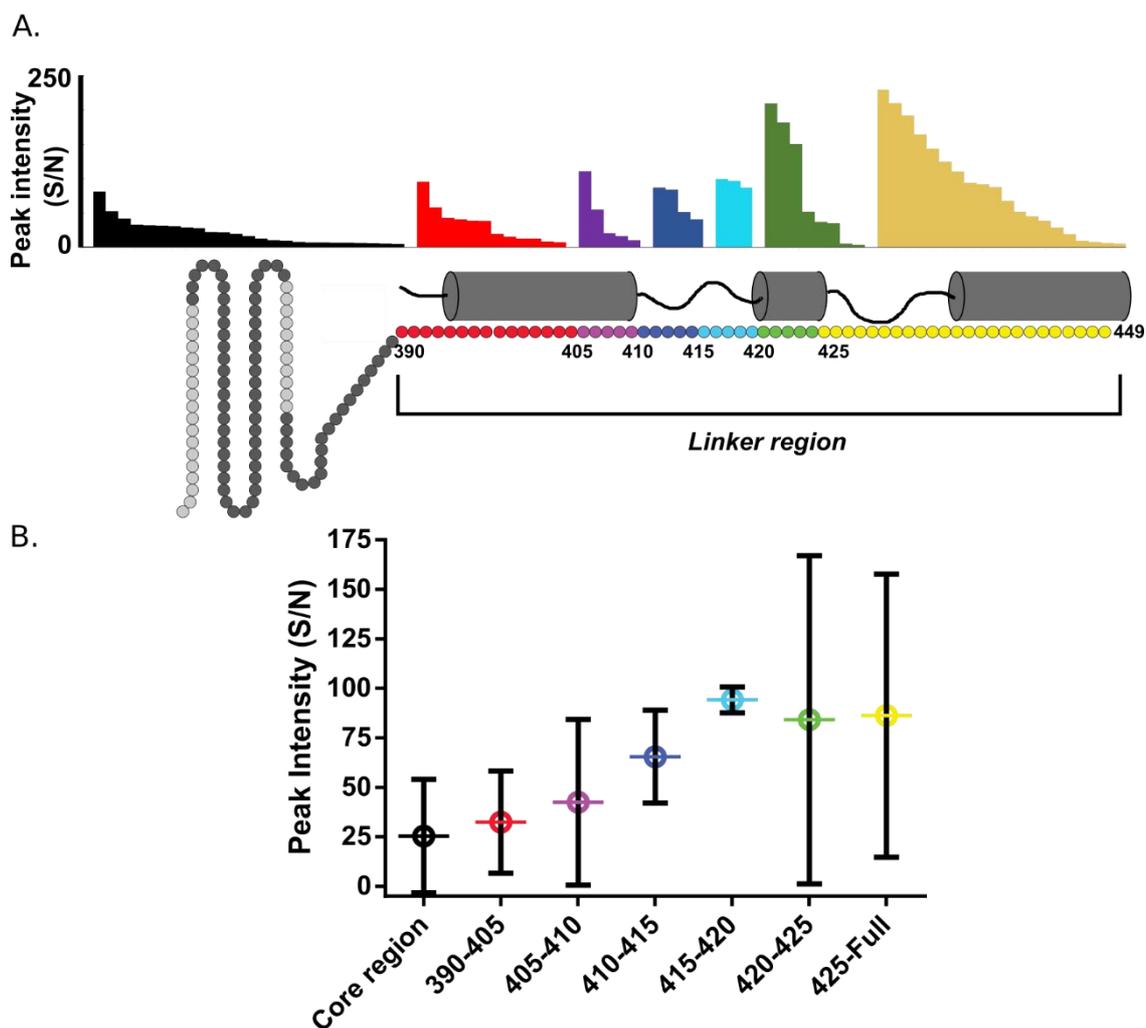


Figure 5.7

Peak intensities for different segments of the luminal domain's linker region and core domain. A. Plotted are the peak intensities (signal to noise (S/N)) for peaks of each segment of the linker region and core domain from a HN TROSY experiment (Section 2.7.1.2). Shown is a cartoon representation of the linker region and of the regions with high helix formation propensity. Within each segment the peak intensities are ordered by intensity. B. A plot of the average signal intensity for each segment of the linker region and core domain with standard deviation bars shown. More distal segments from the core domain have, on average, increasing peak intensities. Those regions that have secondary structure predicted have more heterogeneity in their peak intensities, suggesting differential dynamics.

5.6. Characterising different IRE1 α conformational states by NMR

In Chapter 4 the luminal domain's activation was functionally characterised and conditions to trap individual conformations such as monomers, dimers and oligomers were identified. Solution NMR has relatively non-stringent sample conditions and therefore the luminal domain can be

manipulated and analysed in the desired conformations as it was in Chapter 4. With the partial assignments of the spectrum, solution NMR study will report on the various segments of the linker region and the luminal domain's loop region 2's chemical environment and dynamic processes in the isolated conformations.

To achieve this, analysis of chemical shift perturbations, for information about different chemical environments (Section 1.6.1.3) alongside the previously mentioned analysis of peak intensities (Figure 5.6), for information about the dynamic processes occurring can be used. Chemical shift perturbations are changes in the position of a peak in an NMR spectrum. Such a change represents a change in the chemical environment of that residue, and therefore they are indicative of process such as an internal conformational change or ligand binding (Williamson, 2013). In some cases, a residue's peak disappears from the spectrum, due to a shift in position which makes it impossible to identify the corresponding peak in the reference spectrum, this can be caused by large conformational changes.

5.6.1. Monomeric protein has a differential linker region conformation to dimeric/oligomeric protein

To trap the monomeric state of the luminal domain, the dimerisation-deficient mutant (D123P) (Zhou et al., 2006) was used, this mutation perturbs the dimerisation interface and therefore the protein exists as monomers, as observed by size exclusion chromatography (Figure 5.8A and B). The D123P mutation site is distal to the intrinsically disordered regions (Figure 5.9) that are observed by solution NMR and therefore no chemical shift perturbations are expected to be due to local effects on residues' chemical environments from the mutation.

The D123P mutant's HN TROSY spectrum is compared to *wild-type* protein in Figure 5.8C. As observed in Chapter 4, the *wild-type* protein is expected to be oligomeric at the concentration used for NMR study, therefore this analysis can be considered as a comparison of the monomeric to oligomeric conformations. Unexpectedly, there are multiple chemical shift perturbations throughout the intrinsically disordered regions, some are identified through the peak shifting its position in the spectrum and some from the peak disappearing from the spectrum (Figures 5.8C and D). Due to the location of the D123P mutation, all the perturbations are long-range effects, suggesting that oligomerisation affects the linker region's conformation. The majority of the chemical shift perturbations observed are from residues 390-410 along with two in the 411-415 region and one significant chemical shift perturbation in the most C-terminal segment (residues 426-449). This suggests that the most N-terminal section of the linker region is involved in

dimerisation/oligomerisation of the protein, through a possible allosteric mechanism. The effects of the mutation are summarised in Figure 5.9.

Next, the effect of oligomerisation on peak intensity of the more rigid loop region 2 residues was assessed (Figure 5.8E). As described, these residues are expected to be affected by increased τ_c values (Lee, D. et al., 2006) as shown in Figure 5.6A, and therefore it is expected that the smaller D123P multimers will exhibit increased peak intensities than the oligomeric *wild-type* protein. As expected, the loop region 2 peaks analysed display increases in peak intensity in the D123P construct (1.9-fold increase), suggesting smaller multimeric states, whereas the peak intensities for the dynamic C-terminal linker residues are not significantly affected (Figure 5.8D).

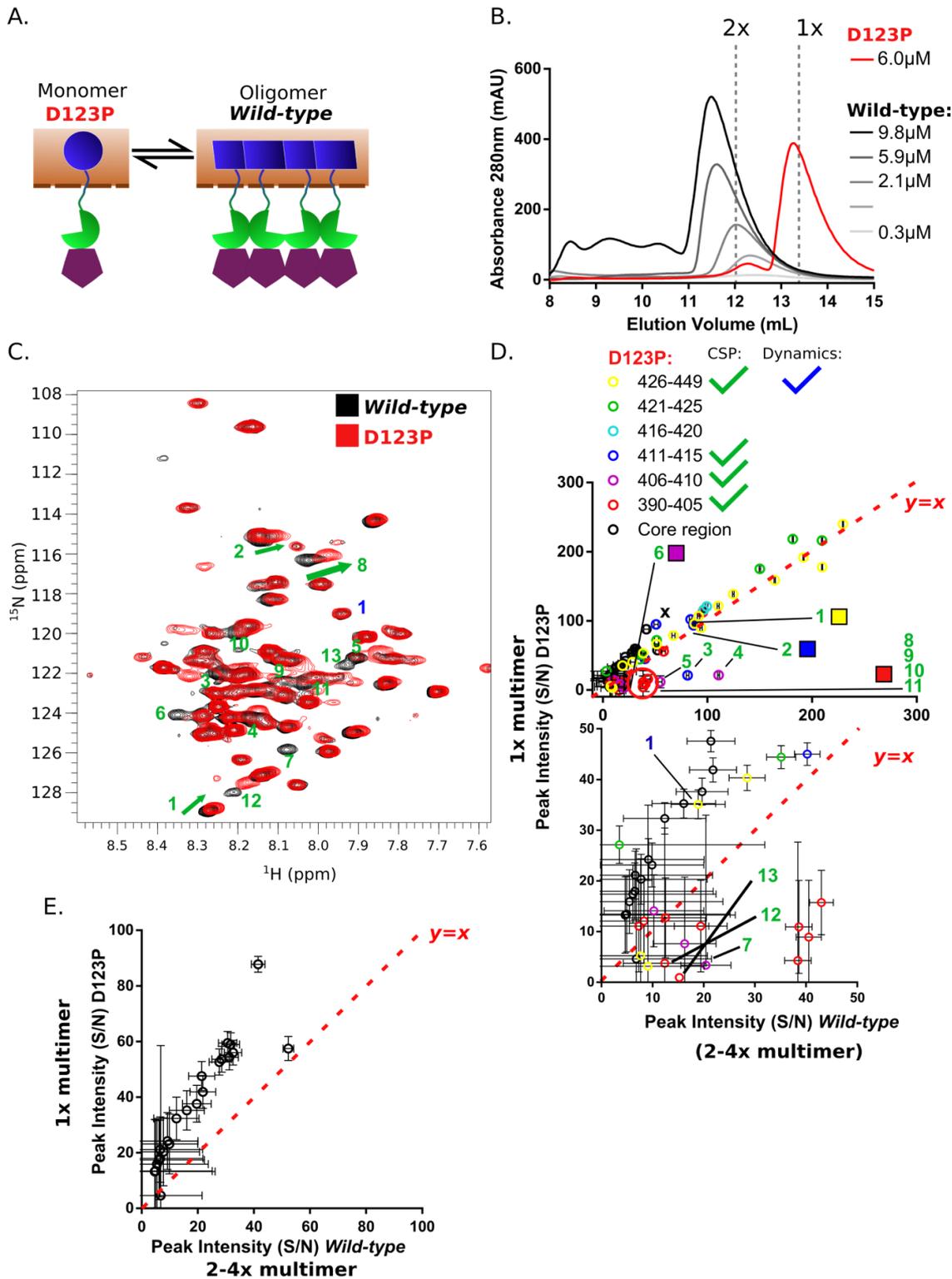


Figure 5.8

The monomeric, dimerisation-deficient D123P mutant analysed by NMR presents many significant changes from oligomeric wild-type protein. A. Model for the differences observed in the spectra shown, the D123P mutant promotes monomeric species of the protein, whereas the wild-type protein is considered to exist as oligomers at the concentration used for NMR analysis (50 μM; Section 2.7.1.2). B. Size exclusion chromatography (Section 2.6.2.1) suggests that the

D123P mutant will behave as a monomer at the concentration used for NMR and the wild-type protein is likely to exist as oligomeric species. C. The overlaid spectra of the *D123P* mutant and wild-type protein with significant chemical shift perturbations shown, the labels correspond to panel D which also details which part of the linker region the residue is from. D. The peak intensities of the spectra of the *D123P* and wild-type luminal domain are compared. Numerous peaks have much reduced intensities in the *D123P* mutant where they appear to disappear on the spectrum, this is due to chemical shift perturbations. Peaks are labelled corresponding to their position in the spectra in panel C. E. The core domain peak intensities for the *D123P* mutant are plotted against wild-type protein with a line for $y=x$ shown. The peaks of the core domain have significantly different intensities in the *D123P* monomeric protein compared to the oligomeric wild-type protein, as analysed by an unpaired *t*-test ($p = 0.0013$).

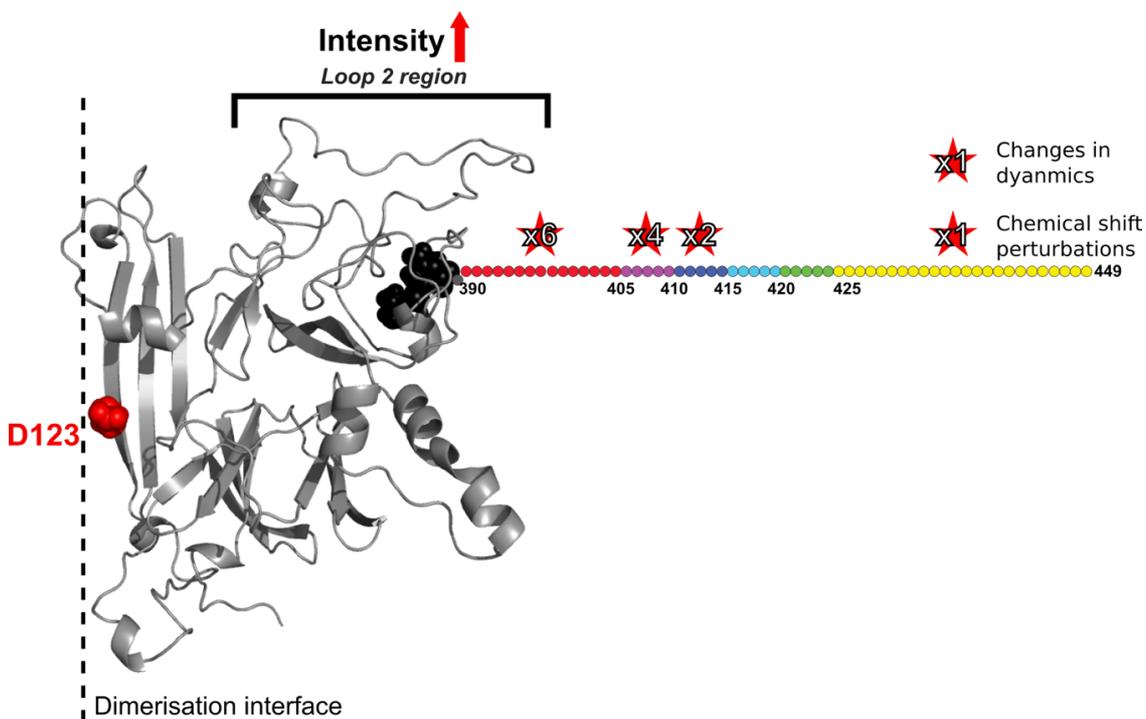


Figure 5.9

The *D123P* mutation shifts equilibrium towards monomeric protein. Effects are observed by NMR in the intrinsically disordered regions of the protein. Shown is the crystal structure model (PDB:2HZ6) (Zhou et al., 2006) with predicted disordered regions (Yang et al., 2015) and a cartoon representation of the linker region. Upon mutation of the *D123* residue at the dimerisation interface to a proline residue, significant changes are observed in NMR peaks of the linker region, suggesting they undergo allosteric changes upon dimerisation/oligomerisation of the protein. Increased peak intensities of residues of loop region 2 suggest a smaller protomer (Lee, D. et al., 2006). Figure prepared using PyMOL version 1.7.

5.6.2. The dimer to oligomer transition involves changes in the protein's disordered regions

To investigate the role of the luminal domain's intrinsically disordered regions in the protein's transition from a dimeric to an oligomeric conformation, the WLLI-GSSG mutant was used (Figure 5.10A). The quadruple mutation has been previously characterised to disfavour luminal domain oligomerisation in the absence of unfolded protein (Karagoz et al., 2017) and in Section 4.3.2.3 the mutation disfavoured the formation of oligomers. Therefore, in the conditions used for NMR study of the protein the WLLI-GSSG mutant is expected to exist as dimers.

Comparison of the WLLI-GSSG mutant spectrum to the *wild-type* spectrum firstly presents three chemical shift perturbations (labelled in green; Figure 5.10B and C). Interestingly, two are from the most distal segment of the linker region (residues 426-449), similar to the D123P mutant this suggests an allosteric interaction of the C-terminal of the linker region involved in conformational changes of the luminal domain. The third is from a low intensity core domain peak (Figures 5.10B and D), this chemical shift perturbation suggests that residues of loop region 2 are affected by changes to the oligomerisation interface residues (Karagoz et al., 2017) and therefore it's possible that the region may have a role in its regulation.

The WLLI-GSSG mutant is expected to exist as smaller multimers than the *wild-type* luminal domain and therefore experience reduced τ_c values and increased core domain peak intensities (Figure 5.6A). As observed in Figure 5.10D and E the WLLI-GSSG loop region 2 peaks have significantly increased peak intensities (1.8-fold), suggesting a smaller luminal domain multimer than in *wild-type* protein, as expected. Two additional significant changes in peak intensity are observed, signifying differential dynamics in residues of the 426-449 segment of the linker region (labelled in red, Figure 5.10B and C), further suggesting long-range communication of this segment with the core domain. Also changes to peak intensities are observed in the 411-415 and 421-425 segments suggesting these regions may also be involved in a potential allosteric network.

Therefore, by study of the WLLI-GSSG mutant an involvement of loop region 2 in oligomerisation has been suggested and more evidence has been provided for the role of the most distal segment of the linker region (residues 426-449) in luminal domain conformation, implying allosteric communication which may affect function (Figure 5.11).

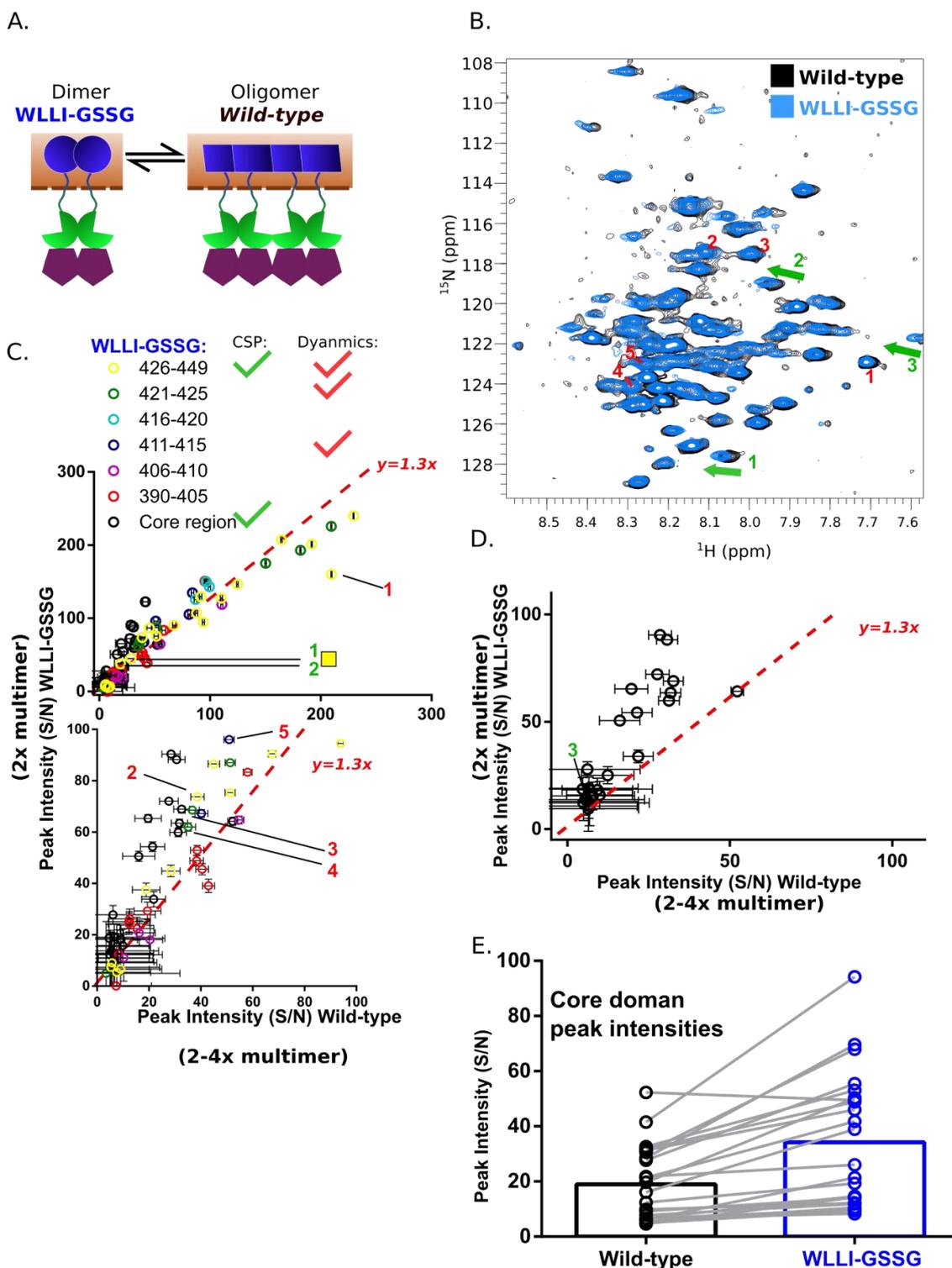


Figure 5.10

The oligomerisation deficient mutant (357-361 WLLI-GSSG) causes changes to residues of loop region 2 and the 426-449 segment of the linker region. A. Comparison of the WLLI-GSSG mutant to wild-type luminal domain allows comparison of the dimeric and oligomeric conformations respectively. B. The HN TROSY (Section 2.7.1.2) spectrum of the WLLI-GSSG mutant overlaid with the wild-type spectrum. Significant chemical shift perturbations occur for three peaks; one in loop region 2 and two of the 426-449 area of the linker region. C. The peak intensities of

different areas of the luminal domain are compared between the wild-type and WLLI-GSSG mutant construct. Line for $y=1.3x$ shown, this difference is thought to be due to different sample concentrations used, subsequent analysis of intensities corrects for this. D. The core domain peaks (from loop region 2) have significantly increased intensities in the dimer than in the oligomeric (wild-type) protein (using an unpaired t -test, $p = 0.001$). E. Comparison of the loop region 2 (core domain) peak intensities of wild-type and WLLI-GSSG constructs. There is a clear trend for increased peak intensities of the loop region 2 residues in the dimeric protein as opposed to oligomeric.

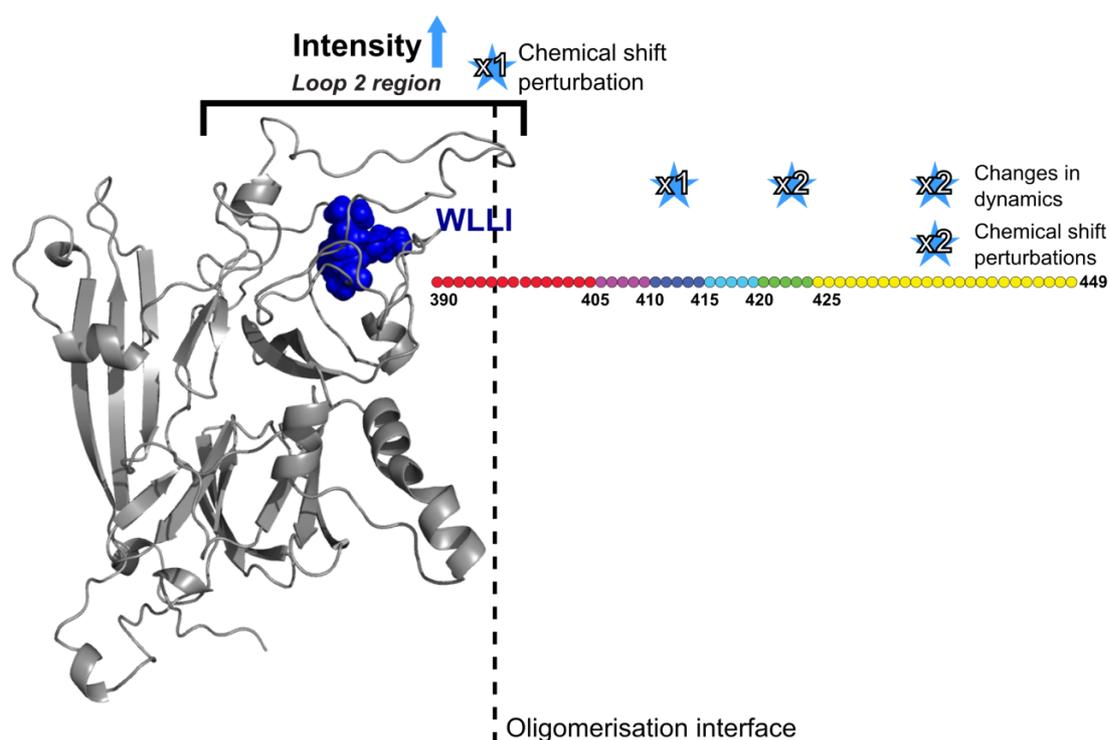


Figure 5.11

Summary of changes observed in solution NMR study of the WLLI-GSSG mutant. The X-ray crystal structure model (PDB: 2HZ6) (Zhou et al., 2006; Yang et al., 2015) shows the WLLI residues of the identified oligomerisation interface in blue (Karagoz et al., 2017). Through NMR study of the disordered regions it is apparent that the WLLI-GSSG mutation increases the peak intensities of residues in loop region 2 that surround the WLLI oligomerisation interface and causes one chemical shift perturbation in loop region 2 and changes in the most distal segment of the linker region (residues 426-449). Figure prepared using PyMOL version 1.7.

5.6.3. There are no significant changes to luminal domain disordered regions upon peptide binding

In order to determine whether peptide binding has an effect on the disordered regions of the protein, *wild-type* luminal domain was incubated with Δ EspP for three hours before being analysed by a HN TROSY experiment and compared to a sample without peptide added (Figure 5.12). The sample without peptide added was at a concentration of $36\mu\text{M}$ and therefore suggested to exist as small oligomers. The sample with $36\mu\text{M}$ peptide added is expected to be bound by peptide but remain soluble, therefore also existing as small oligomers and so any changes observed are expected to be from peptide binding.

It's apparent that there are no observable chemical shift perturbations or changes in peak intensity upon incubation with $36\mu\text{M}$ Δ EspP (Figure 5.12B and C), suggesting that peptide binding to oligomeric luminal domain protein does not prompt any conformational changes to the protein's intrinsically disordered regions.

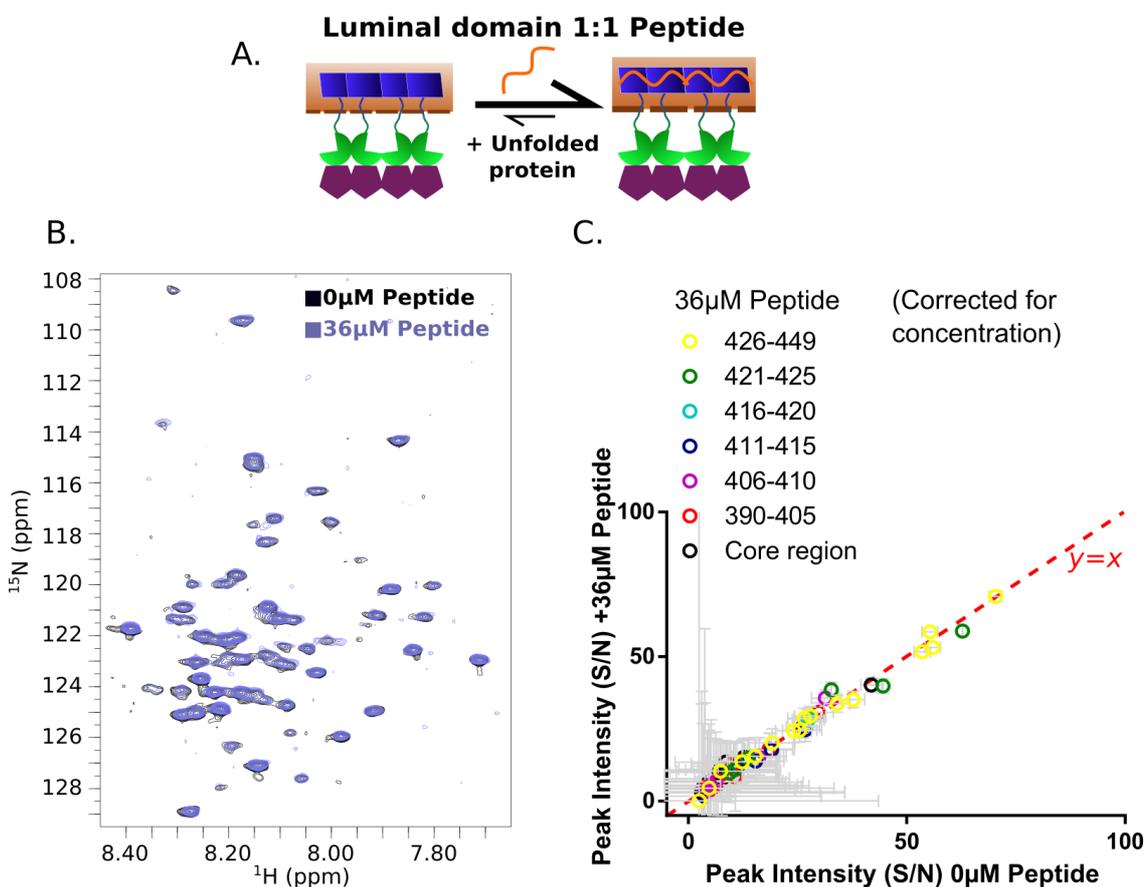


Figure 5.12

Binding of Δ EspP to IRE1 α 's luminal domain doesn't affect the chemical environment of the protein's disordered regions, determined by HN TROSY experiments. A. Cartoon representation of the transition being observed by data in panels B and C. Δ EspP binds to the oligomerised luminal domain. B. Upon incubation of $36.3\mu\text{M}$ Δ EspP with $36\mu\text{M}$ IRE1 α luminal domain for three hours the sample was centrifuged, and the concentration of soluble protein measured as

29.9 μ M, this sample was then analysed using NMR (Section 2.7.1.5). There are no significant chemical shift perturbations in the overlaid spectra. C. The peak intensities of the intrinsically disordered regions, with and without peptide are not significantly different when the change in concentration is accounted for, suggesting no changes in dynamic processes of the luminal domain's disordered regions upon peptide binding.

The inclusion of an equimolar ratio of Δ EspP with the oligomerised *wild-type* protein appeared to cause no changes to the conformation of the linker region. This may be because the protein is already oligomerised, and therefore the protein has already adopted an oligomerisation-active conformation which is unaffected by an increase in oligomeric size. The luminal domain is thought to adopt this conformation transiently in the absence of stress and peptide binding is thought to promote its formation (Karagoz et al., 2017). It was therefore hypothesised that binding of peptide to the oligomerisation disfavoured mutant (WLLI-GSSG) will promote the oligomerisation-active conformation, and therefore may give a spectral pattern resembling *wild-type* protein. The WLLI-GSSG mutant was incubated with an equimolar ratio of MPZ1 peptide, in similar conditions as used in Karagoz et al. (2017), where 50 μ M MPZ1 peptide and 50 μ M luminal domain were used to promote the formation of the oligomer-active state.

The WLLI-GSSG mutant incubated with an equimolar amount of each peptide is shown in Figure 5.13. MPZ1 peptide incubation with the WLLI-GSSG mutant does not appear to significantly perturb the chemical shift or intensities of any observed residues (Figure 5.13B and C). An equimolar ratio of protein and MPZ1 peptide has been previously characterised not to cause an increase in the luminal domain's multimeric state (albeit at 5 μ M) (Karagoz et al., 2017), and doesn't induce an increase in the measured OD400 value (Section 4.3.2.2). Therefore, as no chemical shift perturbations are observed, this indicates that the previously observed changes from dimeric to oligomeric protein (Figure 5.11) are caused by oligomerisation rather than transition to an oligomerisation active conformation of the protein. Additionally, no significant change in loop region 2 peak intensities are observed, this is expected as there is not expected to be a change in multimer size and therefore changes from differential rotational correlation times.

To ensure the formation of the oligomer active state, the WLLI-GSSG protein was also incubated with an equimolar amount of Δ EspP (Figure 5.13D). In Section 4.3.2.1 the Δ EspP has an increased affinity to promote oligomerisation, likely due to its previously characterised increased affinity to bind to the luminal domain compared to the MPZ1 peptide (~6 μ M for Δ EspP, ~24 μ M for MPZ1) (Carrara et al., 2015; Karagoz et al., 2017). Similar to study with the MPZ1 peptide, there are no observable chemical shift perturbations upon incubation with Δ EspP (Figure 5.13E).

A technical fault during acquisition caused the observed peak broadening, preventing accurate analysis of changes in peak intensities.

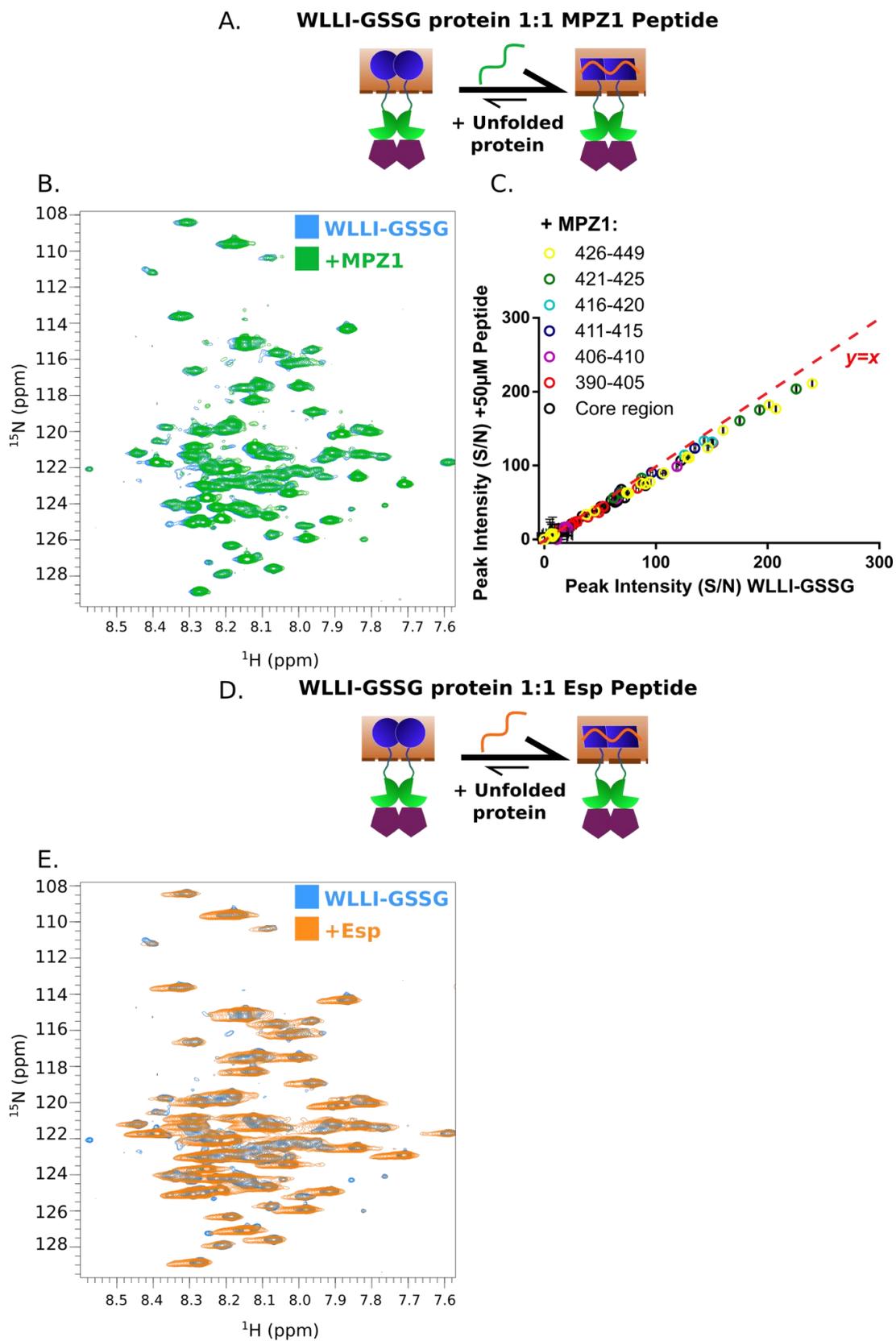


Figure 5.13

The oligomerisation impaired WLLI-GSSG mutant does not experience chemical shift perturbations in its disordered regions upon peptide binding, as observed by NMR HN TROSY experiments. A. Cartoon representation of the transition observed in panels B and C. MPZI peptide binds to the WLLI-GSSG dimeric mutant. B. Upon incubation for 180 minutes with 1:1 (50 μ M) MPZI peptide there are no significant chemical shift perturbations occurring, indicating that peptide binding to the dimeric protein (to promote the oligomerisation active conformation (Karagoz et al., 2017)) does not induce changes in the protein's intrinsically disordered regions as observed by solution NMR (Section 2.7.1.5). C. Binding of MPZI peptide to the WLLI-GSSG mutant doesn't elicit significant changes in peak intensities. D. Cartoon representation of the Δ EspP binding to the WLLI-GSSG dimeric protein. E. Upon incubation for 180 minutes with 1:1 Δ EspP (50 μ M) there does not appear to be significant chemical shift perturbations, however, there was an error whilst running the experiment (leading to 1 H broadening).

The data presented here therefore suggests that unfolded protein binding to the luminal domain only influences structured regions that are not observed by the solution NMR experimentation used here, in agreement with a previous study (Karagoz et al., 2017). The changes in loop region 2 and the most distal section of the linker region observed in Figure 5.11 therefore appear to be a result of oligomerisation of the luminal domain, rather than adoption of the oligomer-active state. Additionally, different peptide sequences binding to the protein do not prompt different conformations of the luminal domain's intrinsically disordered regions.

5.7. Conclusion

In this chapter the luminal domain has been analysed by solution NMR spectroscopy. Previously the spectrum was thought to consist of only intrinsically disordered regions of the protein due to the 1 H dispersion of observed peaks (Okazaki et al., 2018). Here temperature coefficient experiments also suggest that only disordered regions are observed in the luminal domain's NMR spectrum. The majority of residues observed are from the protein's linker region and these were assigned to different segments of the linker region sequence using truncated luminal domain constructs. Although the aforementioned analysis suggests that the linker region is disordered, analysis of the different segments' dynamic processes suggest that some sections of the linker region are involved in differential dynamic processes and may therefore interact with other regions of the protein or adopt secondary structure transiently (Kleckner and Foster, 2011). Interestingly, the regions that present heterogeneity in their dynamic processes are also predicted to adopt secondary structural elements (Shen, Y. et al., 2014; Thevenet et al., 2012; Drozdetskiy et al., 2015).

Results in Chapter 4 offer methods to manipulate IRE1 α 's conformational landscape through mutation (Section 4.3.2.3). Monomeric and dimeric states of the luminal domain were compared to oligomeric protein and for both conformations global effects from the change in protein size were observed in the luminal domain's loop region 2 (Lee, D. et al., 2006). Additionally, differential long-range changes were observed in the protein's disordered regions for both conformations (Table 5.2).

Table 5.2

The transitions of states of IRE1 α 's luminal domain studied in this chapter by NMR spectroscopy. Detailed are the chemical shift perturbations and changes in residue dynamic processes observed in the intrinsically disordered regions of the protein in each transition between conformations.

Transition		Chemical shift perturbations	Dynamics (& oligomerisation)
From:	To:		
<i>Monomer</i>	\rightarrow <i>Oligomer</i>	Region 390-415 (12) Region 426-449 (1)	Loop region 2 (global) Region 426-449 (1)
<i>Dimer</i>	\rightarrow <i>Oligomer</i>	Loop region 2 (1) Region 426-449 (2)	Loop region 2 (global) Region 411-415 (1) Region 421-449 (4)
<i>Dimer</i>	\rightarrow <i>Peptide bound dimer</i>	-	-
<i>Oligomer</i>	\rightarrow <i>Peptide bound oligomer</i>	-	-

The monomer to dimer/oligomer transition causes many chemical shift perturbations in residues between position 390 and 415 of the protein suggesting long-range communication of this area of the linker region and potentially an allosteric mechanism of regulation of dimerisation by the linker region.

Interestingly, removal of the luminal domain's linker region (at residue 389) was reported to reduce BiP binding and therefore the expected BiP-dependent monomerisation of the luminal domain, but reinstatement of residues 390-408 restored normal BiP repression (Oikawa et al., 2009). However, BiP has been shown to bind not to the linker region, but to the core domain (residues 24-389) of the luminal domain, promoted by interaction with ERdj4 (Amin-Wetzel et al., 2017). It's therefore possible that the linker region conformation observed here has a role for fine tuning these interactions.

Upon comparison of the trapped dimeric and oligomeric states of the luminal domain one chemical shift perturbation was observed in loop region 2, suggesting that the region is affected

by the identified oligomerisation interface residues (Karagoz et al., 2017), this may be through an adjoining monomer in oligomerisation or a long-range interaction with the linker region. This is in disagreement with previous NMR study of the luminal domain that suggested no effect was observed in a residue of loop region 2 upon oligomerisation (Karagoz et al., 2017), however, the probe used in this study may have occurred in an area of loop region 2 distal to the affected site observed here.

Unexpectedly, changes in the most C-terminal segment of the linker region were observed in the dimeric to oligomeric transition (Figure 5.11) as well as a chemical shift perturbation in the monomeric protein (Figure 5.9). Interestingly, an amphipathic helix in this segment of the linker region (as predicted in Figure 5.2) has been suggested to be involved in sensing endoplasmic reticulum lipid bilayer stress to promote the unfolded protein response (Halbleib et al., 2017; Kono et al., 2017). Additionally, residues of this region have been implicated as the binding site for the repressing interaction of the Sec61 translocon with the luminal domain (Sundaram et al., 2017). The differential long-range effects of this segment of the linker region in monomeric, dimeric and oligomeric protein may therefore represent an allosteric regulation mechanism for the fine-tuning of the protein's conformational landscape by these interactions.

Binding of peptide to the dimeric and oligomeric states of the luminal domain do not induce changes to the protein's disordered regions, in agreement with previous NMR study of the domain, unfolded proteins are suggested to interact with and alter the conformation of structured regions of the protein (Karagoz et al., 2017; Credle et al., 2005) and are therefore not expected to be observed by the analysis in this chapter. The information gained through solution NMR study of the linker region is summarised in a schematic in Figure 5.14A.

Using the data presented here and previous study of the protein, two models are hypothesised for the role of the N-terminal portion of the linker region in regulation of the luminal domain. Model A (Figure 5.14B) suggests that that BiP binding promotes the monomeric conformation of the linker region observed by solution NMR. Here, the adopted conformation opposes dimerisation, favours BiP/ERdj4 binding and is the most favourable linker conformation when the protein is monomeric (and is therefore observed by solution NMR study). Previous literature has suggested that BiP repression occurs to a lesser extent with the removal of these residues (Oikawa et al., 2009) supporting the hypothesised model.

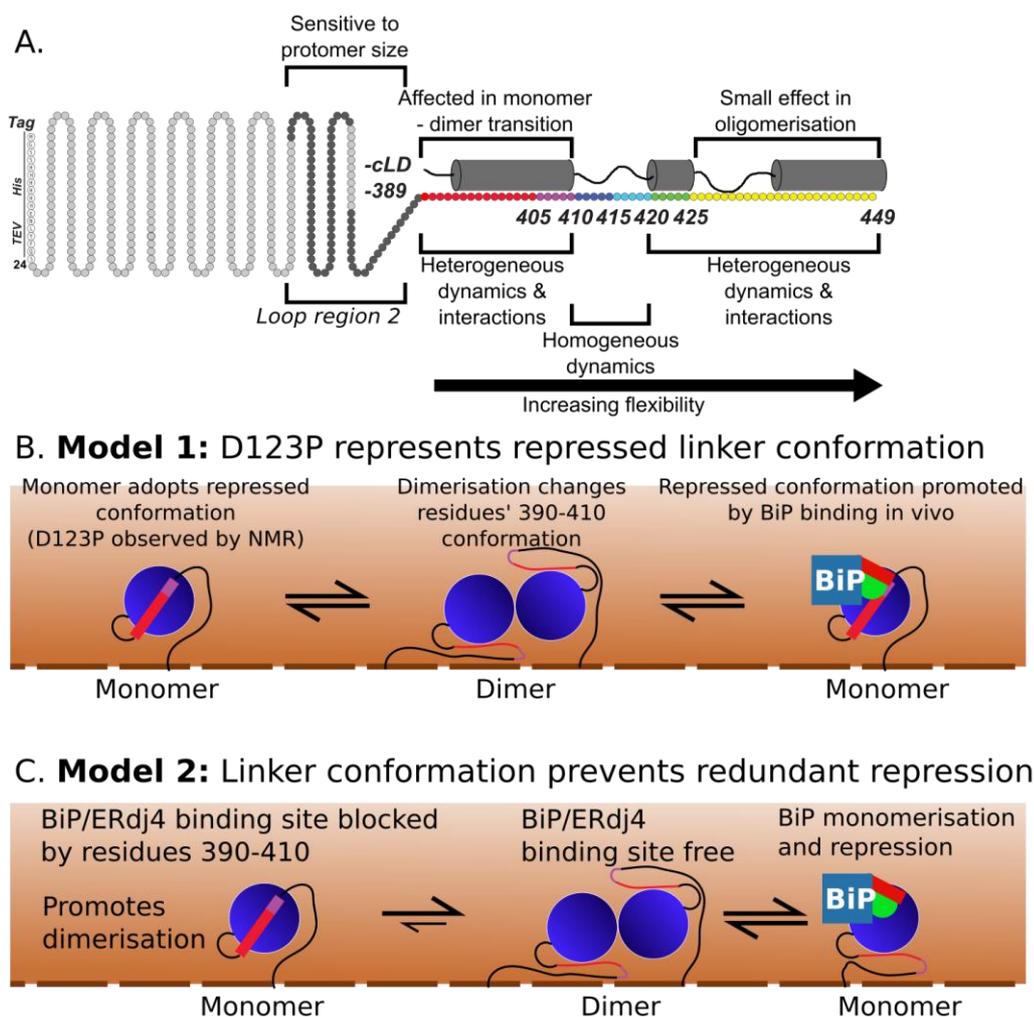


Figure 5.14

Summary of the findings and hypothesised models from solution NMR study of the luminal domain's conformations. A. Schematic of the luminal domain's protein sequence. Shown are the disordered regions and the interactions that they undergo in specific conformations of the protein. B. First hypothesised model for the role of the N-terminal section of the linker region in luminal domain function. The monomeric linker region conformation is favoured in the absence of dimerisation and is promoted upon BiP binding to prevent dimerisation. C. Hypothesised model 2, the monomeric conformation of the linker region promotes dimerisation of the protein and opposes ERdj4/BiP binding. Upon dimerisation the linker region conformation changes and ERdj4/BiP binding is favoured.

The second model hypothesises that the N-terminal region of the linker region in the monomeric conformation disfavors ERdj4/BiP binding (Figure 5.14C), therefore opposing BiP repression (Amin-Wetzel et al., 2017). Dimeric luminal domain is considered to be able to elicit the unfolded protein response (Amin-Wetzel et al., 2017), and therefore in times of low endoplasmic reticulum stress, BiP interacts with dimeric protein to monomerise it. Additionally, in Chapter 4 the

dimerisation affinity for the luminal domain with the linker region was slightly weaker than values reported in the literature for the core luminal domain, albeit by different methodologies (Karagoz et al., 2017). Therefore, in model 2 the monomeric luminal domain conformation prevents redundant repressive interactions and promotes dimerisation, but upon dimerisation the repressive interactions with ERdj4/BiP (Amin-Wetzel et al., 2017) are favourable.

Further investigation is required into the role of this N-terminal segment of the linker region for the roles of its differential conformations. Additionally, the differential long-range effects observed in the C-terminal residues of the linker region require validation for their potential allosteric mechanism.

Therefore, through solution NMR study of the luminal domain, understanding has been gained about the interactions and possible regulatory role of the protein's previously uncharacterised disordered regions in different conformations of IRE1 α 's conformational landscape. Experiments here have also allowed for a deeper understanding of the solution NMR data gained to be potentially used for characterisation of novel conformations identified, interactions of the protein (including potential targeted therapeutic binding) and also the effect of cancer-associated mutations on the disordered regions, which will be examined in the following chapters.

6. Cancer-associated mutants affect the conformational landscape of IRE1 α

6.1. Introduction

The results in Chapter 4 suggest that IRE1 α 's luminal domain's conformational landscape is in a dynamic equilibrium between monomers, dimers and increasingly large oligomers (Sections 4.2 and 4.3). The luminal domain's conformational landscape is influenced by changes to the domain's environment in a number of ways, such as the binding of peptide to shift equilibrium to promote the formation of larger multimeric species and the inclusion of BiP or interface mutations disfavoured oligomerisation. In addition to this, results presented in Chapter 5 suggest long-range effects and a potential allosteric role for the linker region in these conformations.

Numerous cancer-associated mutations of IRE1 α 's luminal domain have been identified, likely owing to the protein's role in adaptive and apoptotic cellular responses (Tam et al., 2014; Han et al., 2009; Chalmers et al., 2019), these mutations may affect the protein's conformational landscape or disrupt interactions of the protein to promote tumorigenesis. Here, four cancer-associated mutations are selected (Table 6.1) due to their occurrence in regions of the luminal domain distal to the protein's identified functional sites (dimerisation interface (Zhou et al., 2006), peptide binding groove (Zhou et al., 2006; Credle et al., 2005; Kono et al., 2017), oligomerisation interface and motifs to promote the oligomer-active state (Karagoz et al., 2017)). The mutations are therefore hypothesised to elicit effects through currently unknown allosteric mechanisms, which can be identified through study of the mutants.

Two of the selected mutations are in the core domain of the protein (N244S and S296F) and two are from the linker region (A414T and V418M), a brief overview of each mutation site is provided here.

The mutated core domain residues are highly conserved compared to the linker region residues. The N244 residue resides in the β -sandwich motif, this motif is suggested to have a role in structurally propagating the signal of peptide binding to promote an oligomerisation-active state of the protein (Karagoz et al., 2017), as shown in Figure 6.1A. Interestingly, in the crystal structure model (PDB: 2HZ6) the N244 residue has a polar contact with and resides in a similar area as the second mutated core domain residue; S296 (Zhou et al., 2006). The S296F mutant has been identified in multiple cases of skin cancer (Sanborn et al., 2015; Pickering et al., 2014) and

resides in a conserved anti-parallel β -sheet region close to the oligomerisation interface, loop region 2 and the β -sandwich region (Zhou et al., 2006). Thus far this conserved anti-parallel β -sheet has not been identified to have a role in luminal domain function, however, its interactions with the functionally important β -sandwich motif (particularly through N244 and S296) and the WLLI oligomerisation motif suggest it may also have a role in regulating peptide-induced stabilisation of the oligomerisation active state (Figure 6.1A).

Table 6.1

Cancer-associated mutations of IRE1 α 's luminal domain studied in this chapter. The conservation of the amino acid mutated (Ashkenazy et al., 2016), location as observed in Figure 6.1, cancer type and reference for the mutations are given.

Mutant	Conservation (Normalised score 0-100)	Location	Cancer-type	Reference
N244S	90.1	Start of β -sandwich domain	Clear cell carcinoma	(Greenman et al., 2007)
S296F	80.4	Conserved anti-parallel β -sheet	Cutaneous squamous cell carcinoma, metastatic melanoma	(Pickering et al., 2014; Sanborn et al., 2015)
A414T	4.5	Linker region	Glioblastoma	(Lhomond et al., 2018)
V418M	6.1	Linker region	Parathyroid carcinoma	(Greenman et al., 2007; Pandya et al., 2017)

The A414T linker region cancer-associated mutation was identified in an aggressive glioblastoma case and has been partially characterised for its cellular role, where the mutation promoted oligomerisation and enhanced *XBP-1* splicing as well as further phenotypical changes in the cells to promote the aggressive tumour observed (Lhomond et al., 2018). The V418M mutation was identified in parathyroid carcinoma (Greenman et al., 2007; Pandya et al., 2017) and has not been studied previously, its proximity to the A414T mutation suggests that it may interfere with the same mechanism that drives the A414T abhorrence. These mutations further suggest an allosteric role for the linker region in regulation of the protein, in particular the central section (Residues 411-420). The mutations are shown in Figure 6.1B alongside the information obtained for the linker region in Chapter 5.

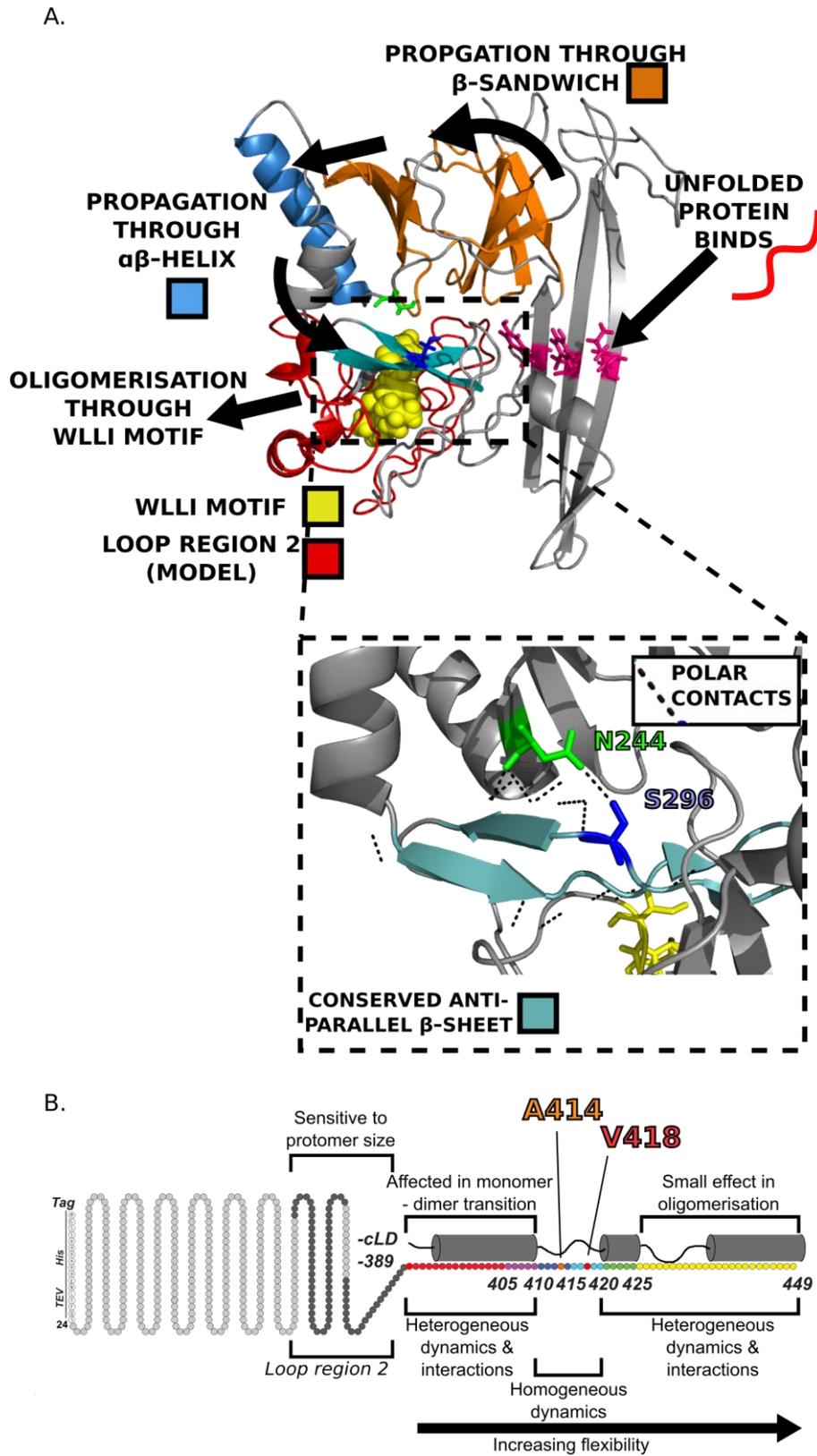


Figure 6.1

The cancer-associated mutants of the luminal domain occur in the linker region and core domain of the protein. A. Shown is a model of IRE1 α 's luminal domain's dimeric conformation based on the crystal structure (PDB: 2HZ6) (Zhou et al., 2006) with modelled intrinsically disordered

regions (Yang et al., 2015). The proposed signal propagation to promote oligomerisation from unfolded protein binding is shown (Karagoz et al., 2017). Additionally, the N244 and S296 residues are shown, these residues reside at an interface between motifs deemed important in signal propagation and a conserved anti-parallel β -sheet which has polar contacts with the oligomerisation interface and these motifs. Figure prepared using PyMOL version 1.7. B. A schematic representation of the protein's linker region with the findings from solution NMR study of the region in Chapter 5. Shown are the locations of the linker region cancer-associated mutations; A414T in orange and V418M in red.

This chapter will therefore explore the effects of these four cancer-associated mutations on IRE1 α 's luminal domain's conformational landscape. The N244S and S296F mutations may interfere with an allosteric pathway for unfolded protein-binding signal propagation to the oligomerisation interface and the linker region is known to adopt different conformations in different luminal domain states (Section 5.6), the cancer-associated mutations may therefore interfere with these conformations to influence the luminal domain's state through allosteric interactions.

6.2. Initial characterisation of the mutants' effects

6.2.1. Three of the mutants can be overexpressed as soluble protein in *E. coli*

In order to investigate the effect of the mutants *in vitro*, the mutant constructs were successfully produced using site directed mutagenesis with the *wild-type* full-length construct. The protocol and primers used are shown in Section 2.3.8.

To initially assess the solubility of the overexpressed mutants, the mutated luminal domain constructs were expressed using the same protocol as for *wild-type* protein (Figure 6.2). Expression of the N244S, S296F and V418M constructs at 37°C gave a significant amount of soluble protein, as with *wild-type* protein (Figure 6.2A and B), however, no soluble expression of the A414T luminal domain mutant was observed (Figure 6.2C). This suggests that all the cancer-associated mutations, excluding A414T can be overexpressed in *E. coli* using the same protocol as *wild-type* protein. The A414T construct is further investigated in Section 6.3. The soluble cancer-associated mutants were therefore expressed, purified and validated by mass spectrometry, showing similar chromatograms as the *wild-type* protein and the expected mass spectrometry results (Sections 2.5.2.2 and 2.6.1.2, respectively).

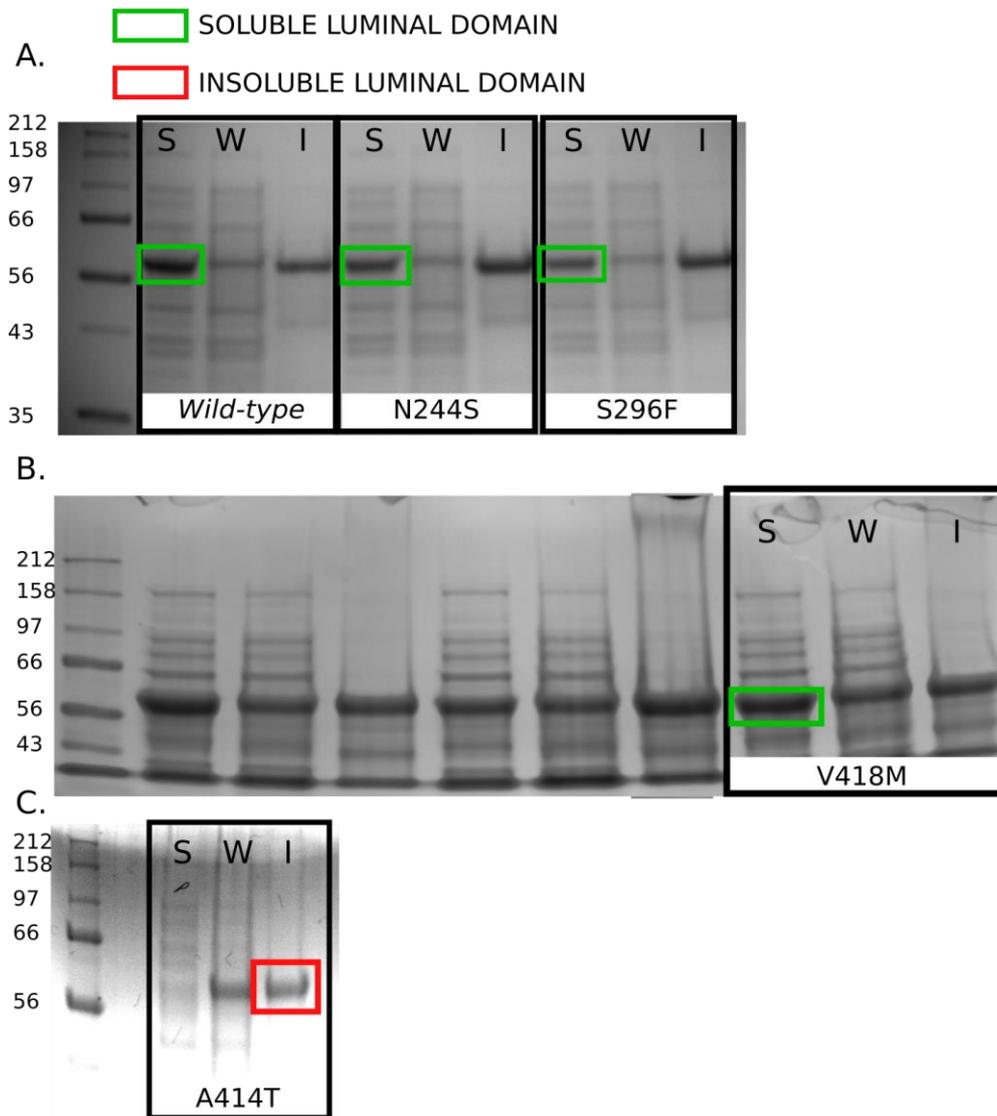


Figure 6.2

Solubility tests for *E. coli* overexpression of the cancer-associated mutants at 37°C with 1mM IPTG for four hours in BL21 DE3 cells. Solubility test protocol in Section 2.5.1, 'S' represents soluble protein, 'W' is whole protein sample and 'I' is the insoluble fraction. Highlighted in green is the soluble protein produced. A. Wild-type protein, N244S and S296F give a significant amount of soluble protein expressed. B. The V418M also gives a significant amount of soluble protein expressed. The solubility test for the V418M construct is shown in the box to the right, wells 2-7 can be discounted. C. A414T gives no significant levels of soluble protein, insoluble protein is highlighted by a red box.

6.2.2. The thermal stability of the expressed soluble cancer-associated mutants

In order to determine whether the mutations studied here affect the luminal domain's stability and therefore possibly perturb stress responses by the protein, thermal stability of the constructs was

measured using differential scanning calorimetry (Figure 6.3). Differential scanning calorimetry increases the temperature of a cell that contains the protein where the amount of energy required to raise the temperature of the cell is monitored. When the protein unfolds there is a peak in the energy required to increase the temperature of the cell due to the energy change upon unfolding. The mutants assessed here do not produce significant changes in the thermal stability of the protein, the largest loss in stability being 2.8°C from the S296F mutant, which still represents only a minor loss in stability. The N244S experiences a decrease of 1.8°C in thermal stability, which is similarly minor. As previously described the two core domain mutations have polar contacts, mutation of the residues may disrupt this interaction between the different areas of the protein and therefore cause the decrease in stability presented here. V418M did not appear to give any significant change in thermal stability to *wild-type* protein (+0.1°C difference).

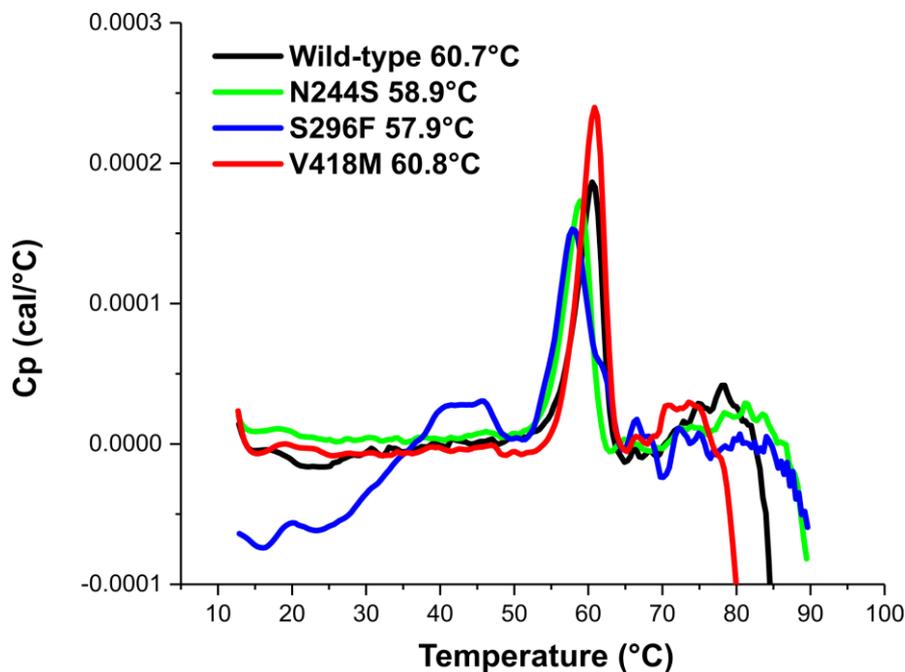


Figure 6.3

The mutants do not significantly perturb the thermal stability of the luminal domain. The mutants shown do not deviate significantly from wild-type protein's thermal stability of 60.7°C through analysis by differential scanning calorimetry (Section 2.9.1, $n=2$ for S296F, $n=3$ for other constructs). The mutants' thermal stabilities: N244S 58.9°C, S296F 57.9°C and V418M 60.8°C.

6.2.3. The cancer-associated mutants do not significantly perturb secondary structure

In order to assess the mutations' effects on the luminal domain's secondary structure, circular dichroism was used (Figure 6.4). Circular dichroism reports on the proportion of secondary structural elements in a protein, where large perturbations in structure cause significant deviations

in the circular dichroism spectrum. The mutations do not significantly alter the proportion of secondary structure elements, as observed in Figure 6.4, suggesting they do not promote large scale changes to luminal domain structure.

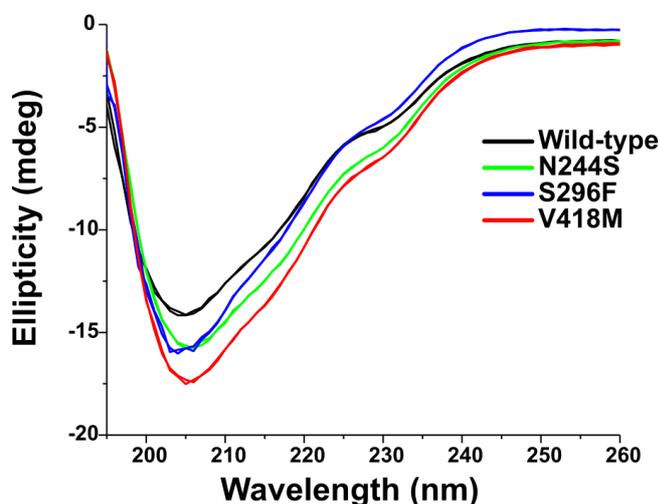


Figure 6.4

Circular dichroism spectra for all mutants show that the mutants do not significantly alter secondary structure components. Compared are the circular dichroism spectra (Section 2.6.3, $n=2$) for the mutants with the spectrum for wild-type protein. The cancer-associated mutants do not appear to significantly alter the secondary structure propensity of the protein, small changes in ellipticity are likely due to variation in sample concentration or the formation of amorphous aggregate of the protein over time (Bustamante et al., 1983).

From the results in this section it appears that the cancer-associated mutants do not significantly alter the stability or secondary structure of the protein. The mutations may therefore affect the protein's conformational landscape or interactions with binding partners to elicit an effect.

6.3. A414T mutation results in the formation of insoluble oligomers

6.3.1. Determination of the linker region's effect on solubility

A414 resides in the luminal domain's linker region (Figure 6.5) which is often removed for *in vitro* study of the protein (Karagoz et al., 2017; Zhou et al., 2006). Expression of the luminal domain with the linker region removed (core luminal domain) in *E. coli* produces less soluble protein than the full-length construct (Figure 6.6). This indicates that the linker region may have a role in promoting soluble luminal domain overexpression, which may explain the insoluble expression observed for the A414T mutant construct. Therefore, multiple truncations to the C-terminal of the linker region were used (Figure 6.5) and their solubility when overexpressed in *E.*

coli at different temperatures analysed to determine the presence of a solubility promoting segment of the linker region (Figure 6.6).

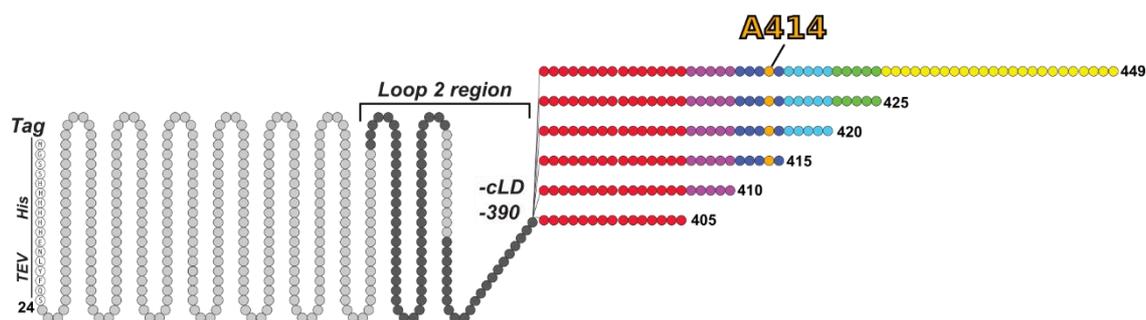


Figure 6.5

Schematic of the truncations made to the luminal domain. Shown also are loop region 2, where the linker region is truncated for the core luminal domain (cLD), the multiple truncations of the linker region used and where the A414T mutation occurs in the linker region.

In all truncated linker region constructs there were significant amounts of soluble protein produced at 20°C, 30°C and 37°C (Table 6.2) and it wasn't possible to determine a specific region that increased the solubility of the protein, thus suggesting that there is no specific segment of the linker region that affects solubility of expressed protein. Only the truncation that also removed part of loop region 2 and the oligomerisation interface (residues 24-356) gives no soluble protein when expressed at the temperatures analysed. Therefore, it appears that the A414T mutation doesn't disrupt a specific region of the linker region required for soluble *E. coli* overexpression.

Table 6.2

For the *E. coli* overexpression of each construct at different temperatures the presence of soluble luminal domain is presented. A '+' indicates the presence of soluble protein where '-' indicates the absence. Truncations of the linker region do not prevent soluble protein being expressed as the A414T mutant does.

CONSTRUCT:	SOLUBLE EXPRESSION		
	20°C	30°C	37C
24-389	+	+	+
24-405	+	+	+
24-410	+	+	+
24-415	+	+	+
24-420	+	+	+
24-425	+	+	+
24-449 (WILD-TYPE)	+	+	+

A414T	-	No data	-
D123P	+	+	-
24-356	-	-	-

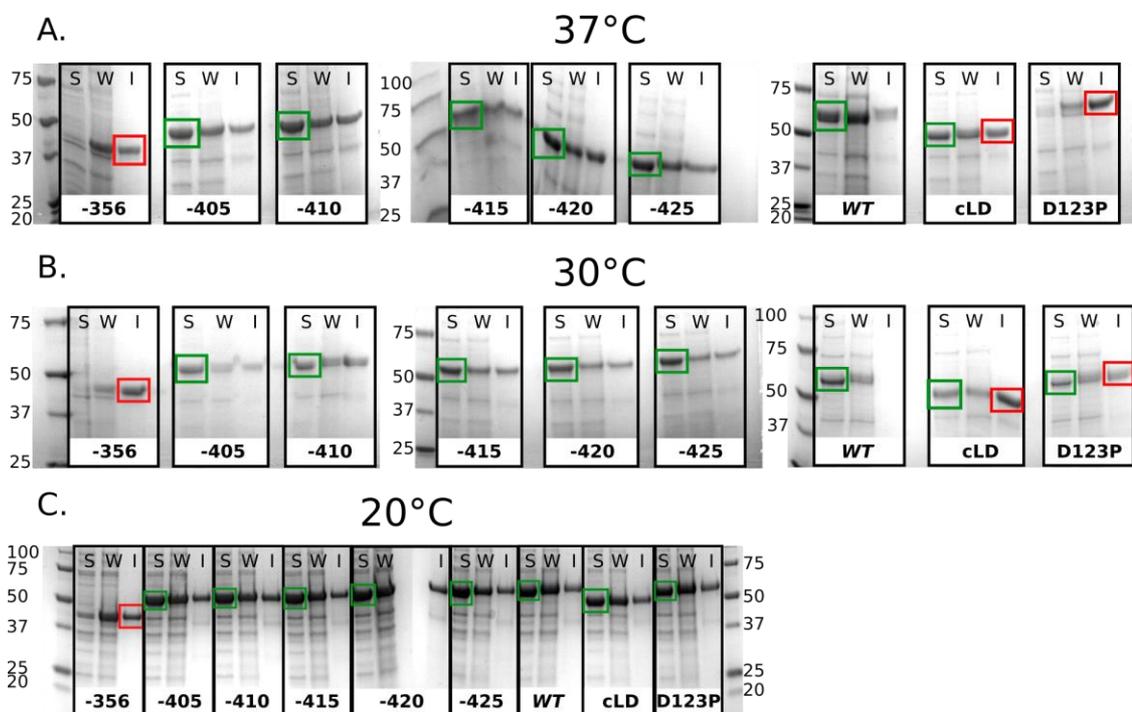


Figure 6.6

There doesn't appear to be a solubility promoting segment in the linker region. 'S' represents soluble protein, 'W' is whole protein sample and 'I' is the insoluble fraction. A-C. SDS-PAGE gels of solubility tests (Section 2.5.1) for expression of different luminal domain constructs at 37°C, 30°C and 20°C. Green boxes show significant soluble luminal domain, red boxes show where there are large amounts of insoluble protein. D123P requires expression at 20°C or 30°C for significant amounts of soluble protein. The core luminal domain lacking the linker region has a reduced amount of soluble protein when expressed above 20°C. The truncations used all show significant amounts of soluble protein at all temperatures trialled. The 24-356 construct shows low solubility under all expression conditions.

6.3.2. Optimisation of expression does not promote soluble expression of the A414T construct

The source of the A414T construct insoluble expression may be due to an adverse intramolecular interaction promoted by the mutation to reduce soluble expression. Therefore, to slow the translation rate and promote protein stability and soluble expression, different conditions for recombinant overexpression were used and further solubility tests carried out (Figure 6.7). The

D123P mutant has lower solubility in *E. coli* overexpression than *wild-type* luminal domain protein and is therefore expressed at 20°C to obtain significant amounts of soluble protein (Figure 6.7A). However, the same 20°C expression to promote stability and reduced translation rates did not yield soluble expression of the A414T construct. Lower IPTG concentrations were also trialled (Figure 6.7B and 6.7C) to reduce the expressed protein concentration alongside a different *E. coli* expression strain (Figure 6.7C) to optimise translation rates for soluble protein expression. However, all conditions used did not give significant soluble expression of the A414T mutant, suggesting that a different mechanism may be promoting insoluble expression of the A414T construct.

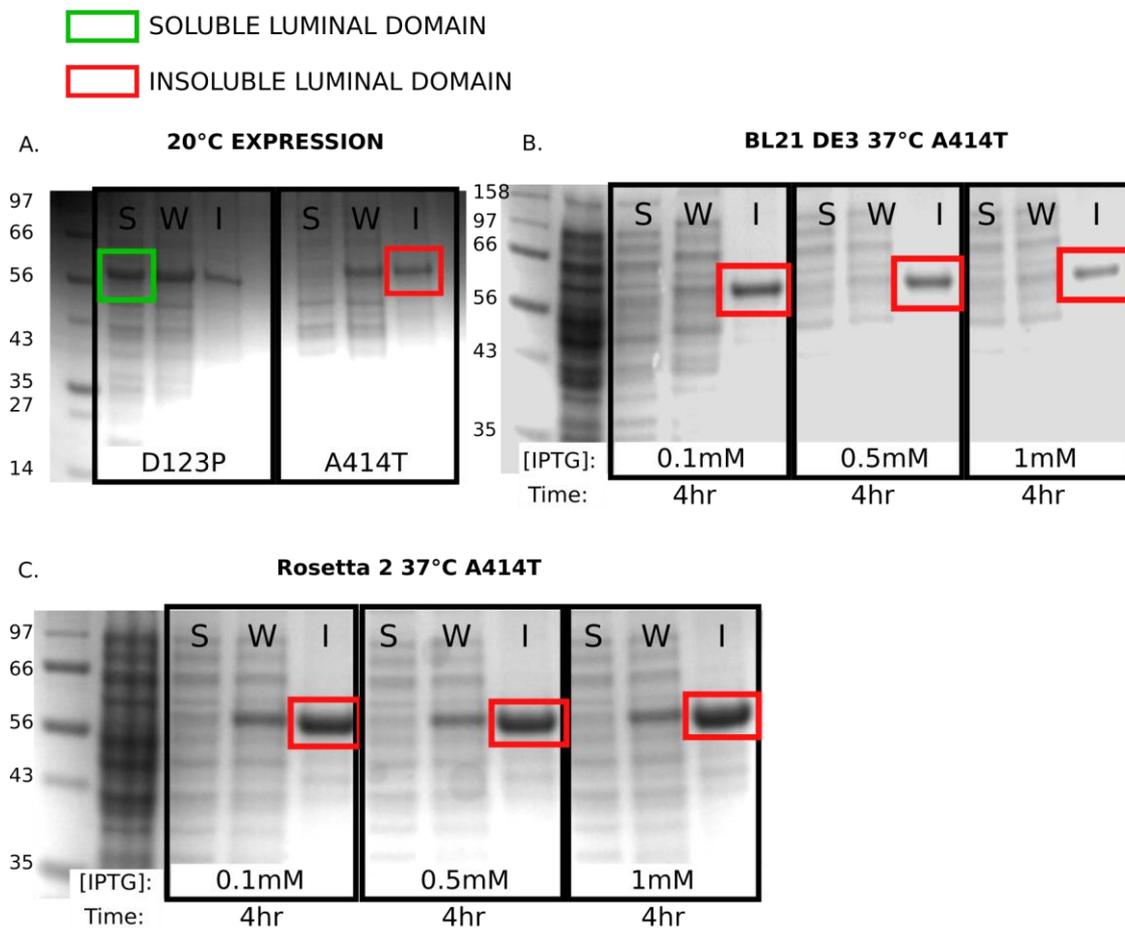


Figure 6.7

The A414T cancer-associated mutation is expressed in the insoluble fraction under all conditions trialled. 'S' represents soluble protein, 'W' is whole protein sample and 'I' is the insoluble fraction. A. Some mutants of the luminal domain require expression at 20°C for 16 hours in *E. coli* for soluble protein expression such as the D123P mutant (Section 2.4.5.4), under these conditions A414T is not produced as soluble protein, as determined by a solubility test (Section 2.5.1). B. Different concentrations of IPTG for induction of expression were trialled in the BL21 DE3 *E. coli* expression strain. However, A414T was not expressed as soluble protein at any

concentration. C. Different concentrations of IPTG for induction of expression were trialled in the Rosetta 2 *E. coli* expression strain. However, A414T was not expressed as soluble protein at any concentration.

6.3.3. Perturbation of the dimer and oligomer interfaces results in soluble A414T expression

The data presented thus far suggest that the A414T mutant's insoluble expression is not caused by a solubility-promoting motif in the linker region being perturbed or interference by the mutation with a different solubility providing motif. The luminal domain is able to form insoluble oligomers upon incubation with peptide, as observed in Section 4.3.2.1 and previous characterisation of the A414T mutant protein in cell-based assays suggested that the mutant promotes increased oligomerisation in the absence of stress (Lhomond et al., 2018). It was therefore hypothesised that the A414T mutation may promote insoluble oligomerisation of the luminal domain when overexpressed in *E. coli*.

In Chapter 4, the dimerisation or oligomerisation deficient mutants (D123P and WLLI-GSSG) (Zhou et al., 2006; Karagoz et al., 2017) disfavoured peptide-induced oligomerisation by affecting the luminal domain's conformational landscape (Section 4.3.2.3). Therefore, inclusion of the mutations to the A414T construct may promote soluble expression if the A414T does indeed promote insoluble oligomer formation. Double and quintuple mutants were produced (A414T+D123P and A414T+WLLI-GSSG) and the solubility of each mutants' *E. coli* overexpression analysed (20°C expression, Figure 6.8). Indeed, the double and quintuple mutants both give significant amounts of soluble luminal domain, suggesting that the A414T mutation affects IRE1 α 's luminal domain's conformational landscape by promoting the formation of insoluble oligomers in the absence of unfolded protein substrate.

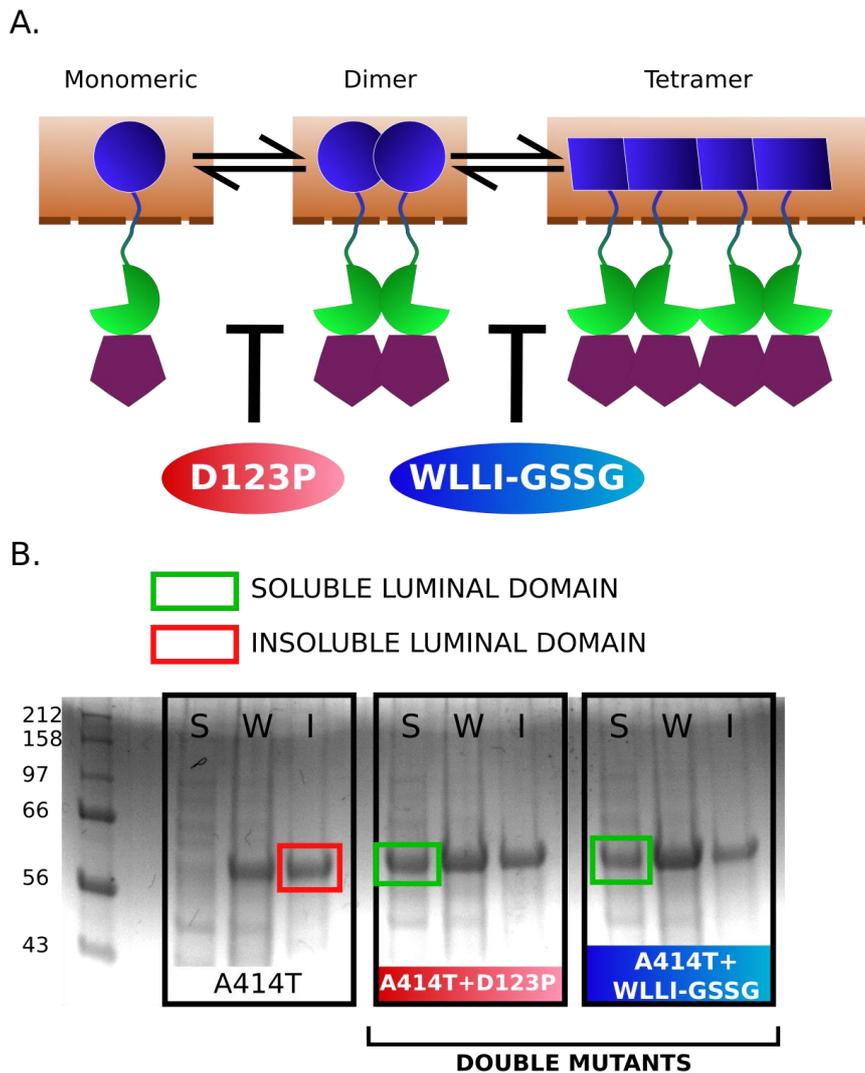


Figure 6.8

The dimerisation and oligomerisation deficient mutants, D123P and WLLI-GSSG influence the A414T mutant's conformational landscape to promote formation of smaller species, therefore giving soluble *E. coli* overexpression. A. Representation of the effect of the D123P mutant and WLLI-GSSG mutant on the luminal domain's conformational landscape in the absence of unfolded protein. D123P inhibits the formation of dimeric protein (Zhou et al., 2006) and WLLI-GSSG inhibits the formation of oligomers (Karagoz et al., 2017). B. SDS-PAGE gel of solubility tests (Section 2.5.1) carried out with the A414T containing double and quintuple mutants when expressed at 20°C for four hours with 1mM IPTG. The A414T mutant alone doesn't produce a significant amount of soluble luminal domain protein. When the A414T construct is mutated with the dimerisation or oligomerisation deficient mutants there is a significant amount of soluble protein produced, as shown in the green boxes. This suggests that A414T promotes oligomerisation, causing large oligomers to precipitate when overexpressed in *E. coli* cultures.

The results presented here also further demonstrate the importance of the linker region in activation of the luminal domain, however, the mode of perturbation that the A414T mutation has on the luminal domain remains unknown and thus requires structural characterisation. The A414T + D123P double mutant was recombinantly expressed in the soluble fraction and used in the subsequent analysis of the mutation's effects on the luminal domain (purification chromatogram and mass spectrometry result for the A414T+D123P double mutant shown in Sections 2.5.2.2 and 2.6.1.2, respectively).

6.3.4. The A414T mutation doesn't cause a significant change in secondary structure of the D123P construct

To observe if the A414T mutation significantly perturbs the secondary structure proportions of the D123P construct, circular dichroism was used (Figure 6.9). No significant differences in the circular dichroism spectra are observed between the D123P and A414T+D123P mutants. As observed in Chapter 5, NMR study of the protein's disordered regions will give more accurate information about the conformational changes conferred by the A414T mutation and are carried out in Chapter 7.

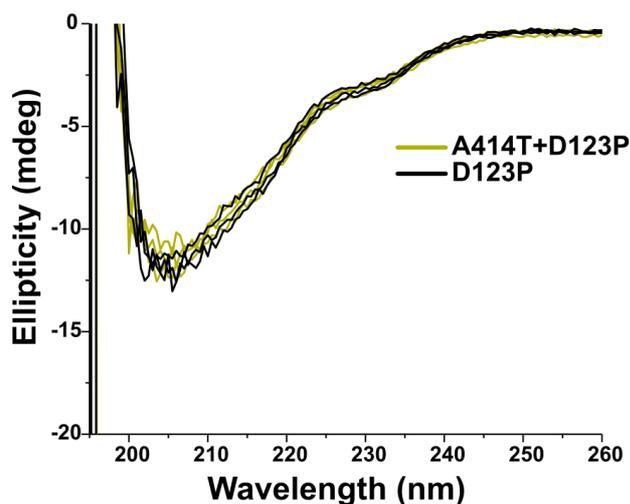


Figure 6.9

Circular dichroism spectra for the A414T+D123P double mutant, the mutation does not significantly alter secondary structure components. Compared are the circular dichroism spectra (Section 2.6.3, n=3) for the A414T+D123P double mutant the spectra for the D123P mutant. Experiment carried out at 4 μ M, the difference observed compared to Figure 6.4 may be due to error in concentration measurement or the decreased stability of the D123P construct.

The A414T mutation appears to promote oligomerisation of the luminal domain in the absence of unfolded protein and requires the D123P mutation in the construct for soluble *E. coli* expression.

This double mutation and the soluble cancer-associated mutations do not appear to significantly perturb protein structure or stability and therefore can be probed for their effects on the luminal domain's conformational landscape and interactions in the following sections.

6.4. How the cancer-associated mutations affect IRE1 α 's conformational landscape

The assays developed in Chapter 4 to study the luminal domain's conformational landscape are used here to identify how the cancer-associated mutations perturb luminal domain function. To study the luminal domain's state in the absence of unfolded proteins, size exclusion chromatography is used. Here, the *wild-type* protein exhibits a sub- μ M dimerisation constant with oligomeric species being formed at higher concentrations (Section 4.2). To assess the constructs' activation in stressed conditions, Δ EspP was incubated with the luminal domain and turbidity, solubility and fluorescence polarisation techniques used to observe oligomerisation of the domain and compare this to *wild-type* protein (Section 4.3).

Therefore, each of the following sections will address the conformational landscape of the luminal domain cancer-associated mutations and how oligomerisation is promoted by a model peptide substrate, using the methods and assays presented in Chapter 4.

6.4.1. No perturbations to luminal domain activation are observed in the N244S mutant

Upon analysis of the N244S mutant's effect on the luminal domain's conformational landscape, the construct appears to behave similarly to *wild-type* protein (Figure 6.10). Figure 6.10A suggests that the protein has a sub- μ M dimerisation constant, similar to *wild-type* protein (Section 4.2). Additionally, the N244S mutant doesn't appear to perturb the luminal domain's peptide-induced oligomerisation or formation of insoluble species, as presented in Figure 6.10B and C where the N244S construct undergoes significant oligomerisation as assessed by fluorescence polarisation and solubility assays. It therefore appears that the N244S mutation does not perturb the luminal domain's conformational landscape in the presence and absence of peptide, or at least to an extent that is observable by the assays used here.

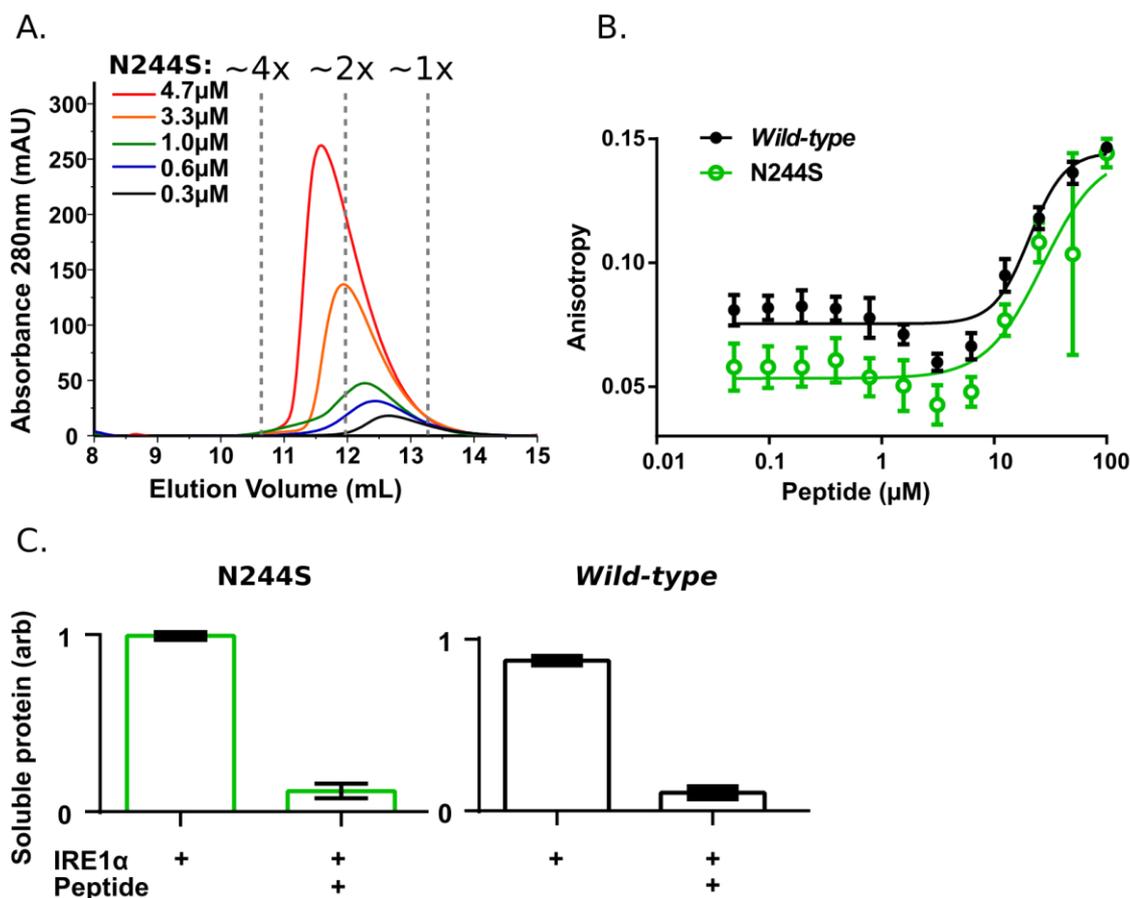


Figure 6.10

The N244S mutation does not appear to perturb IRE1 α 's luminal domain's conformational landscape or its interaction with peptide to induce oligomerisation. A. Size exclusion chromatography (Section 2.6.2) with calculated concentration of N244S protein on the column suggests that the construct exists as an equilibrium of monomers and dimers with oligomerisation at higher concentrations, similar to wild-type protein (Section 4.2). B. Fluorescence polarisation (Section 2.9.4, $n=3$) study of the mutant and wild-type protein oligomerising in response to peptide after 30 minutes of incubation. N244S appears to oligomerise with a similar apparent constant as wild-type protein. C. A solubility test of the protein after incubation with 169 μM Δ EspP for three hours suggests that the mutant has the ability to form insoluble oligomers, similar to wild-type protein (Section 2.9.6.2, $n=3$). Arbitrary units used for amount of soluble protein.

6.4.2. The S296F mutant affects luminal domain oligomerisation and the stability of disulphide bonds

6.4.2.1. The S296F mutant moderately stabilises oligomers

The S296F mutant induces a profound effect to promote tetramerisation when analysed by size exclusion chromatography (Figure 6.11A). The mutant appears to exist in an equilibrium of tetramers and dimers at the same concentrations where *wild-type* luminal domain exists as

monomers and dimers (Figure 6.11B). Upon plotting the size exclusion values, the tetramerisation constant appears to be 1-3 μ M (Figure 6.11B), where the *wild-type* tetramerisation constant is estimated as >10 μ M.

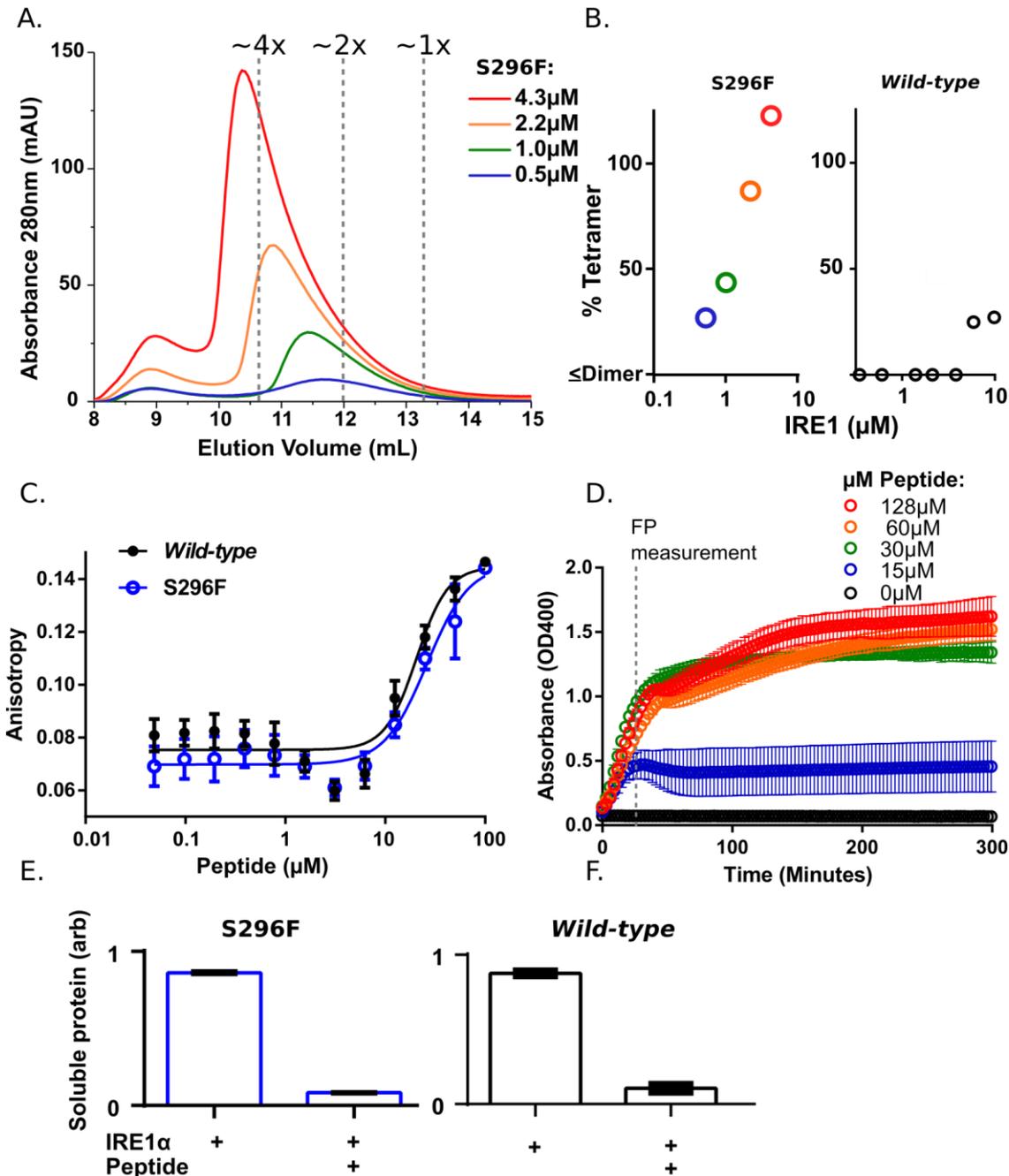


Figure 6.11

The S296F mutation promotes formation of tetrameric luminal domain species in the absence of stress. A. Size exclusion chromatography (Section 2.6.2) with the calculated concentration of S296F protein on the column suggests that the construct exists as an equilibrium between tetramers and dimers (larger species also present). B. Plotting the size exclusion estimated size of multimer (calibrated using D123P mutant mass with calibration curve in Section 2.6.2.2) with

calculated protein concentration suggests a tetramerisation constant of $\sim 3\mu\text{M}$, this is compared to wild-type protein with a constant of $>10\mu\text{M}$. C. Fluorescence polarisation (Section 2.9.4, $n=3$) study of the mutant and wild-type protein oligomerising in response to peptide after 30 minutes of incubation. S296F appears to oligomerise with a similar apparent constant as wild-type protein. D. Analysis of the protein's peptide-induced oligomerisation by turbidity assays (Section 2.9.6, $n=3$) suggests that the mutant does not significantly affect the effect of peptide interaction on the luminal domain's conformational landscape. E. A solubility test of the protein after incubation with $169\mu\text{M}$ ΔEspP for three hours suggests that the mutant forms insoluble oligomers, similar to wild-type protein (Section 2.9.6.2, $n=3$).

Although the S296F mutant appears to promote moderate oligomerisation in the absence of unfolded protein, this did not affect the observed response of the luminal domain to peptide-induced oligomerisation, as measured by fluorescence polarisation and turbidity assays, as well as a solubility assays to measure formation of insoluble oligomers not differing from the response of *wild-type* protein (Figures 6.11C, D (Compare to Figure 4.6A) and E). This suggests that the mutant's propensity to form oligomers does not affect peptide-induced oligomerisation to an observable degree.

As observed in the size exclusion chromatogram of the S296F mutant (Figure 6.11A) there are larger species than the identified tetramers eluting from the column, the formation of these species will be discussed in the following section.

6.4.2.2. The S296F mutant appears to protect formed disulphide bonds

Disulphide bonds can increase the stability of specific protein conformations to promote function (Hogg, 2003), in the case of IRE1 α 's luminal domain they are thought to be involved in stabilising an activated, oligomeric state (Eletto et al., 2014), as discussed in Section 1.4.4. Disulphide bonds can be reduced by small molecule agents such as DTT or by proteins such as PDIs (Wilkinson and Gilbert, 2004) and protection from such agents can be achieved by reducing the bond's solvent exposure or tension on the bond (Hogg, 2003; Zhang, D. et al., 2019; Ohri et al., 2018).

Size exclusion chromatography analysis of the S296F mutant using the same conditions as *wild-type* protein (Sections 4.2 and 2.6.2.1) presents large species that elute in the void volume ($>2000\text{kDa}$), the formation of which do not appear to be dependent on protein concentration, as shown in Figure 6.12A. It was therefore hypothesised that the formation of these large species may be due to more stable disulphide bonds being formed in the S296F construct.

To investigate this, the S296F protein was incubated for 36 hours with reducing agent to overcome the potentially reduced solvent exposure of the construct's disulphide bonds. This sample is compared to a sample incubated with reducing agent for <1 hour, which is the same condition used for the other mutant and *wild-type* luminal domain constructs (Figure 6.12B). In this case the sample incubated for an extended time presented large reduction in the large disulphide oligomers eluting in the void volume to mostly tetrameric protein (similar to those observed in Figure 6.12A). This therefore indicates that S296F stabilises formed disulphide bonds in the protein.

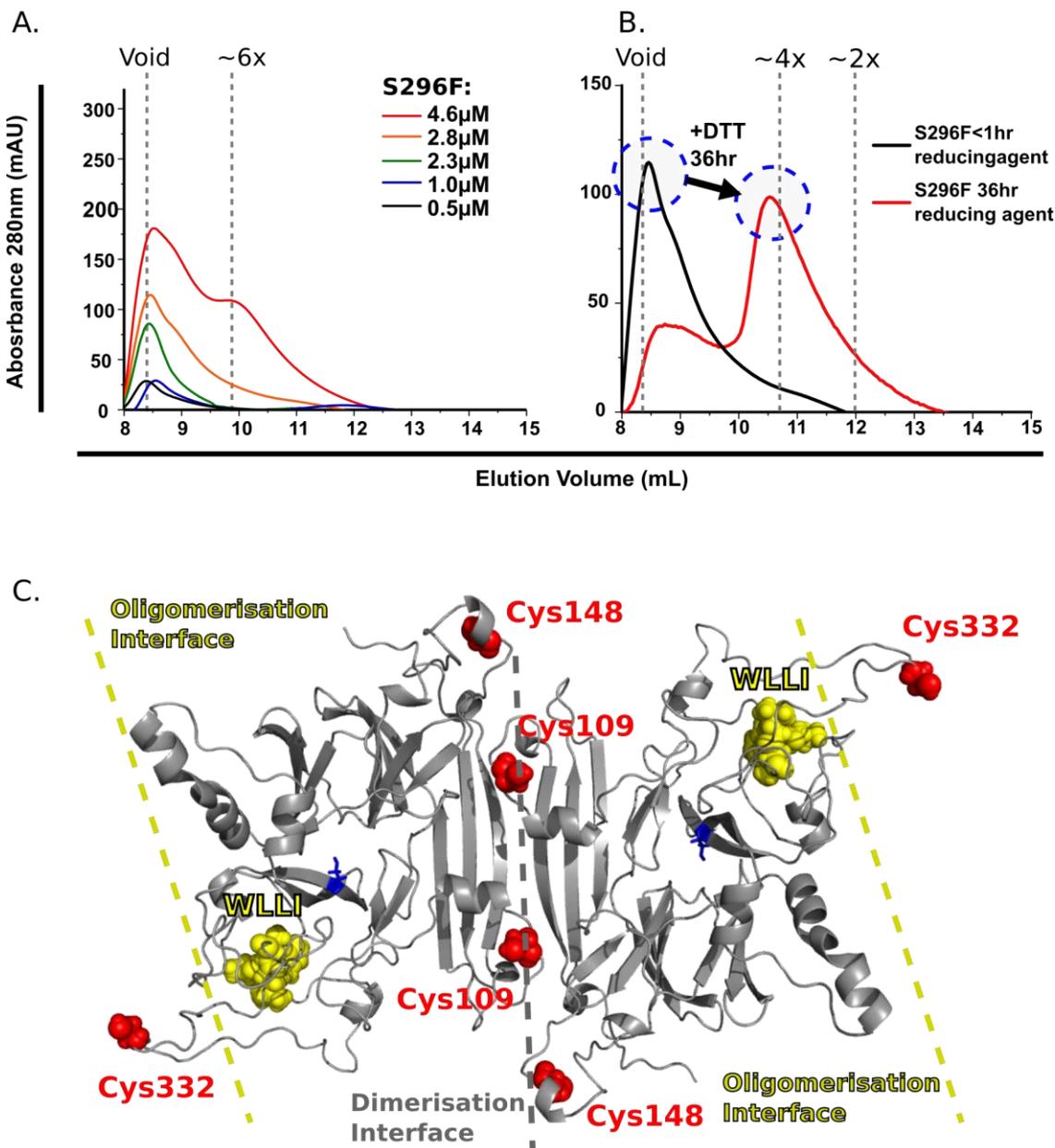


Figure 6.12

The S296F mutant causes disulphide bonds to be protected. Once the disulphide bonds are reduced the protein exists as tetramers. A. Size exclusion chromatogram (Section 2.6.2) of S296F

when purified and prepared as wild-type protein is. Large species eluting in the void volume are present at all concentrations used. B. Two samples of S296F were analysed by size exclusion chromatography. One was incubated with reducing agent as usual (protocol used for previous wild-type size exclusion chromatography experiments, Section 2.6.2.1) and the other was incubated with 5mM of a reducing agent (DTT) for 36 hours before being injected onto the column. After incubation with the reducing agent, a large portion of the disulphide bonded void volume peak was reduced and a tetrameric peak emerged. C. Crystal structure model (PDB: 2HZ6) (Zhou et al., 2006; Yang et al., 2015) with cysteine residues shown. Figure prepared using PyMOL version 1.7.

It's therefore apparent that the S296F mutant promotes formation of IRE1 α oligomers in the absence of unfolded protein. The cause of the mutant's stabilisation of disulphide bonds is not clear, the effect may be due to the construct's increased oligomerisation promoting disulphide bond formation, or the mutation may promote adoption of a conformation that reduces solvent access to formed disulphide bonds (Zhang, D. et al., 2019). Investigation of this mechanism also requires further understanding of the *wild-type* protein's propensity to form disulphide bonds when oligomerised. Interestingly, the C332 residue (which is thought to be involved in disulphide bond formation in the luminal domain, work carried out by Sam Dawes) resides close to the oligomerisation interface and S296F mutation site in the flexible loop region 2 (Figure 6.12C), it's possible that the S296F mutation promotes the protection of disulphide bonds through this residue by stabilising oligomeric complexes.

After this finding the S296F mutant was purified in the presence of reducing agent to remove disulphide bonds (Section 2.5.2.2.2) and so for all other results presented in this thesis (including Figure 6.11) the S296F mutant is in a reduced state.

6.4.3. The A414T mutation promotes peptide-induced oligomerisation in the D123P construct

The size exclusion chromatogram for the A414T+D123P double mutant is presented in Figure 6.13A. The elution volumes are similar to what is expected for the dimerisation deficient D123P mutant (Section 4.2, Figure 4.2), suggesting the construct is monomeric. The A414T + D123P construct was assessed for its ability to oligomerise, promoted by interaction with peptide (Figure 6.13B and C). Fluorescence polarisation analysis of peptide-induced oligomerisation suggests that the A414T + D123P construct oligomerises with a similar constant as with the D123P protein. However, the double mutant has an increased anisotropy with lower concentrations of peptide, this suggests formation of larger species. It's worthy of note that the fluorescence polarisation

values are normalised to the value obtained with the highest concentration of peptide added, and this may be different between the A414T + D123P and D123P samples. Further characterisation of the mutant's peptide-induced oligomerisation by turbidity assays corroborates the fluorescence polarisation data, suggesting that the A414T mutant has a higher propensity to form oligomers upon addition of peptide (Figures 6.13C and D) than the D123P mutant alone. Figure 6.14E presents the A414T+D123P double mutant's ability to form insoluble oligomers upon incubation with peptide, similar to *wild-type* protein.

The A414T mutation promotes formation of insoluble oligomers upon overexpression in *E. coli* and the addition of the dimerisation deficient D123P mutation (Zhou et al., 2006) promotes soluble expression by disfavoured oligomerisation in the luminal domain's conformational landscape. The results here further suggest the model of the protein's dynamic conformational landscape, as incubation with peptide promotes the formation of large oligomers in the A414T + D123P construct, opposing the effect of the D123P mutation. Additionally, the A414T mutation appears to moderately increase the propensity of the D123P protein to form oligomers in the presence of peptide (Figure 6.13C and D). This further corroborates the results in Section 6.3.3, where the A414T mutant construct appears to promote oligomerisation of the luminal domain and previous literature, suggesting that the mutation promotes oligomerisation of the whole protein (Lhomond et al., 2018).

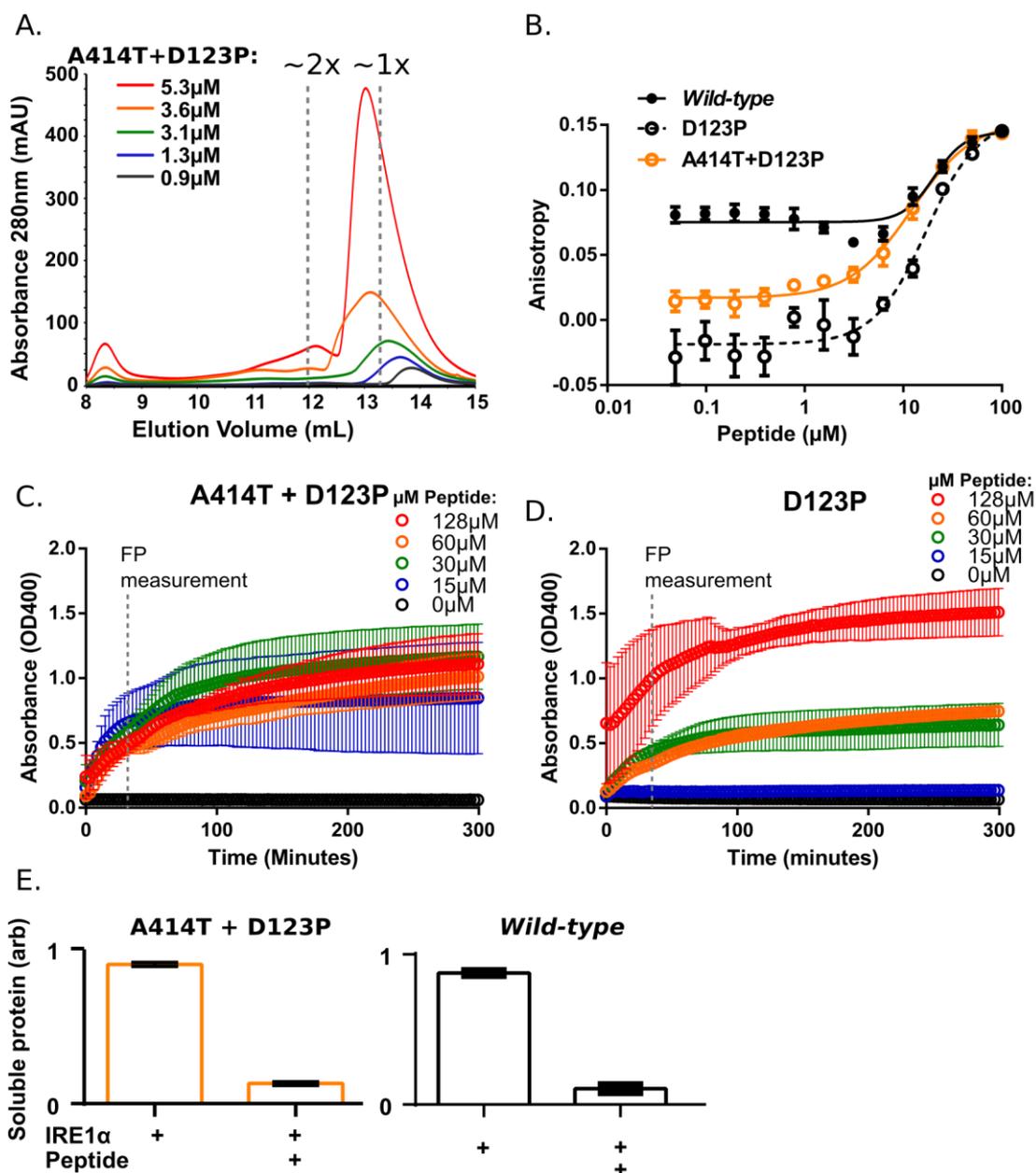


Figure 6.13

The A414T+D123P double mutation promotes peptide-induced oligomerisation. A. Size exclusion chromatography (Section 2.6.2) with the calculated concentrations of A414T + D123P protein on the column suggests that the construct elutes similarly to the D123P construct. The protein appears to be slightly larger, which requires further investigation to determine significance. B. Fluorescence polarisation (Section 2.9.4, $n=3$) study of the A414T+D123P double mutant, the D123P construct and wild-type protein oligomerising in response to peptide after 30 minutes of incubation. The A414T+D123P construct appears to promote the formation of larger species with less peptide added when compared to the D123P mutant, however this may be a result of the data being normalised to the data point with the highest concentration of peptide added. C. Analysis of the A414T+D123P construct's peptide-induced oligomerisation by turbidity assays (Section 2.9.6, $n=3$) suggests that the mutant promotes increased oligomerisation

with lower concentrations of peptide added compared to the D123P mutant alone (Panel D). D. Turbidity assay with the D123P mutant alone when incubated with different Δ EspP concentrations for comparison to the data in panel C. E. A solubility test of the protein after incubation with $169\mu\text{M}$ Δ EspP for three hours suggests that the A414T+D123P double mutant forms insoluble oligomers, similar to wild-type protein (Section 2.9.6.2, $n=3$).

6.4.4. The V418M mutation does not perturb the luminal domain's conformation landscape

Analysis of the V418M mutation suggests that in all aspects observed by the assays used, it behaves similarly to *wild-type* protein (Figure 6.14). The V418M mutation does not appear to affect the protein's sub- μM dimerisation constant as analysed by size exclusion chromatography (Figure 6.14A). Additionally, peptide-induced oligomerisation and insoluble oligomerisation occurs in a similar manner as *wild-type* protein, suggesting the process is not perturbed upon mutation (Figures 6.14B-D).

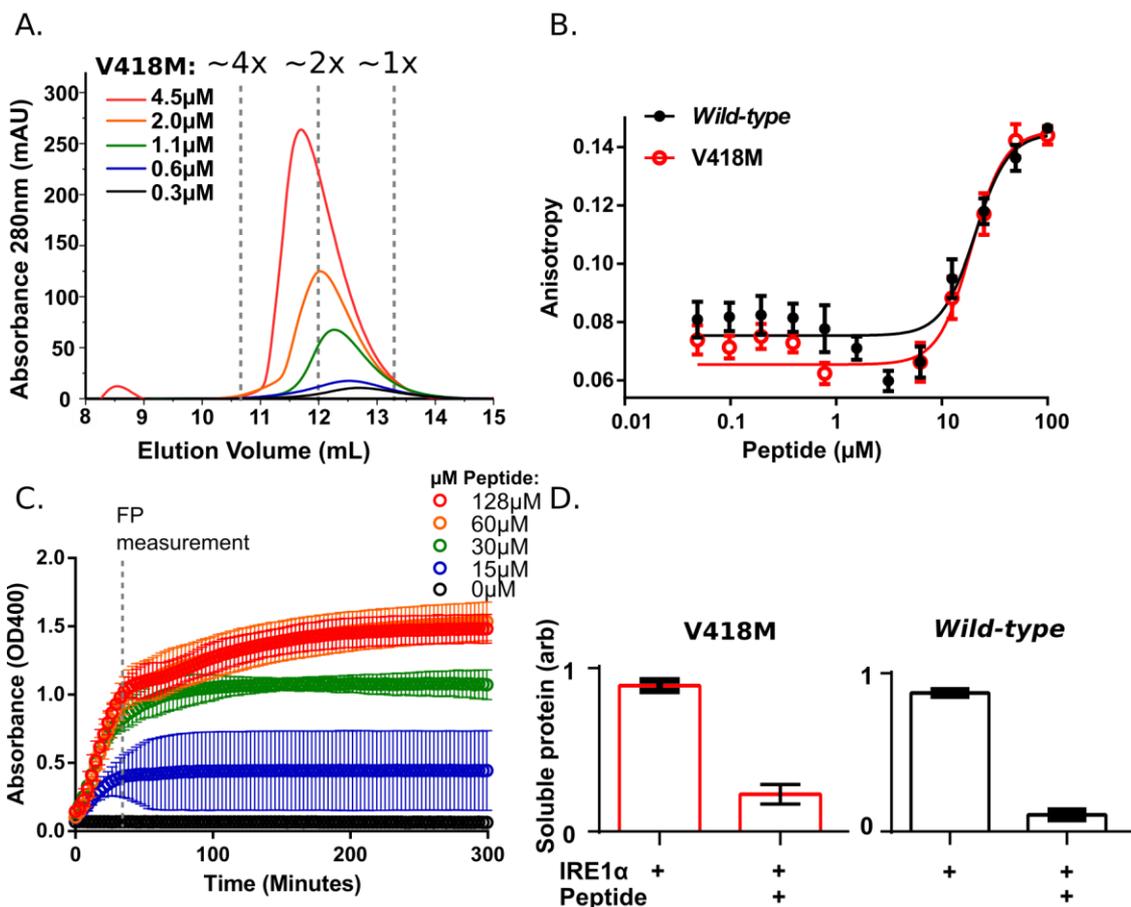


Figure 6.14

The V418M mutation does not appear to affect the luminal domain's dynamic equilibrium in the assays used. A. Size exclusion chromatography (Section 2.6.2) with the calculated concentrations of V418M protein on the column suggests that the construct exists in an equilibrium similar to

wild-type protein. B. Fluorescence polarisation (Section 2.9.4, n=3) study of the mutant and wild-type protein oligomerising in response to peptide after 30 minutes of incubation. The V418M mutant appears to oligomerise with a similar apparent constant as wild-type protein. C. Analysis of the protein's peptide-induced oligomerisation by turbidity assays (Section 2.9.6, n=3) suggests that the mutant does not significantly affect the effect of peptide interaction on the luminal domain's conformational landscape. D. A solubility test of the protein after incubation with 169 μ M Δ EspP for three hours suggests that the V418M mutant has the ability to form insoluble oligomers, similar to wild-type protein (Section 2.9.6.2, n=3).

6.5. The cancer-associated mutations do not perturb BiP's interaction with oligomeric luminal domain

In Section 4.4, BiP's activity to bind to and re-solubilise luminal domain oligomers was observed, as this activity likely acts to reduce IRE1 α activation, the cancer-associated mutations may oppose it to promote the adaptive response and aid tumorigenesis (Chalmers et al., 2019). Therefore, to determine whether the cancer-associated mutants inhibit BiP binding and re-solubilisation of insoluble oligomers, the peptide-oligomerised mutants were incubated with BiP and BiP with ATP and the soluble and insoluble fractions were assessed (Figures 6.15 and 6.16). SDS-PAGE gel samples shown in Figure 6.15 suggest that BiP is able to bind to the insoluble mutant luminal domain oligomers and re-solubilise them, as with *wild-type* protein in Section 4.4. The soluble and insoluble gel samples corroborate the soluble protein measurements presented in Figure 6.16. Therefore, the cancer-associated mutants assessed here do not appear to interfere with the re-solubilising activity of BiP on insoluble luminal domain oligomers.

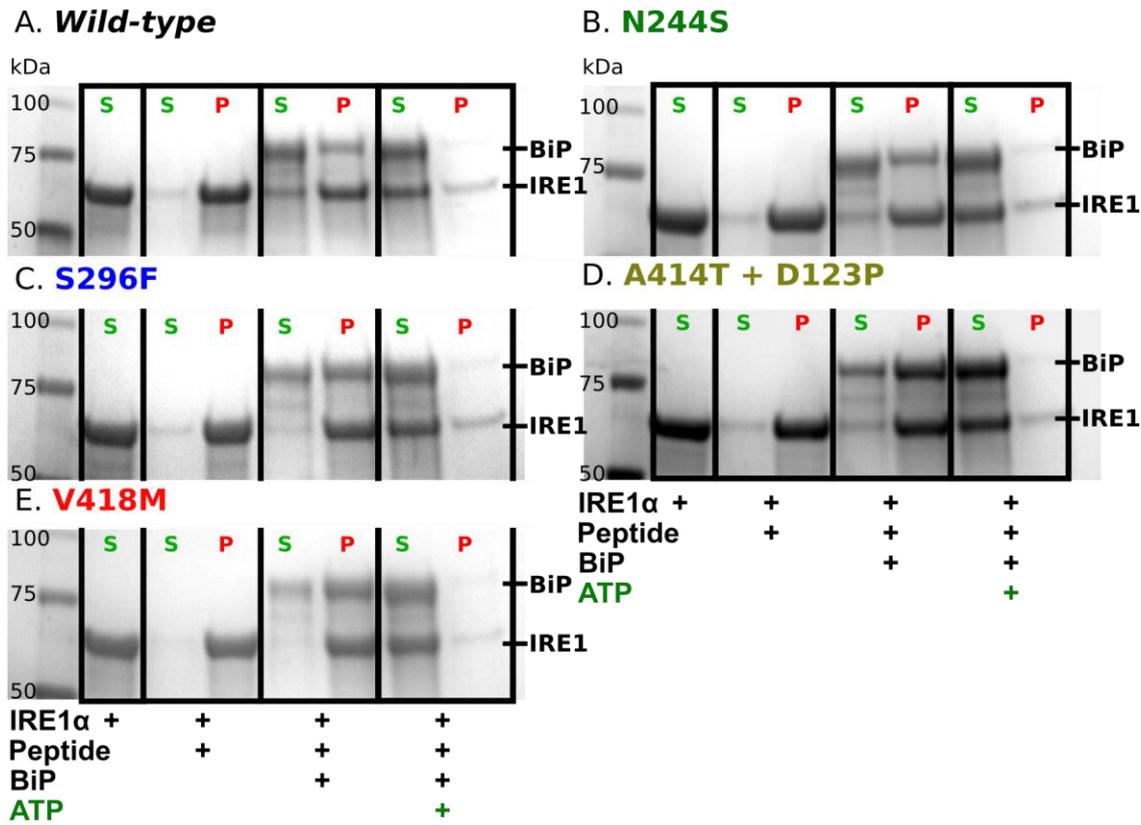


Figure 6.15

SDS-PAGE gels of the BiP-dependent termination assays carried out for each cancer-associated mutant suggest that the mutants do not interfere with BiP's re-solubilising activity. 'S' represents soluble protein and 'P' represents the pellet (insoluble) fraction of protein present. A-E. In all cases almost all luminal domain is insoluble after incubation with 169 μ M Δ EspP. Upon addition of BiP, some BiP also appears in the insoluble fraction along with the majority of luminal domain, suggesting BiP binding to the insoluble luminal domain oligomers. After three hours of incubation with BiP, ATP and the insoluble oligomeric protein, the vast majority of BiP and luminal domain are present in the soluble fraction (Section 2.9.7, n=3).

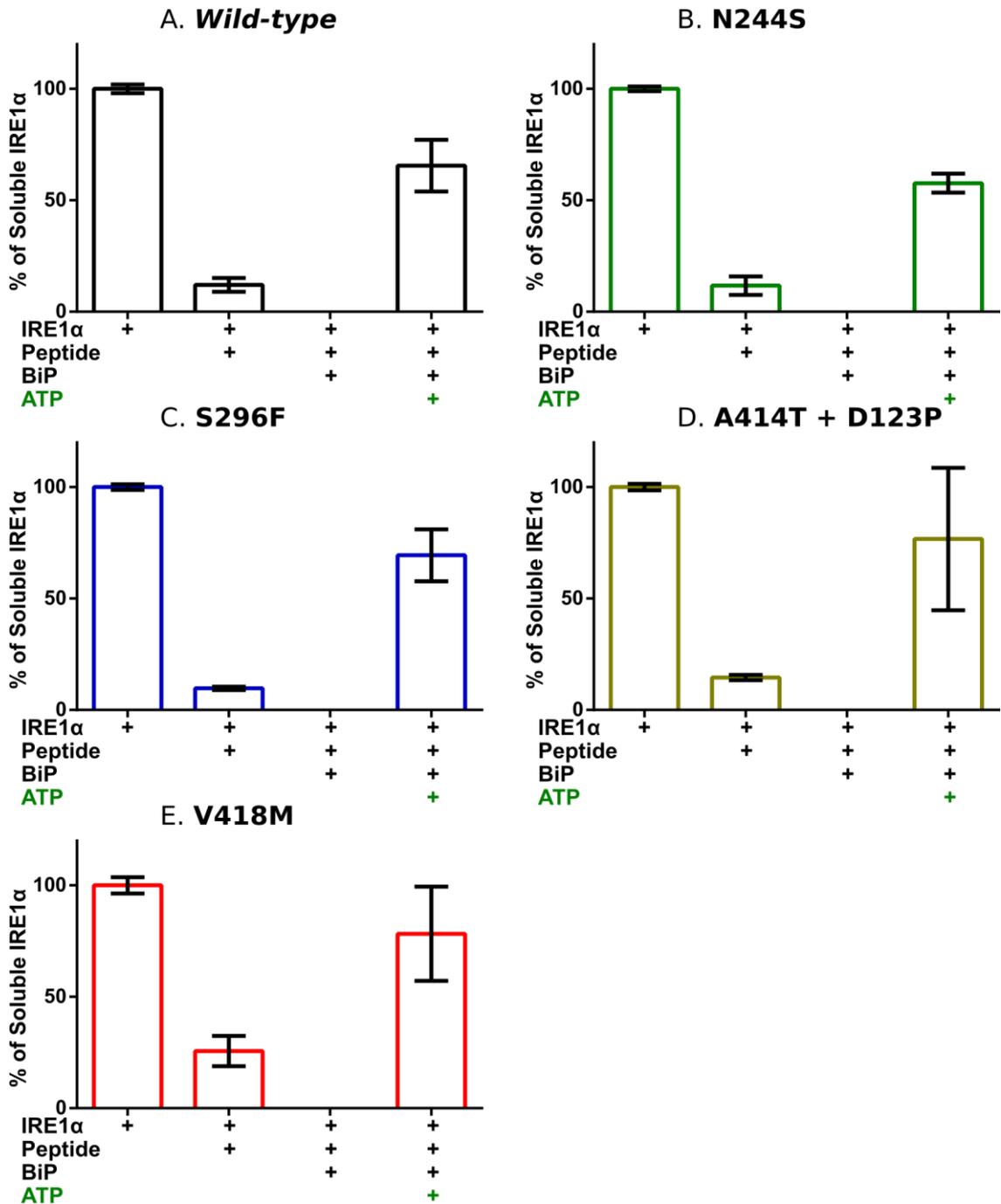


Figure 6.16

BiP's ability to solubilise the luminal domain cancer-associated mutants after peptide-induced oligomerisation is not significantly different to wild-type protein. 20 μ M of the luminal domain was incubated with 169 μ M Δ EspP for three hours before 24 μ M of BiP was added and incubated for three hours. Finally, 40mM of ATP was added to the samples for three hours. Before the addition of each species and at the end of the ATP incubation, samples were taken and centrifuged to remove insoluble protein and the soluble protein concentration measured (Section 2.9.7, n=3). A-E. All of the cancer-associated mutants have similar amounts of insoluble protein after incubation with peptide. After addition of BiP the total soluble protein in the sample is increased

due to soluble BiP present. After incubation with ATP all mutants show similar rescue of insoluble luminal domain as with wild-type protein. This suggests that the cancer-associated mutants do not interrupt BiP-dependent termination of their oligomerised states. For reference, the wild-type protein was repeated here, differences in amount of luminal domain rescue are likely due to pipetting error in mixing samples. All measurements taken with Bradford assays and the percentage of soluble luminal domain calculated.

6.6. Conclusion

In this chapter, the effects of four selected cancer-associated mutations on the luminal domain's conformational landscape were investigated. Two of the mutations reside in the domain's core region (N244S and S296F) and two from the domain's flexible linker region (A414T and V418M). Through use of methods optimised in Chapter 4, two cancer-associated mutations (S296F and A414T) that promote luminal domain oligomerisation in the absence of endoplasmic reticulum stress have been identified (Figure 6.17). These mutations occur in distal sites to the functionally important regions of the protein previously identified for dimerisation and oligomerisation (Karagoz et al., 2017; Zhou et al., 2006), suggesting that their effects may be through allosteric mechanisms.

The A414T mutation promotes insoluble overexpression of the protein in *E. coli*. However, the subsequent soluble overexpression of the mutant upon inclusion of the dimerisation and oligomerisation disfavouring mutants suggests the A414T mutation's role in promoting oligomerisation in the absence of unfolded proteins and further displays the dynamic equilibrium of the luminal domain's conformational landscape. The domain's dynamic conformational landscape is further demonstrated by analysis of the peptide-induced oligomerisation of the A414T + D123P construct, where the A414T mutation has a higher propensity to form peptide-induced oligomers than the D123P construct alone (Figure 6.13C and D). The findings presented here for the A414T mutant corroborate with a previous study where the A414T mutant exhibited increased IRE1 α clustering and *XBP-1* splicing when compared to *wild-type* protein in a mammalian cell line (Lhomond et al., 2018). This finding also further suggests a role for the linker region in influencing the luminal domain's conformation, corroborating results presented in Chapter 5.

The S296F mutation's influence on oligomerisation appears more moderate than that of the A414T mutant (Figure 6.11A). Cell-based assays are required to investigate the effect this mutation elicits to potentially promote tumorigenesis. However, an increased propensity to

oligomerise in the absence of unfolded proteins suggests that the mutant may promote an increased basal adaptive response similar to that identified with the A414T mutation (Lhomond et al., 2018) to give a selective advantage to cells containing the mutant in the harsh growth conditions that tumours experience (Kaufman et al., 2002; Chalmers et al., 2019).

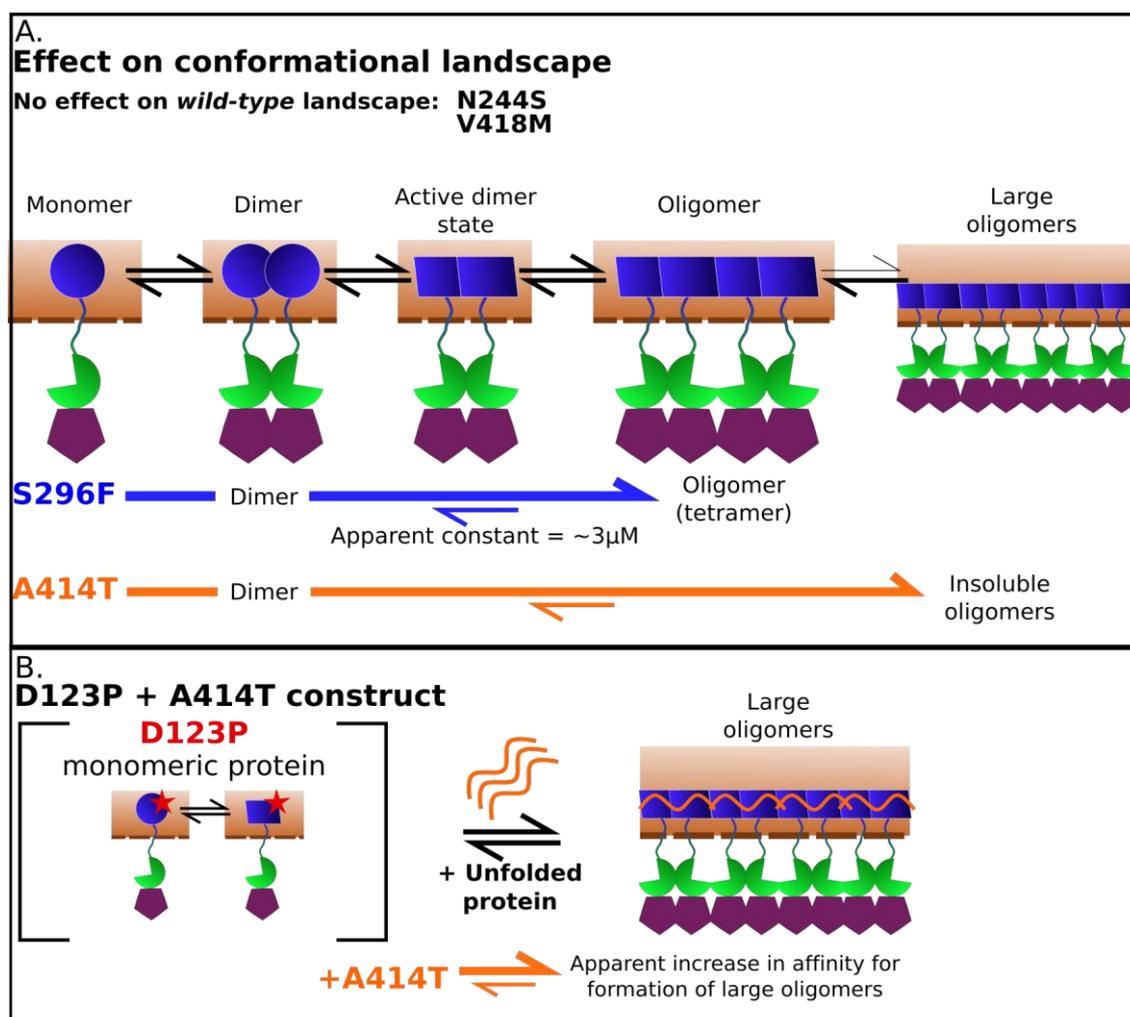


Figure 6.17

Summary of the effects of the luminal domain cancer-associated mutants on the protein's conformational landscape. A. In the absence of unfolded protein (peptide) the S296F mutation appears to promote formation of tetramers, and the protein exists as an equilibrium between these and dimers at the concentrations studied. The A414T mutation appears to have a strong effect of promoting formation of insoluble oligomers when recombinantly over expressed in *E. coli* culture. The N244S and V418M mutations have no significant effect detectable in the assays used. The S296F mutation still requires interaction with peptide to promote formation of larger and insoluble oligomers. B. The A414T + D123P double mutant appears to display an increased propensity to form oligomers upon addition of peptide than the D123P mutant, where less peptide is required to promote an increase in the measured OD400 values.

Further to this, the S296F mutation appears to stabilise formed disulphide bonds. Disulphide bonds formed by the luminal domain are suggested to increase IRE1 α sensitivity to endoplasmic reticulum stress (Eletto et al., 2014) and therefore stabilising formed disulphide bonds would also promote the adaptive response to potentially aid abhorrent cancer cell growth (Kaufman et al., 2002; Chalmers et al., 2019). It's not clear as to the mechanism by which S296F stabilises disulphide bonds. It's possible that the S296F mutation reduces solvent exposure of the disulphide bonds by adoption of a differential conformation, similar to examples in the literature where disulphide bond reduction is dependent on protein conformation (Iyer and Klee, 1973; Hong et al., 2009). It's also possible that apparent stabilisation of disulphide bonds is due to the increased oligomerisation of the S296F protein, promoting disulphide bond formation. It is not clear which of the luminal domain's cysteine residues are involved in this process (Figure 6.13C), further validation of disulphide bond formation in luminal domain oligomers is first required to fully understand the implications of disulphide bond stabilisation *in vivo* by the S296F mutation. Mutation of specific cysteine residues in the S296F construct may elucidate the culprit cysteine residues for this process.

One possible mechanism for the S296F mutation's role in promoting oligomerisation is through perturbation of interactions between the conserved anti-parallel β -sheet that the residue resides in, the identified oligomerisation interface and β -sandwich motif that is thought to be involved in propagating an oligomerisation-active conformation after unfolded protein binding (Karagoz et al., 2017). This would therefore suggest a role for the S296 residue and the antiparallel β -sheet in regulation of oligomerisation, further studies of the region may elucidate its potential regulatory role.

Interestingly, in the solved crystal structure (Zhou et al., 2006) the N244 residue also appears to interact with the conserved anti-parallel β -sheet motif from the β -sandwich domain, however, the N244S mutation doesn't affect the luminal domain's conformational landscape in the analysis employed in this chapter. Similarly, the V418M mutation occurs close to the impactful A414T mutation but did not elicit any influence on the luminal domain's conformational state. These mutations may therefore influence luminal domain activation in a different manner not assessed here, such as inhibiting the interactions of ERdj4, Hsp47, PDIA6 or other luminal domain binding partners (Amin-Wetzel et al., 2017; Eletto et al., 2014; Sepulveda et al., 2018; Sundaram et al., 2017). It's also possible that the effects of the mutations on equilibrium are too subtle to be detected by the assays used here or even that the mutations don't have an effect on protein function and are passenger mutations due to the DNA instability caused by tumour progression (Jeggio et al., 2016).

As discussed in Section 4.5 the luminal domain may be amenable to structural studies using solid-state NMR, cryo-electron microscopy as well as X-ray crystallography, in which case more information about the changes to the mutants' structural regions can be understood. As observed in Chapter 5, solution NMR presents a powerful technique to characterise the effects of the cancer-associated mutations on the protein's disordered regions. This is particularly useful for characterising the effects of luminal domain linker region mutations (A414T and V418M) but can also provide information about allosteric effects of all the mutants and may suggest further importance for the linker region in the luminal domain's conformational landscape. Therefore, in the next chapter the cancer-associated mutations will be assessed using solution NMR.

7. The effects of cancer-associated mutations on the luminal domain's intrinsically disordered regions

7.1. Introduction

As presented in Chapter 5, solution NMR study of the luminal domain can provide information about the protein's disordered regions and the partial assignments of the linker region allow for sequence specific information to be gained. NMR study of the cancer-associated mutants will therefore allow elucidation of the mutants' effects on the protein's disordered regions and how these changes may be important for the mutants' influence on protein function.

As discussed in Sections 1.6.1.3 and 5.5, NMR can report on the changes experienced by a protein in multiple ways. Each peak in the spectra presented here represents one HN group in the luminal domain's backbone, the position of the peak in the spectrum is dependent on the chemical environment of this HN group. Therefore, a change in position in the spectrum in a mutant construct represents a change in chemical environment. A peak's intensity reports on the dynamic processes experienced by that residue. As observed in Chapter 5, the luminal domain's loop region 2 reports on global multimeric changes of the protein (larger species cause a decrease in intensity due to line broadening). Further to this, the increased flexibility of the linker region increases local ns dynamics, producing increased peak intensities in more C-terminal sections of the linker region. In addition to these dynamic processes, heterogeneity in peak intensities reports on μ s-ms dynamics or adoption of secondary structure (Section 5.5).

The cancer-associated mutations may elicit different effects on the luminal domain's intrinsically disordered regions. Local changes to intrinsically disordered regions are due to residues proximal to the mutation site in the protein sequence experiencing changes to their chemical environments (~6 changes expected), this is therefore expected to be observed for only the linker region mutations, where these proximal residues will be observed. In contrast, the S296F and N244S mutations occur in structured regions that are not observable by solution NMR and therefore only long-range effects can be observed in the spectra obtained.

In this chapter, the effect of the cancer-associated mutations on the luminal domain's intrinsically disordered regions will be investigated.

7.2. Effects of the core domain cancer-associated mutations on luminal domain disordered regions

The core domain cancer-associated mutants (N244S and S296F) occur in the same area of the protein and have polar contacts in the solved crystal structure (Zhou et al., 2006). However, no functional effect of the N244S mutation was observed in Chapter 6, whereas the S296F mutant moderately promoted oligomerisation and protected intermolecular disulphide bonds. As discussed in Section 6.1, the mutated residues may act to bridge the conformational rearrangements of substrate binding to the protein's MHC-like groove (Karagoz et al., 2017). A model of the solved crystal structure including disordered regions (Zhou et al., 2006; Yang et al., 2015) suggests that the solution NMR observable loop region 2 is proximal to the core domain mutants (Figure 7.1), and therefore changes to this region may be observed in obtained spectra.

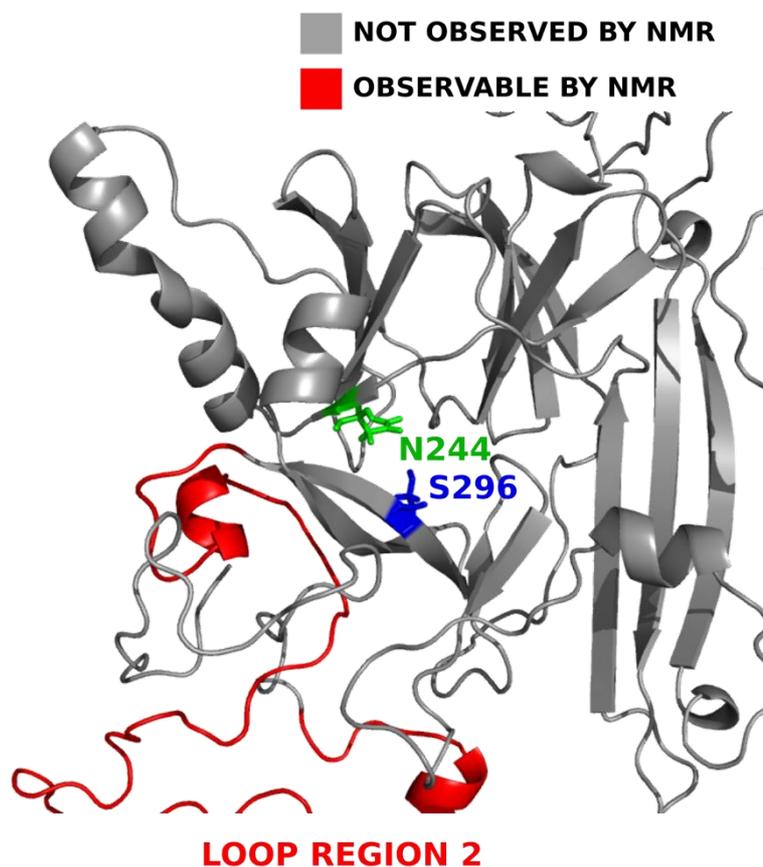


Figure 7.1

The N244 and S296 residues occur in structured regions of the luminal domain, proximal to loop region 2. Shown is a model based on the solved luminal domain crystal structure (PDB: 2HZ6) (Zhou et al., 2006) with loop region 2 modelled (Yang et al., 2015) and shown in red, this region is observed by solution NMR study of the protein. Figure prepared using PyMOL version 1.7.

7.2.1. The N244S mutant has no observable effects on the disordered regions of the protein

The N244S mutant's HN TROSY spectrum was compared to *wild-type* protein (Figure 7.2) and no apparent chemical shift perturbations or changes to dynamic processes in the protein's disordered regions were observed (Figures 7.2A and B). The N244S mutation therefore only affects the structured regions of the protein that aren't visible by solution NMR. Although no changes are observed, NMR analysis of the N244S mutant demonstrates the precision of solution NMR for study of any alterations to the chemical environments and dynamic processes experienced by residues of the other mutants.

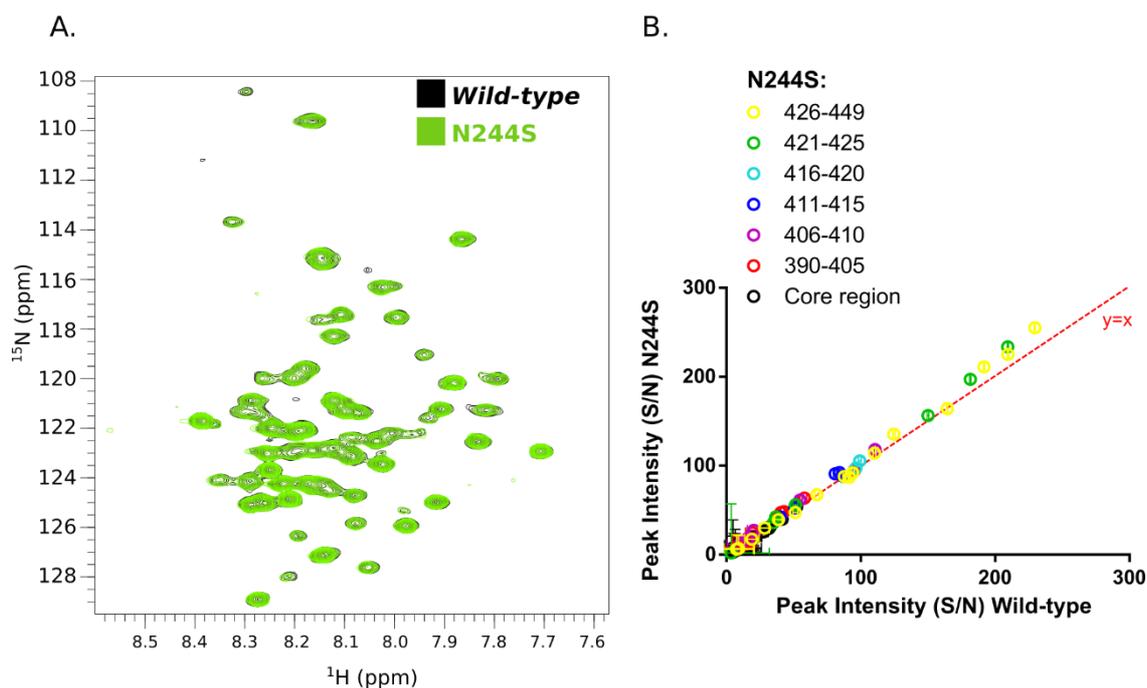


Figure 7.2

The N244S mutation does not cause any significant chemical shift perturbations or changes to dynamic processes in the disordered regions of the luminal domain that are visible by solution NMR. A. The HN TROSY spectrum of wild-type luminal domain protein overlaid with that of the N244S mutant at 50 μ M shows no significant perturbations in the protein (Section 2.7.1.2). B. Plotting the signal to noise (S/N) of peaks of the N244S construct to wild-type protein shows no significant changes in peak intensities and therefore dynamic processes.

7.2.2. The S296F mutation alters disordered region conformation

In contrast to the N244S mutant, the S296F mutation causes profound changes in the spectrum acquired (Figure 7.3). Two chemical shift perturbations are observed for residues in loop region 2 (Figure 7.3A and B), these peaks effectively disappear from the spectrum. This suggests that

the S296F mutation alters loop region 2's conformation. The fact that the S296F mutation promotes oligomerisation suggests that loop region 2 may be involved or affected in the regulation of oligomerisation, with the S296F mutation interfering with this process.

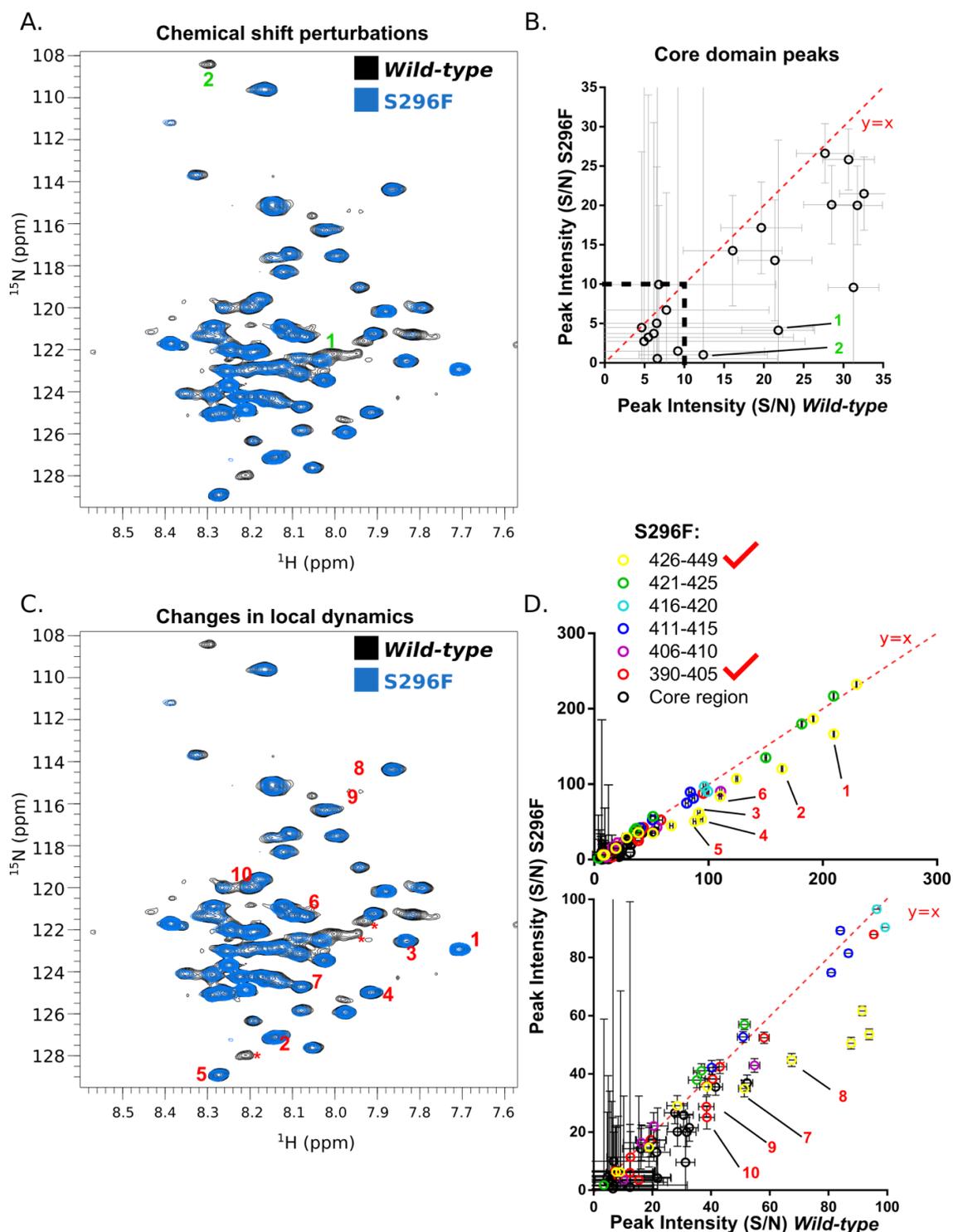


Figure 7.3

S296F causes chemical shift perturbations in loop region 2 and differential dynamic processes in the linker region. A. Overlaid HN TROSY spectra of the S296F mutant and wild-type protein

(Section 2.7.1.2, 50 μ M protein). Shown are the chemical shift perturbations caused by mutation as in panel B. B. Peak intensities for loop region 2 peaks of the S296F construct and wild-type protein are compared. Peaks with a S/N value of <10 in both constructs are ignored in the chemical shift perturbation analysis. Highlighted are the two loop region 2 residues that experience chemical shift perturbations. C. Overlaid HN TROSY spectra for the S296F mutant and wild-type luminal domain protein with the peaks that experience differential dynamic processes in the S296F mutant labelled with number corresponding to panel D. Shown also are three peaks labelled with '*' that are not significant by the analysis used but present a trend for reduced peak intensities in the 390-410 area of the linker region. D. Plotting the S/N of peaks of S296F to wild-type protein shows changes to peak intensities of residues from the linker region, the residues that experience these changes are highlighted and numbered, and can be observed in the spectra shown in panel C.

Additionally, multiple residues of the linker region appear to experience differential dynamic processes in the S296F construct (Figures 7.3C and D). Interestingly, these residues reside in the most C-terminal segment of the linker region (residues 426-429). However, it is not clear whether the changes observed are due to the mutant's increased oligomerisation causing differential conformations of this region or due to direct long-range interactions with the mutation site. There are also significant reductions in the intensities of two residues in the 390-410 segment of the linker region (peaks 9 and 10, Figure 7.3C) as well as three smaller changes in this area (marked with '*', changes occur to low intensity peaks (<15S/N), Figure 7.3C), therefore this segment of the linker region also experiences the effects of the S296F mutant.

To further understand the changes that the S296F mutation confers, the peak intensities of its partially assigned spectrum were compared to *wild-type* protein and the oligomerisation deficient WLLI-GSSG mutant in Figure 7.4. Comparison of loop region 2 peak intensities suggest that the S296F mutation promotes formation of larger species than *wild-type* protein due to an observable trend for reduced peak intensities in the mutant (Figure 7.4A; 0.71-fold decrease in peak intensity), this result is consistent with the size exclusion chromatography analysis carried out in Section 6.4.2.

Analysis of the other segments of the linker region suggest no difference for residues 390-425 of the linker region (Figure 7.4B and C). Additionally, although changes to dynamic processes are observed in residues 426-449 (Figure 7.4D, Section 5.6.2) for both mutants, there doesn't appear to be a clear trend. The effects observed in this region may be due to this most C-terminal section of the linker region interacting with the oligomerisation interface and therefore being affected by these mutants that reside close to there. It must be considered that the most C-terminal region of

the linker region consists of 23 residues and there may be more than one allosteric site within the assigned section. To validate and determine the specificities of the role of this section of the linker region in the luminal domain's oligomeric state further investigation is required.

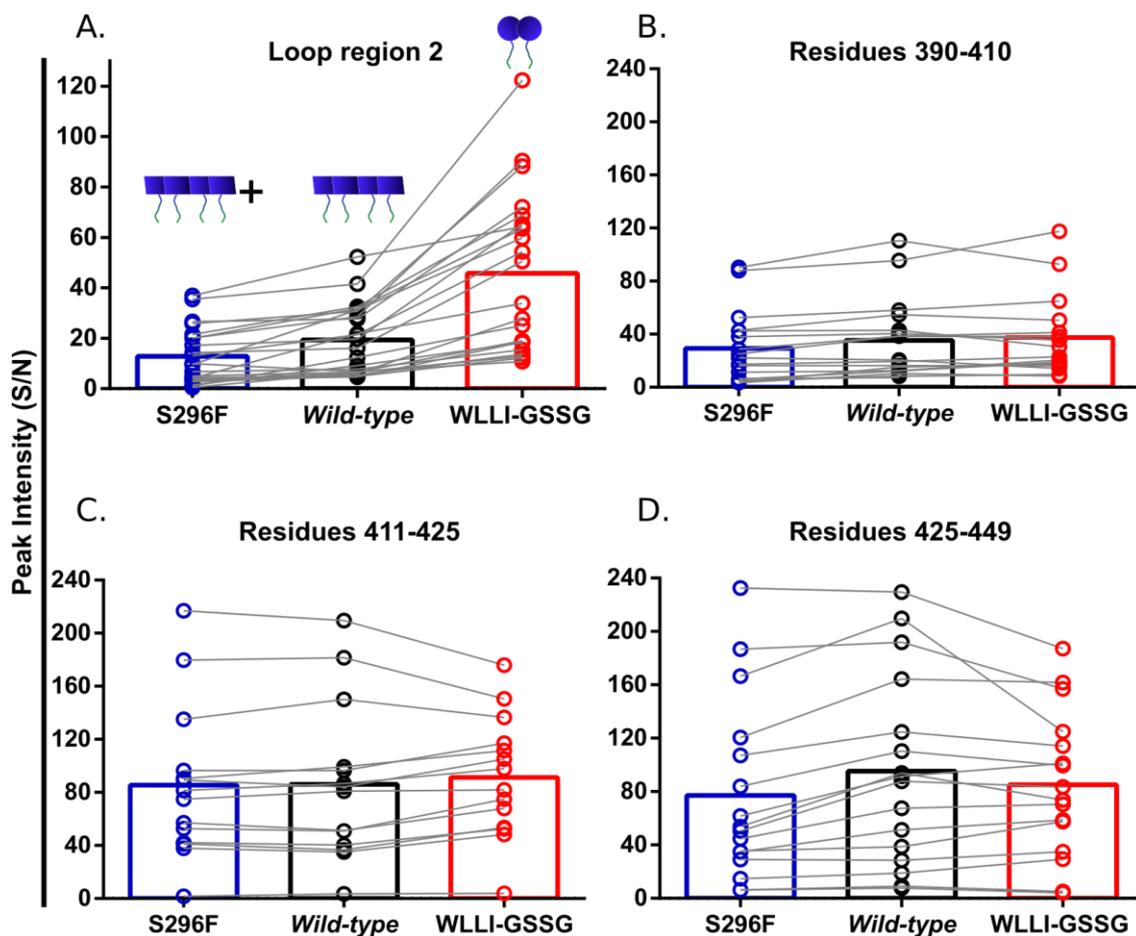


Figure 7.4

Comparison of the peak intensities for S296F mutant residues with wild-type and WLLI-GSSG protein in different disordered regions. A. The core domain (loop region 2) residues' peak intensities are plotted for each construct. Lines are drawn between the same residues in each construct. There is a trend between multimeric size and peak intensity, where larger multimers reduce the peak intensities of these residues (0.71-fold decrease in core domain peak intensity; but not significant by unpaired *t*-test, $p=0.061$). B. Comparison of the peaks of the most N-terminal segment of the linker region (residues 390-410) suggests no trend between the size of multimers and peak intensity. C. Comparison of peaks of residues 411-425 suggests no trend between multimer size and peak intensity. D. Comparison between the residues of the most distal segment of the linker (residues 426-449) for each construct suggests that some residues of this segment exhibit altered peak intensities in each mutant, but no clear trend in intensity is observed.

7.3. How the linker region mutations affect the luminal domain's disordered regions

The A414 and V418 residues are likely to have peaks visible in the obtained spectra and therefore more information about their effects on the linker region and the long-range communication of the protein's disordered regions can be garnered. Changes to the chemical environments of adjacent residues in the protein's sequence to the mutation sites will cause chemical shift perturbations and are considered local changes. Whereas, mutation-induced conformational changes in the linker region will be observed in more distant regions of the linker region.

7.3.1. The A414T mutation only causes local perturbations in the linker region

The HN TROSY spectrum for the A414T + D123P construct was compared to the D123P mutant's spectrum (Figure 7.5) to observe changes that the A414T mutation elicits on the disordered regions of the protein. Only four chemical shift perturbations were observed, all of which occur in residues 411-420, local to the A414 mutation site. No significant changes in peak intensities were observed suggesting no changes to linker region dynamic processes were made upon mutation.

The D123P mutant is thought to exist as monomeric protein in the solution NMR conditions used with a small amount of dimeric protein. As observed in Section 5.6.1, the D123P mutation exhibits a differential linker region conformation, affecting residues 390-415. It's therefore possible that this linker region conformation affects the A414T mutation's environment and therefore allosteric effects of the A414T mutation that promote oligomerisation do not occur and so only local changes are observed for the A414T+D123P construct. This hypothesis is in agreement with Section 6.4.3, which suggests that upon oligomerisation, and therefore an expected different linker region conformation, more similar to that observed for *wild-type* protein, the A414T mutation has its functional effect to promote oligomerisation restored. Another possibility is that the A414T mutation elicits its effect through interaction with protein's structured regions, which are not observable by solution NMR. Solution NMR study of the A414T mutant coupled with the WLLI-GSSG mutation (to allow soluble overexpression in *E. coli*) may provide more information about the mutation's effects. Additionally, the V418M cancer-associated mutation occurs near to the A414T mutation and does not include the D123P mutation in its construct. Solution NMR study of the V418M mutation may therefore elucidate long-range effects of the 411-420 area of the linker region in the oligomeric protein.

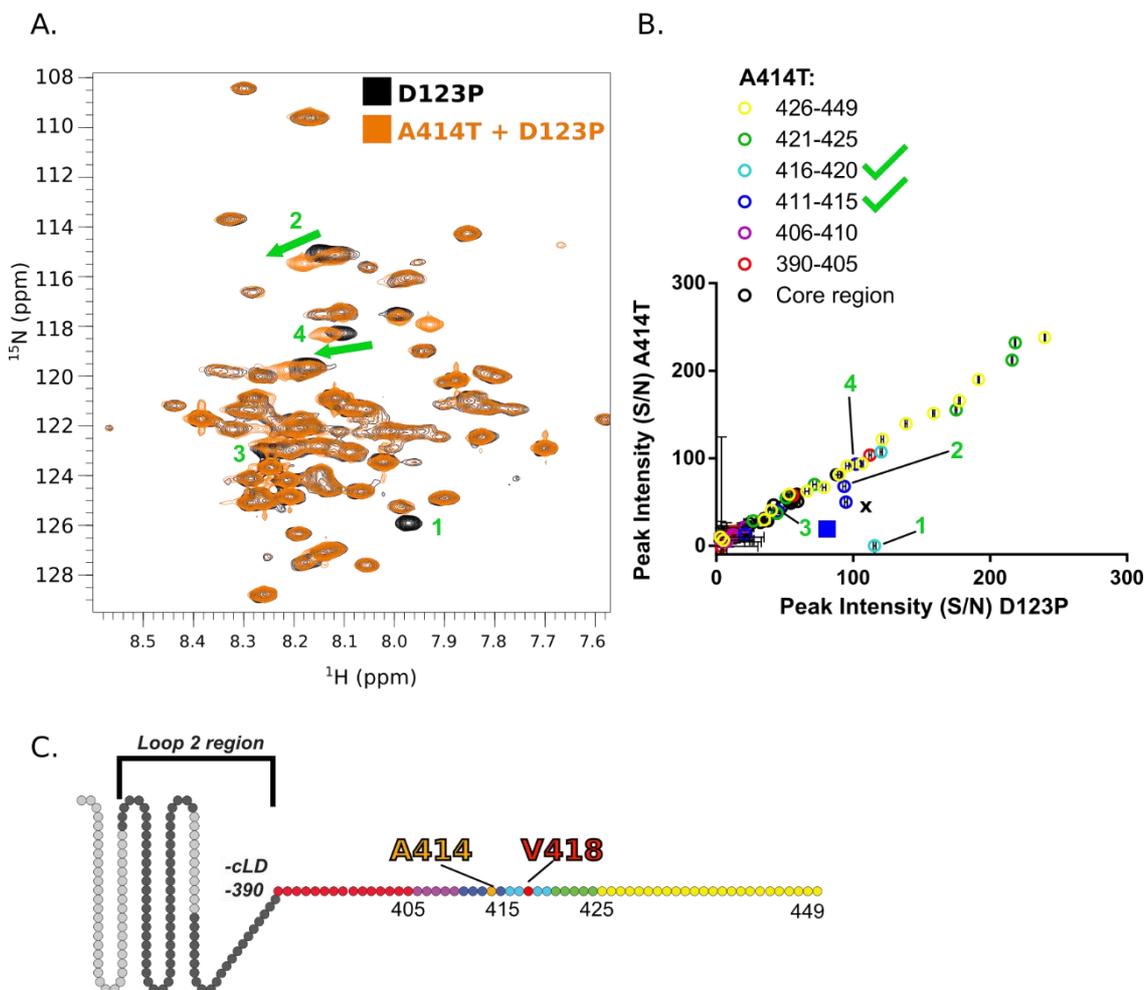


Figure 7.5

The A414T mutation causes significant chemical shift perturbations in local residues of the linker region, as observed by a HN TROSY solution NMR experiment. A. The HN TROSY spectra of the D123P mutant luminal domain overlaid with that of the A414T + D123P double mutant (Section 2.7.1.2, 50 μ M protein). The A414T+D123P spectrum shows four significant chemical shifts, all of which are from local regions of the protein (segment 411-420) to the A414T mutation site. B. Plotting the S/N of peaks of the A414T + D123P spectrum and comparing them to D123P protein shows some significant changes in peak intensity which are due to chemical shift perturbations. Labelled are the corresponding peaks that experience chemical shift perturbations as in panel A. Labelled with an 'x' is an apparent change in peak intensity caused by the chemical shift perturbation of a neighbouring unassigned peak. C. Schematic of the luminal domain's loop region 2 and linker region, displaying the cancer-associated mutations of the linker region.

7.3.2. The V418M mutation causes long-range perturbations of the linker region

The V418M mutant causes six significant chemical shift perturbations when compared to the *wild-type* protein's spectrum (Figures 7.6A and B). However, four of the perturbations are from

residues 410-425 in the linker region and are therefore considered local changes, close to the V418 mutation site.

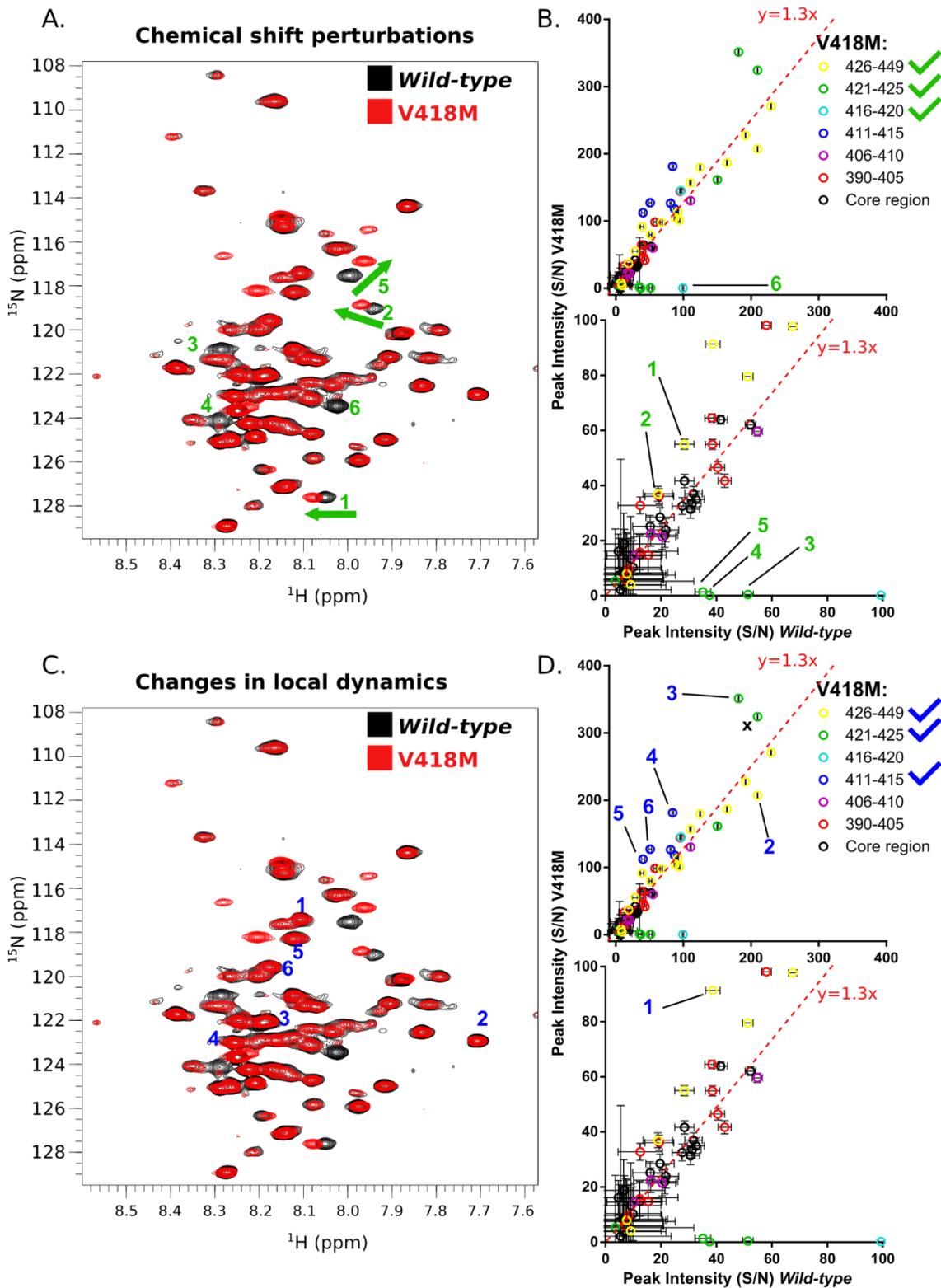


Figure 7.6

The V418M mutation causes numerous chemical shift perturbations and changes in dynamic processes. A. The HN TROSY spectrum of wild-type luminal domain in black is overlaid with that of the V418M mutant in red (Section 2.7.1.2, 50 μ M protein). Chemical shift perturbations are labelled with numbers. Chemical shift perturbations 1 and 2 are from the most C-terminal segment of the linker region (residues 426-449). It is worthy of note that some new peaks are also observable in the spectrum for the V418M mutant, these are likely shifted peaks, however, it is not possible to definitively assign them to peaks in the wild-type spectrum. B. Plot of the S/N of peaks of the V418M mutant compared to wild-type protein. Peaks that experience chemical shift perturbations are labelled with numbers corresponding to the labels in panel A. C. Overlaid HN TROSY spectra for the wild-type and V418M constructs. Labelled are the peaks that experience changes in intensity in the mutant construct, therefore suggesting these residues experience differential dynamic processes. D. A plot of the S/N of peaks in the wild-type and V418M constructs, as in panel B. Here, the residues that experience differential dynamics upon mutation are labelled with numbers corresponding to the labels in panel C. One residue of the 421-425 region is labelled with an 'x' as its increased peak intensity appears to be from an overlapping peak that experienced a chemical shift perturbation in the mutant construct (thought to be from chemical shift 4) and is therefore not considered a change in peak intensity resulting from differential dynamic processes. Here, reference lines for $y=1.3x$ are shown as the V418M mutation has an overall change in peak intensity, this may be due to a different sample concentration being used. Corrections for this have been made in analysis of peak intensities.

Interestingly, two significant chemical shift perturbations are observed in the C-terminal segment of the linker region, suggesting long-range communication between the 416-420 site that V418 resides in and residues 426-449 (Figure 7.6A and B, peaks 1 and 2). This may suggest that the V418M mutation alters long-range interactions within the linker region and linker region conformation. Unexpectedly, similar chemical shift perturbations are observed in the WLLI-GSSG mutant also (Figure 7.7), suggesting that the V418M mutation promotes a dimer-like conformation of this section of the linker region. This may indicate an interaction between the middle section and most C-terminal section of the linker region that is disrupted by the V418M mutant or by mutation of the oligomerisation interface (WLLI-GSSG). It is possible that residues 426-449 of the linker region communicate with other motifs of the linker region (specifically residue 418) as well as the area around the oligomerisation interface (from changes observed in WLLI-GSSG and S296F constructs) to regulate luminal domain oligomerisation.

Changes in peak intensities are also observed for residues of the most C-terminal segment of the linker region (residues 426-449), validating that the V418 mutation site and these residues communicate. Additionally, three residues experience different dynamic processes between

residues 411 and 415 and one residue in the 421-425 segment of the linker region. This indicates that the V418M mutation may interfere with the conformations adopted by these local segments of the linker region. Interestingly, the three 411-415 residues affected surround the A414T mutation site, suggesting that although no changes to their dynamic processes or interactions are observed by A414T mutation in the D123P construct, these residues may have interactions with other areas of the linker region in oligomeric protein.

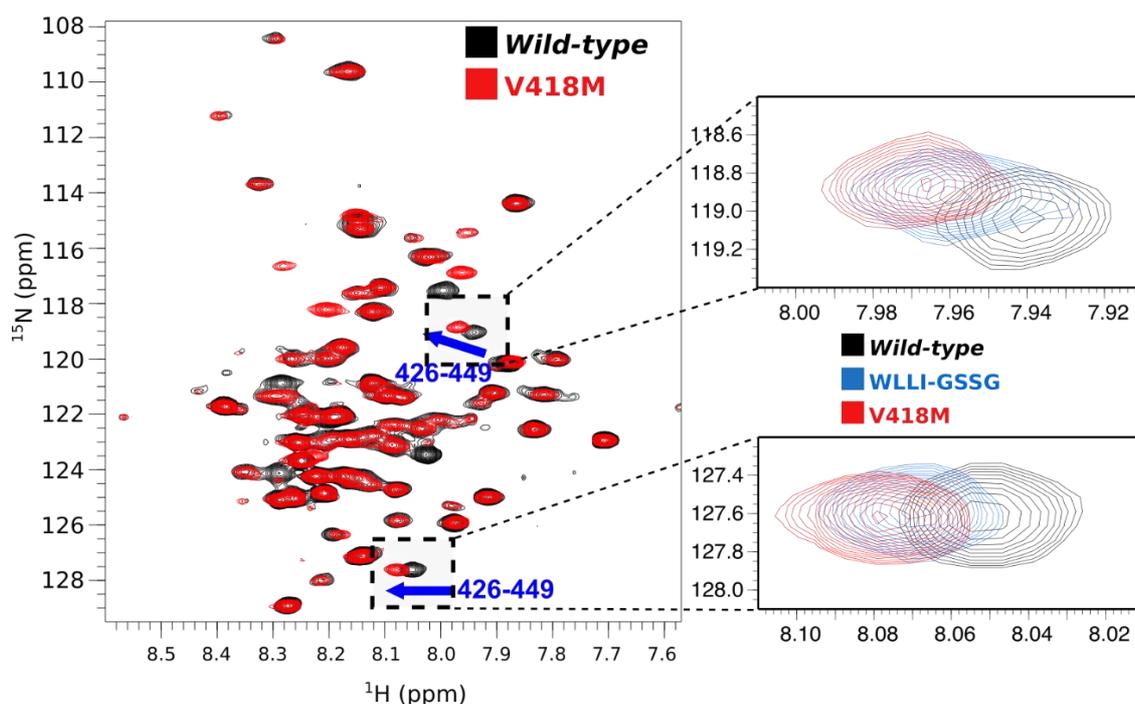


Figure 7.7

Similar chemical shift perturbations are observed in the V418M mutation and the WLLI-GSSG oligomerisation-deficient mutation. Shown are the V418M and wild-type protein HN TROSY spectra overlaid. Expanded are the two chemical shift perturbations observed for the 426-449 segment of the linker region and compared to the perturbations observed in these residues for the WLLI-GSSG mutant (Section 5.6.2). This suggests that the V418M mutation affects the conformation of these regions in a similar manner as the WLLI-GSSG mutant, suggesting an effect of the V418M mutation on oligomerisation, although no changes were determined in the functional assays carried out in Chapter 6.

7.4. Conclusion

In this chapter, solution NMR study was used to determine alterations to the conformations adopted by the luminal domain's disordered regions caused by cancer-associated mutations. In line with results presented in Chapter 6, where the S296F and A414T mutants altered the luminal

domain's conformational landscape, NMR study of the mutants has revealed long-range communication between their mutation sites, suggesting an allosteric network between the oligomerisation interface, residues in the 411-420 segment and the most C-terminal section of the linker region. The changes in the mutants' solution NMR spectra are summarised in Figure 7.8.

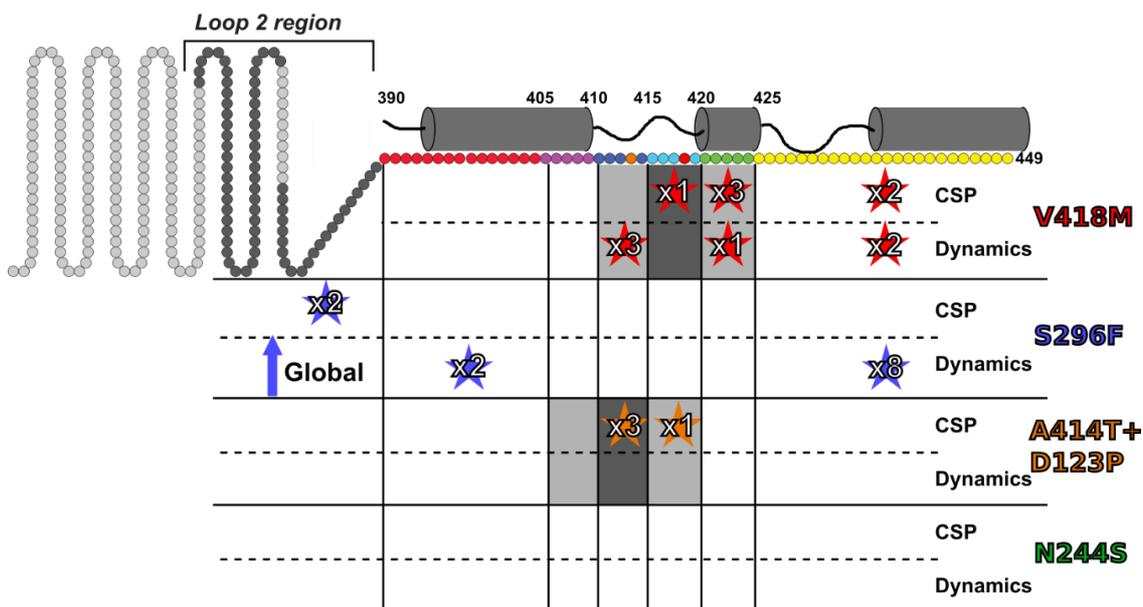


Figure 7.8

Summary of chemical shift perturbations and changes in dynamic processes from the introduction of cancer-associated mutations to the luminal domain identified by solution NMR study. The schematic shows the different assigned sections of the protein's disordered regions and the changes from each mutation. Included are the regions that are predicted to adopt secondary structure for reference (Drozdetskiy et al., 2015; Shen, Y. et al., 2014; Thevenet et al., 2012). Dark grey symbolises the five-residue region where the mutation occurs and in light grey are the adjacent segments in the luminal domain's disordered regions, suggesting local effects.

Both functional assays in Chapter 6 and solution NMR analysis of the N244S mutant's disordered regions have not identified any influence of the mutation on the luminal domain. The mutant was part of a previous study for the effects of cancer-associated mutations on IRE1 α signalling, using cell-based assays but was removed for technical reasons, and therefore there is no data available for its effect (Xue et al., 2011). However, the results here suggest that any conformational changes incurred through this mutation are to the structured regions of the protein and therefore future research of the mutant's effect can be directed towards these regions and for interactions not observed in the analysis carried out in Chapter 6.

Solution NMR study of the S296F mutant validates findings in Chapter 6 that the mutation favours oligomerisation (Figure 7.4A). The S296F mutation also causes chemical shift perturbations in loop region 2, suggesting there may be an allosteric mechanism by which loop region 2 is involved in luminal domain oligomerisation. It's also possible that the chemical shift perturbations observed in loop region 2 are indicative of an altered conformation that stabilises disulphide bonds formed by the C332 residue that resides there, as observed in Section 6.4.2.2, but further validation is required.

The A414T mutation causes the most profound functional changes but only local chemical shift perturbations are observed by solution NMR study of the D123P + A414T double mutant (Figure 7.8). However, study of the V418M mutant presents changes to the dynamic processes experienced by residues surrounding the A414T mutation site (Figure 7.6C and D). This suggests that the residues around A414 have interactions with the V418 residue or are involved in similar processes and therefore may also be influenced by the most C-terminal region (linker region C, Figure 7.9A) of the linker region (as the V418M mutation affects residues here; Figure 7.9C). As the A414T mutant construct also contains the D123P mutation that causes differential linker region conformations (Section 5.6.1), the interactions of A414T with other segments of the linker region that promote oligomerisation may be inhibited in the double mutant and therefore not observed by solution NMR.

In addition to the changes in dynamics of residues in section 411-415 of the linker region, the V418M mutation causes two chemical shift perturbations and affects two residues' dynamic processes in linker region C (Figures 7.8 and 7.9C). This suggests that there are long-range interactions in the linker region B and C. Additionally, the two chemical shift perturbations are similar to perturbations observed in the WLLI-GSSG mutation (Section 5.6.2), implying that the V418M mutation's long-range interactions with this region are similar to that of dimerised protein as opposed to oligomerised *wild-type* protein. Interestingly, peak broadening of linker region C is observed with the S296F mutation (Figures 7.3 and 7.9C), further suggesting long-range interactions linker region C, with not only other parts of the linker region but also the core domain of the protein which may be part of an allosteric mechanism to regulate oligomerisation. Interestingly, linker region C has been implicated as having a regulatory role in activation of the luminal domain by lipid bilayer stress (Volmer et al., 2013; Kono et al., 2017). However, the effect on the globular region of the luminal domain from lipid-bilayer stress is not yet clear. Additionally, this area of the linker region is involved in Sec61-dependent regulation of activation (Sundaram et al., 2017) and therefore the interactions between the three sites identified may be involved in this regulatory interaction.

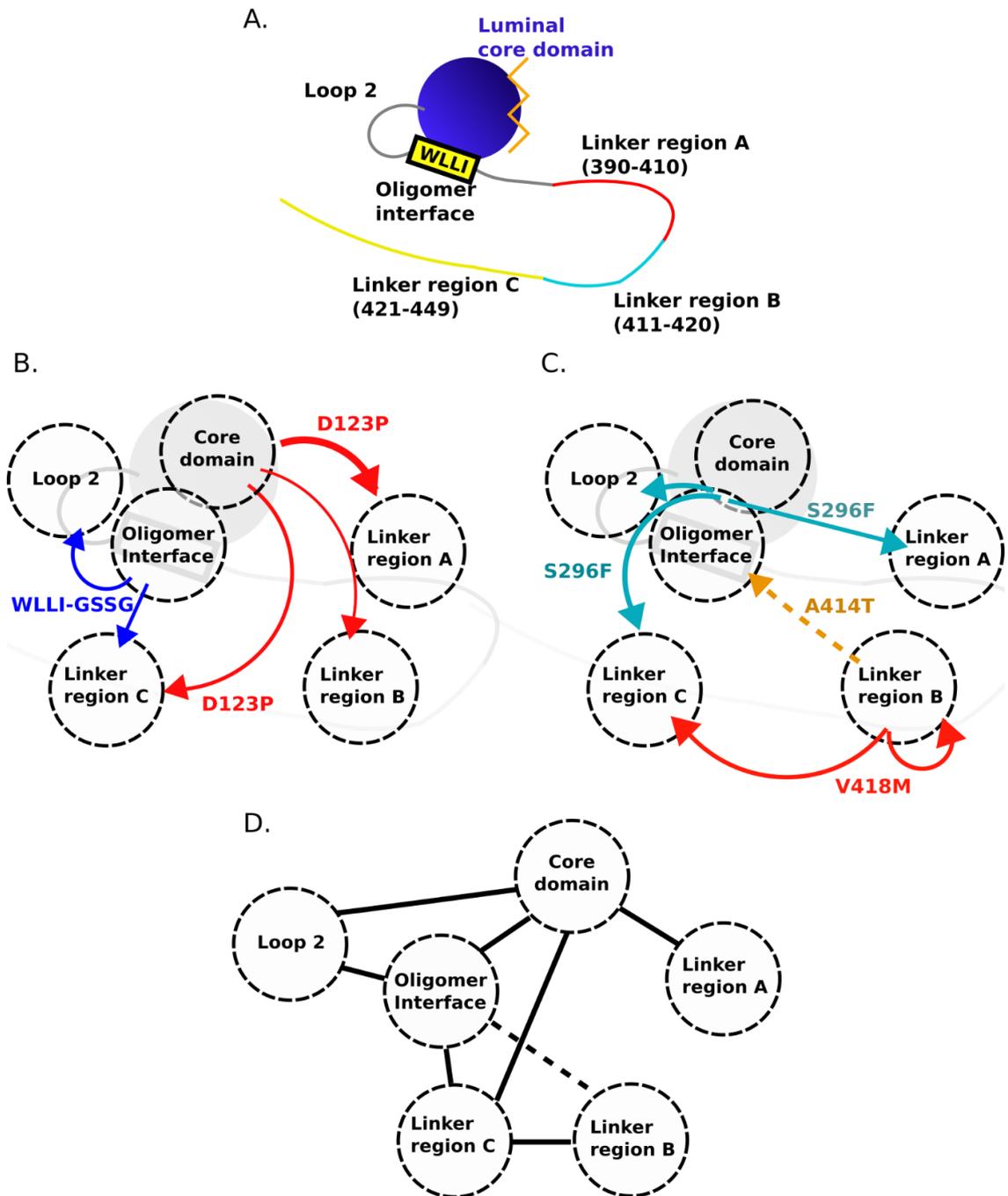


Figure 7.9

The identified allosteric network of the luminal domain. A. Schematic of the luminal domain with key features labelled. B. Connections between different nodes of the proposed allosteric network observed in Chapter 5. C. Connections between different nodes of the proposed allosteric network observed in Chapters 6 and 7. A dotted line is used for A414T as this is an observed effect of the mutation to promote oligomerisation. D. Overall proposed allosteric network of the luminal domain.

Solution NMR study of cancer-associated mutants of IRE1 α 's luminal domain has therefore provided information about an allosteric network between the protein's oligomerisation interface, residues 411-420 (linker region B) and linker region C. Data from the analysis of the cancer-associated mutants can be used in combination with information about the conformation of the linker region from Chapter 5 to better understand long-range communication between the protein's regions and their potential allosteric roles in regulation of activation, in this way an allosteric network in the luminal domain can be defined (Figure 7.9D). The luminal domain's intrinsically disordered regions were previously largely uncharacterised, but the data presented here has suggested a potentially important role for them in regulating the luminal domain, the proposed allosteric network can now be further investigated for its role in luminal domain activation.

8. Conclusions and future perspectives

IRE1 α 's activation cascade is a complex and multi-step process which consists of multiple conformations adopted by the protein's cytoplasmic and luminal domains (Korennykh, A. and Walter, 2012; Walter and Ron, 2011). In response to endoplasmic reticulum stress the protein's endoplasmic reticulum luminal domain is first activated before the signal is propagated to promote cytoplasmic domain activation. The luminal domain responds to fluctuating stress levels in the endoplasmic reticulum, through stress-related factors influencing its conformation such as unfolded protein binding to promote oligomerisation (Karagoz et al., 2017) and BiP binding to dimeric luminal domain protein to repress its activation (Amin-Wetzel et al., 2017). However, the complexity of luminal domain activation is not fully understood causing discrepancies between the two models for luminal domain IRE1 α activation (Section 1.4.5) (Preissler and Ron, 2019). In addition to this, the effect of cancer-associated mutations on the domain are not understood and their occurrence in structural regions distal to those considered to be functionally important hints that the protein may contain allosteric regulation mechanisms. The aim of the research in this thesis was therefore to unravel and investigate the luminal domain's activation mechanism in response to stress and its potential allosteric regulatory sites.

The luminal domain's conformational landscape was explored in Chapter 4, giving evidence for its existence as a conformational equilibrium that is influenced by mutations (Sections 4.2 and 4.3.2.3), peptide binding (Section 4.3) and the inclusion of BiP (Section 4.4); a balance that likely dictates cellular responses of the protein. The effect of increasing peptide concentration in dictating oligomer size was also further explored, suggesting a mechanism for IRE1 α 's luminal domain to initiate a proportionate response to stress levels (Section 4.3.2). Formation of large insoluble oligomers of the luminal domain were characterised and observed for the first time (Section 4.5.1), suggesting a mechanism to initiate the large scale IRE1 α clustering observed *in vivo* (Li, H. et al., 2010). Additionally, a stable interaction of BiP with luminal domain oligomers was revealed, alongside characterisation of an uncommon chaperone and clathrin-like interaction (Section 4.4) (Bocking et al., 2011; Clerico et al., 2015) to de-oligomerise the luminal domain, leading to the presentation of a novel model for the termination of unfolded protein stress activated IRE1 α (Figure 8.1A).

Further investigation to validate this model by *in vivo* study is required as well as determining whether BiP's novel interaction involves the same interactions as observed in Amin-Wetzel et al. (2017), therefore the effect of ERdj4 on the process can be assessed. However, the transient nature of this interaction, requiring a much-reduced amount of BiP in comparison to IRE1 α (Section 4.4,

Figure 4.14) is likely why BiP's interaction with IRE1 α under stressed conditions is observed to a lesser degree (Bertolotti et al., 2000; Oikawa et al., 2009) and may represent a differential mechanism to the ERdj4 mediated interaction. Further to this, the fate of the bound unfolded protein is yet to be identified after BiP-dependent de-assembly to smaller oligomers. However, this represents a mechanism by which IRE1 α may respond to changes in endoplasmic reticulum stress after being severely activated by binding to unfolded proteins (Karagoz et al., 2017) and even a mechanism to further bridge the direct-binding and chaperone repression models currently in the field (Amin-Wetzell et al., 2017; Preissler and Ron, 2019), therefore adding to our understand of IRE1 α 's response to stress.

The findings presented here also provide preliminary data for in-depth conformational and structural study of the domain's peptide- and BiP-bound oligomeric states by cryo-electron microscopy and solid-state NMR. Further study by these techniques can provide BiP's interaction site with the protein, the mode of unfolded protein binding to the luminal domain and the downstream conformational changes that these events incur in the protein's structure. Use of solid-state NMR has already identified potential BiP binding sites in the luminal domain's disordered regions (Section 4.5.2) which now require assignment to specific residues.

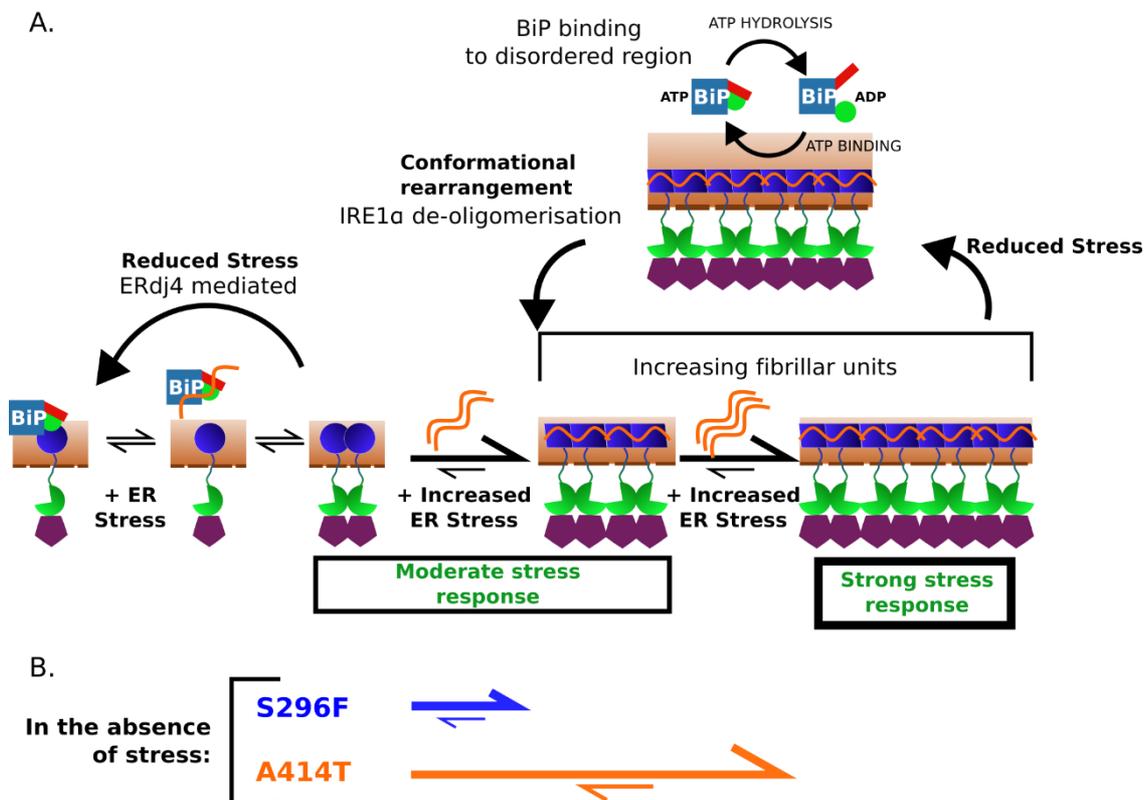


Figure 8.1

Summary of findings for the functional activation of the luminal domain and its cancer-associated mutants. A. Hypothesised model for IRE1 α activation in response to unfolded protein stress and its subsequent termination through interaction of luminal domain oligomers with BiP. B. The effects of the S296F and A414T mutations. The S296F mutation may cause a more moderate cellular response than the A414T mutant and requires validation in cell-based assays.

In Chapter 6, a mechanistic analysis of four luminal domain cancer-associated mutations was carried out leading to the characterisation of the oligomerisation driving A414T (Section 6.3) and oligomerisation biasing S296F (Section 6.4.2) mutations' effects (Figure 8.1B). Study of the mutations adds further support to the model of the luminal domain's conformational landscape being influenced to promote IRE1 α activation. However, it is not yet clear whether the moderately influential S296F mutation drives tumorigenesis like the A414T mutation or requires compounding changes to the cell, such as p53 mutation, similar to the previously studied P336L cancer-associated mutation (Lhomond et al., 2018). Therefore, cellular study of the mutants is required and may give further understanding about how the luminal domain's multimeric state influences the downstream unfolded protein response. Additionally, *in vivo* study of the N244S and V418M mutants' cellular effects can be targeted by the findings presented here, that suggest the mutations' likely interference with interactions of the luminal domain's binding partners (Amin-Wetzel et al., 2017; Sundaram et al., 2017; Sepulveda et al., 2018; Eletto et al., 2014). The effect of the A414T and S296F mutations appear to be through obstruction of previously unrecognised allosteric pathways; the conserved anti-parallel β -sheet close to functionally important oligomerisation motifs (Karagoz et al., 2017) may present a novel regulatory mechanism for oligomerisation and therapeutic target as well as the luminal domain's linker region in the case of the A414T mutation.

Due to evidence for the luminal linker region's and other disordered regions' potential importance in regulating IRE1 α function (Oikawa et al., 2009; Lhomond et al., 2018; Kono et al., 2017; Halbleib et al., 2017; Sundaram et al., 2017), solution NMR was used to study how these were affected by different states in the luminal domain's conformational landscape. Results suggest that monomeric, dimeric and oligomeric states of the protein promote differential linker region conformations (Chapter 5), suggesting that the linker region is influenced by the luminal domain's state through a novel allosteric network. Investigation of the cancer-associated mutants' disordered regions (Chapter 7) provides further understanding about the protein's allosteric network (Section 7.4, Figure 7.9D) and evidence for its role in regulation.

Firstly, analysis of the WLLI-GSSG and S296F mutations, that inhibit and promote oligomerisation respectively revealed long-range changes in linker region C and loop region 2,

suggesting the role of these regions in oligomerisation. The highly influential A414T mutation (Sections 6.3 and 6.4.3) (Lhomond et al., 2018) didn't appear to alter the linker region's interactions, suggesting that its effect is abrogated by the differential linker region conformation adopted by monomeric protein in the A414T+D123P double mutant, further suggesting that linker region conformation may act to regulate the protein's activation. It is also possible that the residue is involved in interactions with the structured regions of the protein, future study of the A414T+WLLI-GSSG construct will therefore provide further information about the mutant's effects. Interestingly, analysis of the V418M mutation in linker region B elucidated alterations of the oligomerisation sensitive region C and in residues surrounding the A414T mutation site, suggesting A414 may be communicate with this site, similar to the S296F and WLLI-GSSG mutations.

Results from analysis of the differential conformational states and cancer-associated mutants' effects on the protein's disordered regions therefore suggest an allosteric network involving the different regions of the linker region, core domain, loop region 2 and oligomerisation interface (Section 7.4). Now that a network between these regions has been observed, their roles in regulating luminal domain activation and precise interactions can be further investigated. Briefly described are models for the potential influence of these regions on IRE1 α activation.

The signal of unfolded protein binding to the luminal domain's MHC-like groove is thought to be propagated through the β -sandwich and $\alpha\beta$ -helix motifs to the protein's oligomerisation interface (residues 357-361 WLLI motif; Figure 8.2A) (Karagoz et al., 2017). A conserved anti-parallel β -sheet motif containing the S296 residue resides between these functionally important motifs and the oligomerisation interface and is close to loop region 2. The first model presented proposes that the anti-parallel β -sheet element is involved in the regulation of oligomerisation interface formation (Figure 8.2B), whereby S296F disrupts regulatory contacts to mimic signal propagation from unfolded protein binding to promote oligomerisation and rearrangement of loop region 2 and linker region C. Our results suggest that the WLLI residues are not the only component of the oligomerisation interface, as oligomerisation is observed in the WLLI-GSSG mutant (Section 4.3.2.3), and therefore rearrangements of loop region 2 and linker region C may also be elements of the formed interface.

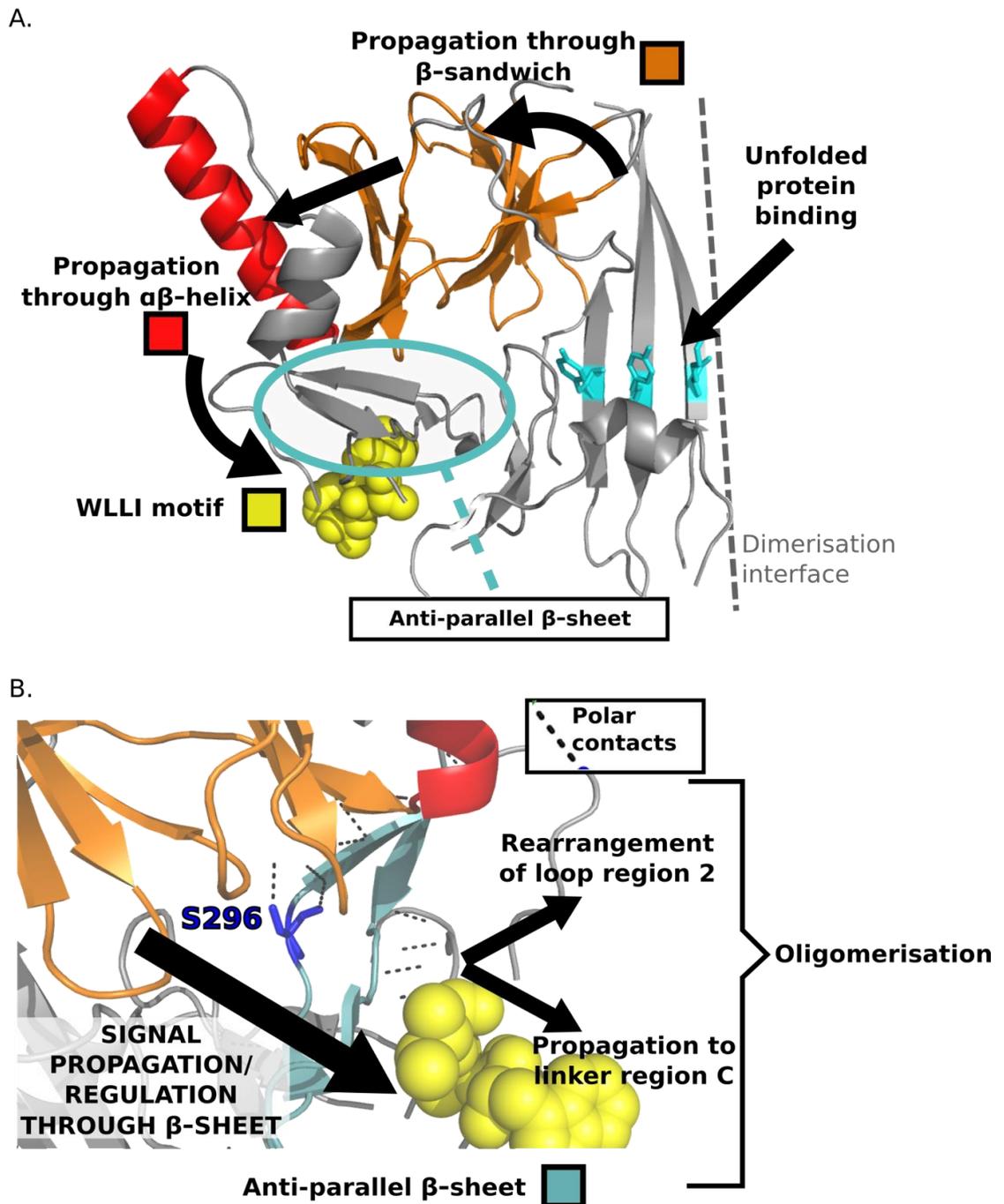


Figure 8.2

Proposed regulation of the oligomerisation-active conformation of the protein by the conserved anti-parallel β -sheet element, shown with the solved crystal structure model of the luminal domain (PDB: 2HZ6) (Zhou et al., 2006). A. Proposed propagation of conformational rearrangements from unfolded protein binding to trigger oligomerisation of the luminal domain, therefore formation of the oligomerisation-active state (Karagoz et al., 2017). B. The S296F mutation resides in the conserved anti-parallel β -sheet that has interactions with regions of importance in propagating the unfolded protein binding signal and the oligomerisation interface (WLLI). This region is surrounded by loop region 2. The hypothesis presented suggests that the anti-parallel

β -sheet has a role in regulating the oligomerisation interface, where the S296F mutation unhinges regulation to promote oligomerisation through the WLLI motif, the effect of which also affects loop region 2 and linker region C, which aren't resolved in the crystal structure but may be components in the oligomerisation interface. Figure prepared using PyMOL version 1.7.

The differential conformations of the linker region observed in different luminal domain states form the basis of the second model posed here (Figure 8.3), whereby the linker region's conformations act to regulate luminal domain activation. Potential roles of linker region A's conformational shift upon dimerisation have been suggested in Section 5.7, potentially interfering with ERdj4 or BiP's interaction with dimeric protein (Amin-Wetzel et al., 2017) or representing the BiP bound and repressed conformation of the protein due to the importance of the 390-408 residue region in BiP regulation, described in a previous study of the protein (Oikawa et al., 2009).

Further to this, the communication between linker region B and C (Section 7.3.2), and linker region C's differential conformation in oligomerised protein (Section 5.6.2) may also represent conformations that regulate luminal domain activation or interactions (Figure 8.3B). Interestingly, residues 434-443 in linker region C have been identified as the binding site for the regulatory Sec61 translocon channel (Sundaram et al., 2017; Plumb et al., 2015). Sec61 is suggested to bind to IRE1 α in low stress conditions to inhibit oligomerisation, but in high stress conditions, Sec61 interaction no longer inhibits luminal domain activation (Sundaram et al., 2017), this may be due to linker regions B + C's 'Conformation 2' (Figure 8.3B) interfering with this process. Additionally, linker region C's different conformation may be involved in sensing lipid bilayer stress, as region C has been implicated in the mechanism previously (Kono et al., 2017; Halbleib et al., 2017). One study observed that inclusion of a dimerisation disfavoured mutation prevents a lipid bilayer stress response (Kitai et al., 2013) suggesting that there may be a role for the core luminal domain and linker region C in sensing lipid bilayer stress, and therefore the different linker region conformations may couple lipid bilayer stress sensing and the core luminal domain. However the allosteric network's potential role in this process requires further research, especially as one study observed no requirement for the core luminal domain in the lipid bilayer stress response (Volmer et al., 2013).

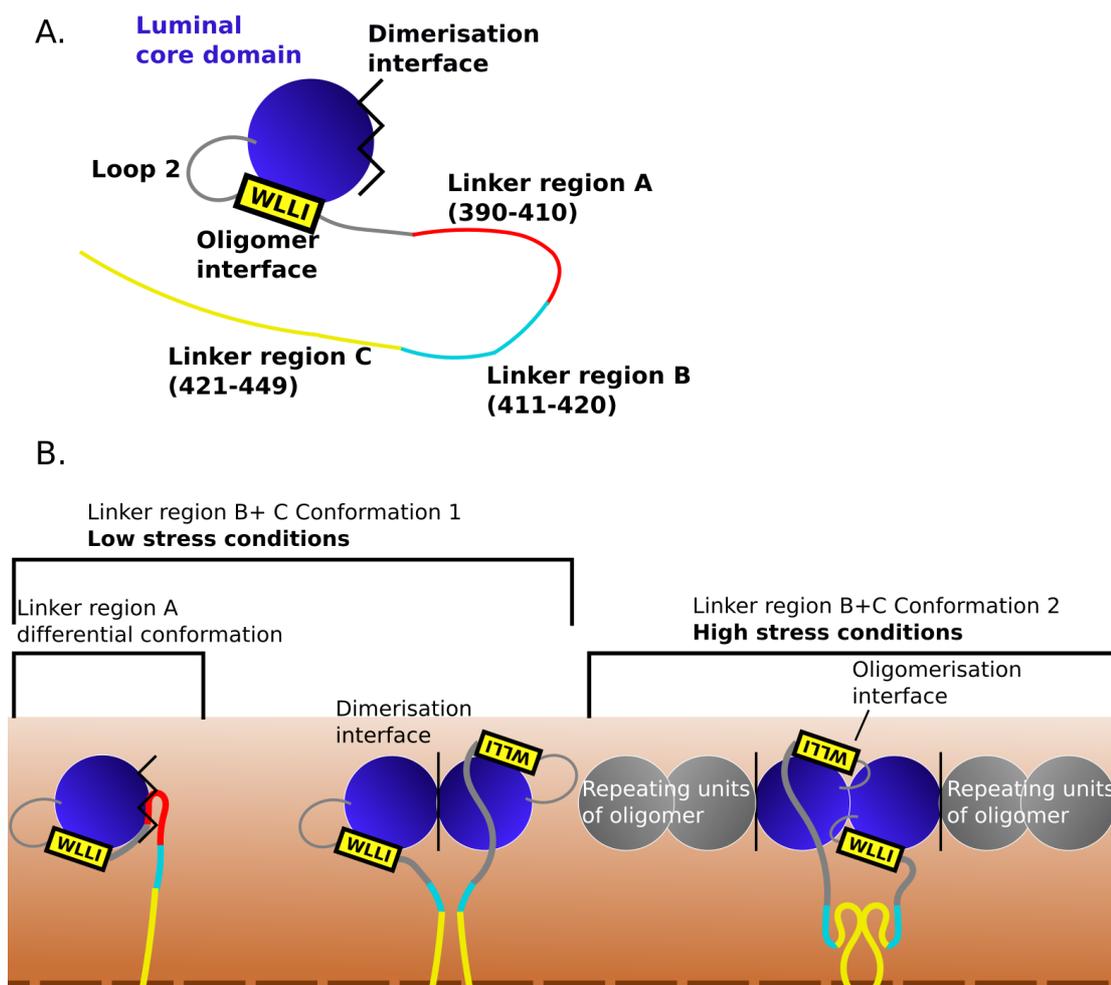


Figure 8.3

Proposed model for the role of the different linker region conformations identified. A. Schematic summary of the luminal domain's regions studied here. B. Proposed model for differential conformations of the linker region in different luminal domain states. Linker region A is altered in the monomer to dimer transition. Linker region B + C 'Conformation 1' in dimeric protein and 'Conformation 2' in oligomeric protein may represent a regulatory role of the allosteric network identified here, possibly through the linker region's interaction with Sec61 to inhibit activation (Sundaram et al., 2017) or in lipid bilayer stress sensing (Halbleib et al., 2017; Kono et al., 2017).

The two models posed here are not mutually exclusive and its possible that elements of each model may be valid and therefore further *in vitro*, *in vivo* and structural assays are required, in part by using the optimised and pilot-tested techniques utilised in this thesis. Further to this, techniques such as fluorescence resonance energy transfer (FRET) (Haas, 2012; LeBlanc et al., 2018), paramagnetic relaxation enhancement (PRE) (Silvestre-Ryan et al., 2013) and hydrogen-deuterium exchange (Faustino et al., 2019) can give further understanding about the observed allosteric network, particularly of the linker region nodes.

Interestingly, the proposed allosteric mechanisms and novel sites of interest, such as residues implicated in the BiP binding site occur in the luminal domain's disordered regions. The targeting of disordered regions of proteins is a newly progressing field in therapeutics, made desirable by the increased number of disordered regions perturbed in diseases/disorders, described as the D² concept (Uversky et al., 2008). Our understanding of optimal drug design and approaches to influence disordered regions is ever improving but requires understanding of the interactions within, and conformational ensembles of these regions to identify potential druggable sites (Ruan et al., 2019). Therefore, the results presented here form the basis for investigation of the conformational ensembles of the identified allosteric network to aid in their targeting to influence the luminal domain's activation to give a desired outcome.

IRE1 α is the most conserved endoplasmic reticulum stress sensor, its activation appears to be influenced by many factors to give the appropriate response to stress (Amin-Wetzel et al., 2017; Sundaram et al., 2017; Kono et al., 2017; Karagoz et al., 2017). Therefore, a multifaceted approach is required to appreciate the protein's complexity, from studying the protein's complex interactions collectively *in vivo* to a reductionist investigation into specific mechanisms. Here, novel methods, mechanisms and allosteric networks of the protein have been presented to further our understanding of the protein's regulation and response and how it can be influenced for future targeting by therapeutic approaches for the numerous disease states it is implicated in.

9. Appendices

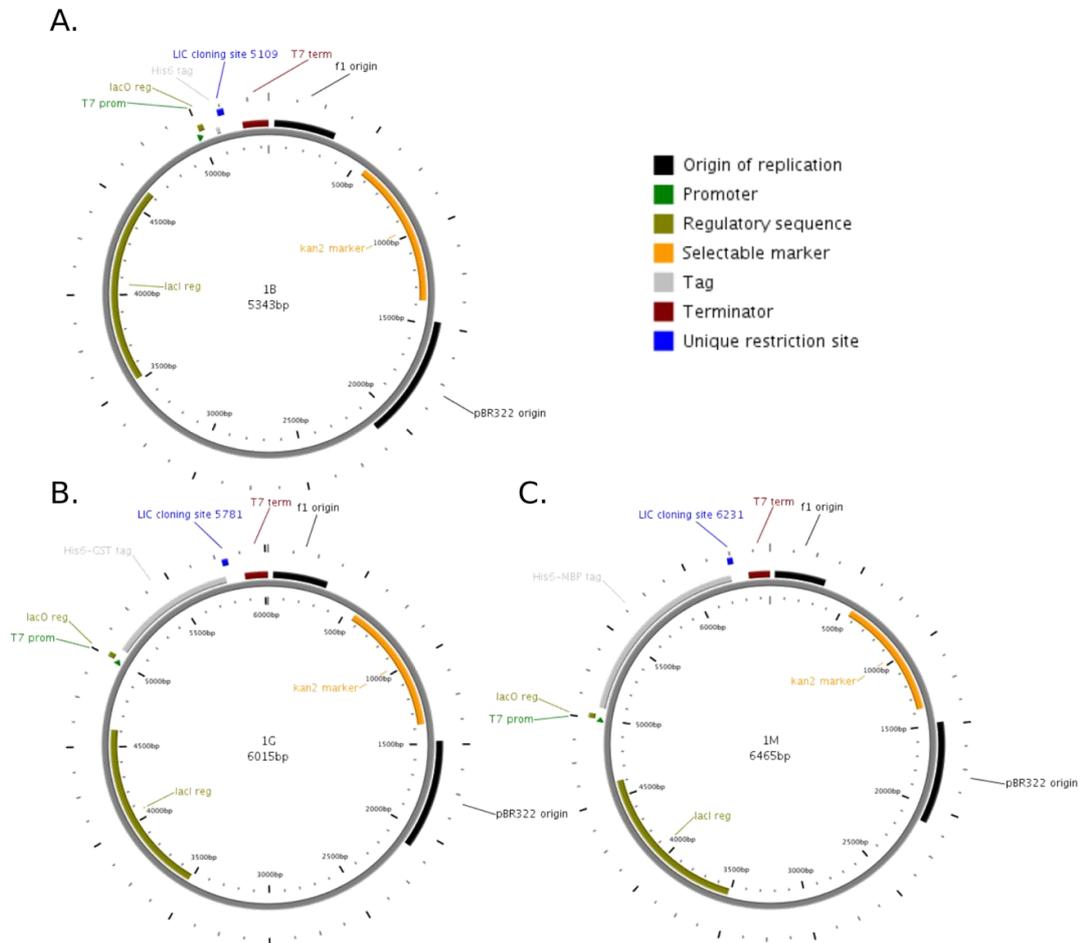


Figure 9.1

Plasmid maps of the 1B (A.), 1G (B.) and 1M (C.) plasmid used for subcloning of the cytoplasmic domain construct. Figures adapted from www.addgene.org. Plasmids 1B, 1G and 1M were gifts from Scott Gradia (1B: Addgene plasmid #29653; <http://n2t/addgene:29653> ; RRID:Addgene_29653, 1G: Addgene plasmid #29655; <http://n2t/addgene:29655> ; RRID:Addgene_29655, 1M: Addgene plasmid #29656; <http://n2t/addgene:29656> ; RRID:Addgene_29656).

Cytoplasmic domain expressed construct

GSSHHHHHHENLYFQSNAGSSPSLEQDDGDEETSVVIVGKISFCPKDVLGHGAEGTIVYRGMFDNRDV
 AVKRILPECFSFADREVQLLRESDEHPNVIRYFCTEKDRQFOYIAIELCAATLQEYVEQKDF AHLGLE
 PITLLQOTTSGLAHLHSLNIVHRDLKPHNILISMPNAHGKIKAMISDFGLCKKLAVGRHSFSRRSGVP
 GTEGWIAPEMLS EDCKENPTYTVDIF SAGCVFYYVISEGSHPF GKSLQRQANILLGACSLDCLHPEKH
 EDVIARELIEKMIAMDPQKRPSAKHV LKHPFFWSLEKQLQFFQDVSDRIEKESLDGPIVKQLERGGRA
 VVKMDWRENITVPLQTDLRKFR TYKGGSVRDLLRAMRNKKH HYRELPAEVRETLGSLPDDFVCYFTSR
 FPHLLAHTYRAMELCSHERLFQPYFHEPPEPQPPVTPDAL

■ 6His-tag

■ TEV-cleavage site

Figure 9.2

The recombinantly expressed cytoplasmic domain construct's protein sequence including N-terminal methionine cleavage (residues 548-977).

Cytoplasmic domain expressed construct

MGSSHHHHHHENLYFQSNAGSSPSLEQDDGDEETSVVIVGKISFCPKDVLGHGAEGTIVYRGMFDNRD
 VAVKRILPECFSFADREVQLLRESDEHPNVIRYFCTEKDRQFOYIAIELCAATLQEYVEQKDF AHLGLI
 EPITLLQOTTSGLAHLHSLNIVHRDLKPHNILISMPNAHGKIKAMISDFGLCKKLAVGRHSFSRRSGV
 PGTEGWIAPEMLS EDCKENPTYTVDIF SAGCVFYYVISEGSHPF GKSLQRQANILLGACSLDCLHPEK
 HEDVIARELIEKMIAMDPQKRPSAKHV LKHPFFWSLEKQLQFFQDVSDRIEKESLDGPIVKQLERGGRA
 AVVKMDWRENITVPLQTDLRKFR TYKGGSVRDLLRAMRNKKH HYRELPAEVRETLGSLPDDFVCYFTS
 RFPHLLAHTYRAMELCSHERLFQPYFHEPPEPQPPVTPDAL

Sequencing Primer 1 result

GATCGGCGGCTGAAC TGTGCTGCCACTGCCAGCTTCTTGCAGAGGCCAAAGTCGGAGATCATGGCCTGATCTTGCCGTGTGCA
 TTGGGCATGGATATGAGGATGTTGTGTGGCTT TAGGTCCTCTGTGAACGATGTTGAGGGAGTGGAGGTGGGCCAGGCCGAGGTGGT
 CTGCTGCAGCAAGGTGATGGCTCCAGGCCGAGATGCGCAAAGTCTTCTGCTCCACATACTCTTGCAGGGTGGCTGCACACAGCT
 CGATGGCAATGFACTGGAATTGCCGGTCTTCTCCGTGCAGAAGTAGCGGATCACGTTCCGGTGTCTATCCGATCTCGCAACAGC
 TGGACCTCACGGTCTGCGAAGCTAAAACACTCGGGGAGGATCCTCTTACCGCCACGTCGCGGTTGTCAAACATGCCCCGTACAC
 AATTGTGCCCTCAGCTCCATGGCCAGGACATCCTTGGGACAGAAGAAATTTCCCACTATCACCACGCTGGTTTCTCATCTC
 CATCGTCTTGTCCAGGGAGGGCTGCTTGCATTGGATTGGAAGTACAGTTTTTCATGGTGATGGTGATGGTGAGAAGAACCATG
 GTATATCTCCTTCTTAAAGTTAAACAAAATTATTTCTAGAGGGGAATGTTATCCGCTCACAATTTCCCTATAGTGAGTCGTATTA
 ATTTCCGGGATCGAGATCTCGATCTCTACGCCGACGCATCGTGGCCGGCATCACCGGCCACAGGTGCGGTTGCTGGCGNCT
 ATATCGCCGACATCACCGATGGGAAGATCGGGCTCGCCACTTCGGGNTCATGAGCGCTNTGTNT

Sequencing Primer 2 result

CCNNCAANNCTCATATCCATGCCCAATGCACACGGCAAGATCAAGCCATGATCTCCGACTTTGGCCTCTGCAAGAAGCTGGCA
 GTGGGCAGACACAGTTTCAGCCCGGATCTGGGGTGCCTGGCACAGAAGGCTGGATCGTCCAGAGATGCTGAGCGAAGACTGTAA
 GGAGAACCCTACCTACAGGTGGACATCTTTTCTGCAGGCTGCGTCTTTTACTACGTAATCTCTGAGGGCAGCCACCCCTTTGGCA
 AGTCCCTGCAGCGGCAGGCCAATCCTCCTGGGTGCCTGCAGCCTTACTGCTTGCACCCAGAGAAGCACGAAGACGTCATTGCA
 CGTGAATTGATAGAGAAGATGATTGCGATGGATCCTCAGAAACGCCCTCAGCGAAGCACGTCGCTCAAACACCCGTTCTTCTGGAG
 CCTAGAGAAGCAGCTCCAGTTCTTCCAGGACGTGAGCGACAGAATAGAAAAGGAATCCCTGGATGGCCCGATCGTGAAGCAGTTAG
 AGAGAGCGGGGAGAGCCGTGGTGAAGATGGACTGGCGGGAGAACATCACTGTCCCCCTCCAGACAGACTGCGTAAATTCAGGACC
 TATAAAGGTGGTTCTGTGTCAGAGATCTCCTCCGAGCCATGAGAAATAAGAAGCACCACTACCGGGAGCTGCCTGCAGAGGTGCGGGA
 GACGCTGGGGTCCCTCCCGACGACTTCGTGTGCTACTTACGCTCTCGCTTCCCCACCTCCTCGCACACACCTACCGGCCATGG
 AGCTGTGCAGCCACGAGACTCTTCCAGCCCTACTACTTCCAGAGCCCCAGAGCCCCAGCCCCAGTACTCCAGACGCCCC
 TAATAACATTGGAAGTGATAACGGATCCGAATTCGAGCGCCGTCGACAAGCNTGCGGCCGCACTCGAGCACCACCACCACCA
 CTGANNNCCGGCTGCTAACAAAGCCGAAAGNAGCTGAGTTGGNTGCTGCCACCGCTNANCAATNACTAGCNNTAACCCNTGGG
 GNNNNNANGNNCTTGNNGGGTTTTTGTCTGAAAGGNNNTATTATNCNNNTNNAATGGNNNGCCNNTANCGNCNCATTANCN
 CNNNGNNNNNNNCNCNCNA

Figure 9.3

The cytoplasmic domain construct is shown with sequencing data from the forward and reverse primers used. Together the primers give complete coverage of the protein construct. The region

of the protein construct covered by each sequencing result is represented by red for the forward primer and green for the reverse.

Luminal domain expressed construct

GSS **HHHHH** **ENLYFQS** NASTVTLPETLLFVSTLDGSLHAVSKRTGSIKWTLKEDPVLQVPTHVEEPAF
 LPDPNDGSLYTLGSKNNEGLTKLPFTIPELVQASPCRSSDGILYMGKKQDIWYVIDLLTGEKQOTLSS
 AFADSLCPSTSLLYLGRTEYTIMYDTKTRELRWNATYFDYAASLPEDDVDYKMSHFVSNNGDGLVVTV
 DSESGDVLWIQNYASPVVAFYVWQREGLRKVMHINVAVETLRYLTFMSGEVGRITKWKYPPFKETEA
 SKLTPTLYVVGKYSTSLYASPSMVHEGVAVVPRGSTLPLEGPQTDGVTIGDKGECVITPSTDVKFDPG
 LKSKNKLNYLRNYWLLIGHHETPLSASTKMLERFPNNLPKHRENV **IPADSEKKSFEVINLVDQTS**
EN **APT**TVSRDVEEKPAHAPARPEAPVDSMLKDMATIILS

■ 6His-tag ■ Linker region
 ■ TEV-cleavage site

Figure 9.4

The recombinantly expressed luminal domain construct's protein sequence including N-terminal methionine cleavage (residues 24-450). Highlighted in grey is the linker region of the protein that is removed in the core luminal domain (cLD) construct.

Luminal domain expressed construct

GSSHHHHHHENLYFQSNASTVTLPELTLFVSTLDGSLHAVSKRTGSIKWTLKEDPVLQVPTHVEEPAF
 LPDPNDGSLYTLGSKNNEGLTKLPFTIPELVQASPCRSSDGILYMGKKQDIWYVIDLLTGKQOTLSS
 AFADSLCPSTSLLYLGRTEYITITMYDTKTRELNRWNATYFDYAASLPEDDVYKMSHFVSNGLVVTV
 DSESGDVLWIQNYASPVVAFYVWQREGLRKMVHINVAVETLRYLTFMMSGEVGRITKWYPPFKETEAK
 SKLTPTLVYGKYSTSLYASPSMVHEGVAVVPRGSTLPLLEGPQTDGVTIGDKGECVITPSTDVKFDPG
 LKSKNKLNYLRNYWLLIGHHETPLSASTKMLERFPNNLPKHRENVIPADSEKKSFEEVINLVDQTSN
 APTTVSRDVEEKPAHAPARPEAPVDSMLKDMATIILS

Sequencing Primer 1 result

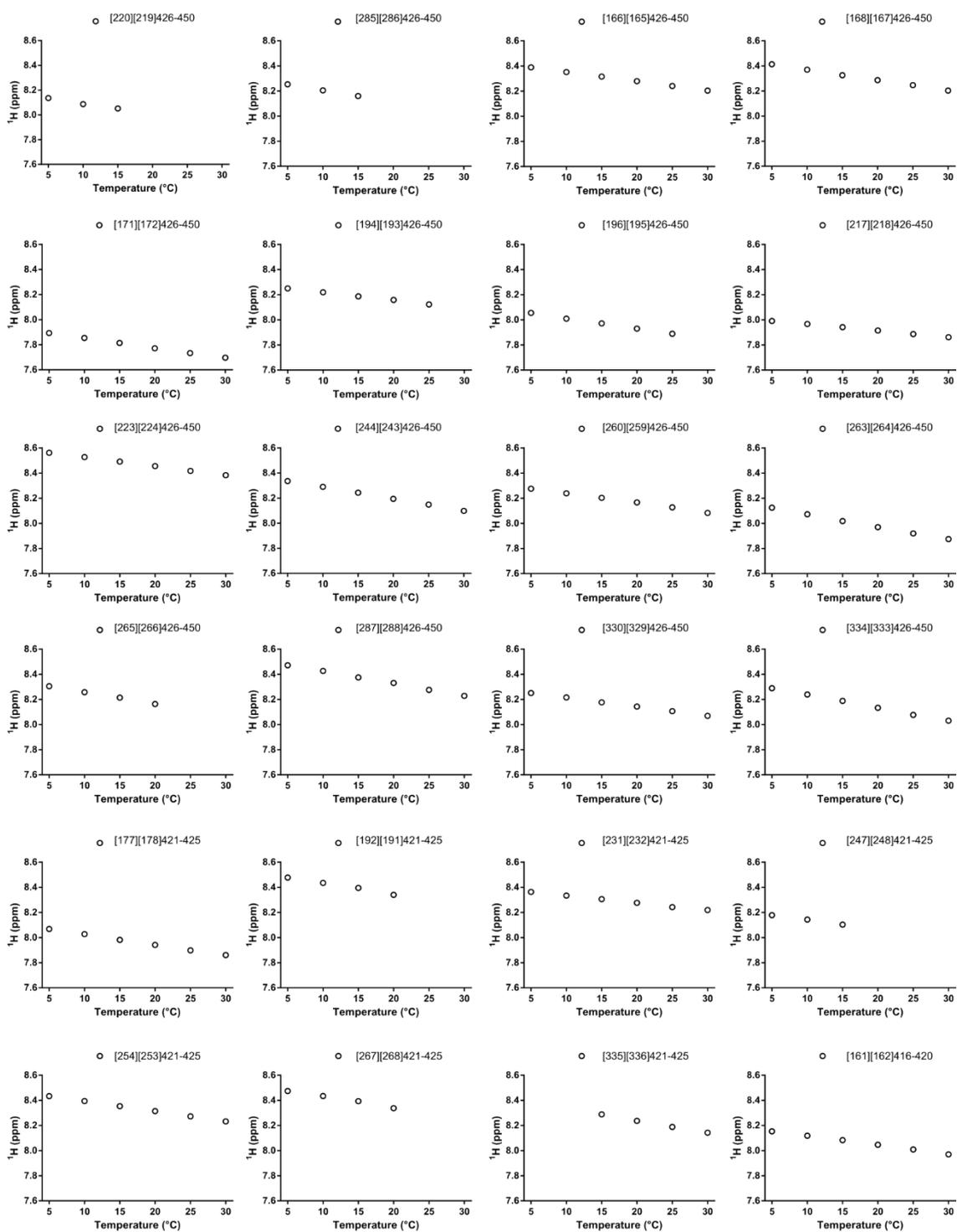
TTTGTTTAACTTTAAGAAGGAGATATACCATGGGTTCTTCTCACCATCACCATCACCATGAAAACCTGTACTTCCAATCCAATGCA
 AGCACAGTGACGCTTCTCGAAACCTTGTGTGTGTGTCAACGCTGGATGGAAGTTTGCATGCTGTCAGCAAGAGGACAGGCTCAAT
 CAAATGGACTTTAAAAGAAGATCCAGTCTGCAGGTCCTCAACACATGTGGAAGAGCCTGCCTTTCTCCAGATCCTAATGATGGCA
 GCCTGTATACGCTTGAAGCAAGAATAATGAAGCCTGACGAACTTCTTTTACCATCCCAGAATTGGTGCAGGCATCCCATGC
 CGAAGTTCAGATGGAATCCTCTACATGGGTAAAAAGCAGGACATCTGGTATGTTATTGACCTCTGACCGGAGAGAAGCAGCAGAC
 TTTGTATCGGCCTTTGCAGATAGTCTCTGCCATCAACCTCTTCTGTATCTTGGGCGAACAGAATACACCATCACCATGTACG
 ACACCAAACCCGAGAGCTCCGGTGAATGCCACCTACTTTGACTATGCGGCCTCACTGCCTGAGGACGACGTGGACTACAAGATG
 TCCCCTTTGTGTCCAATGGTATGGGCTGGTGGTACTGTGGACAGTGAATCTGGGGACGCTCTGTGGATCCAAAACCTACGCCTC
 CCTGTGGTGGCCTTTTATGTCTGGCAGCGGGAGGGTCTGAGGAAGGTGATGCANATCA

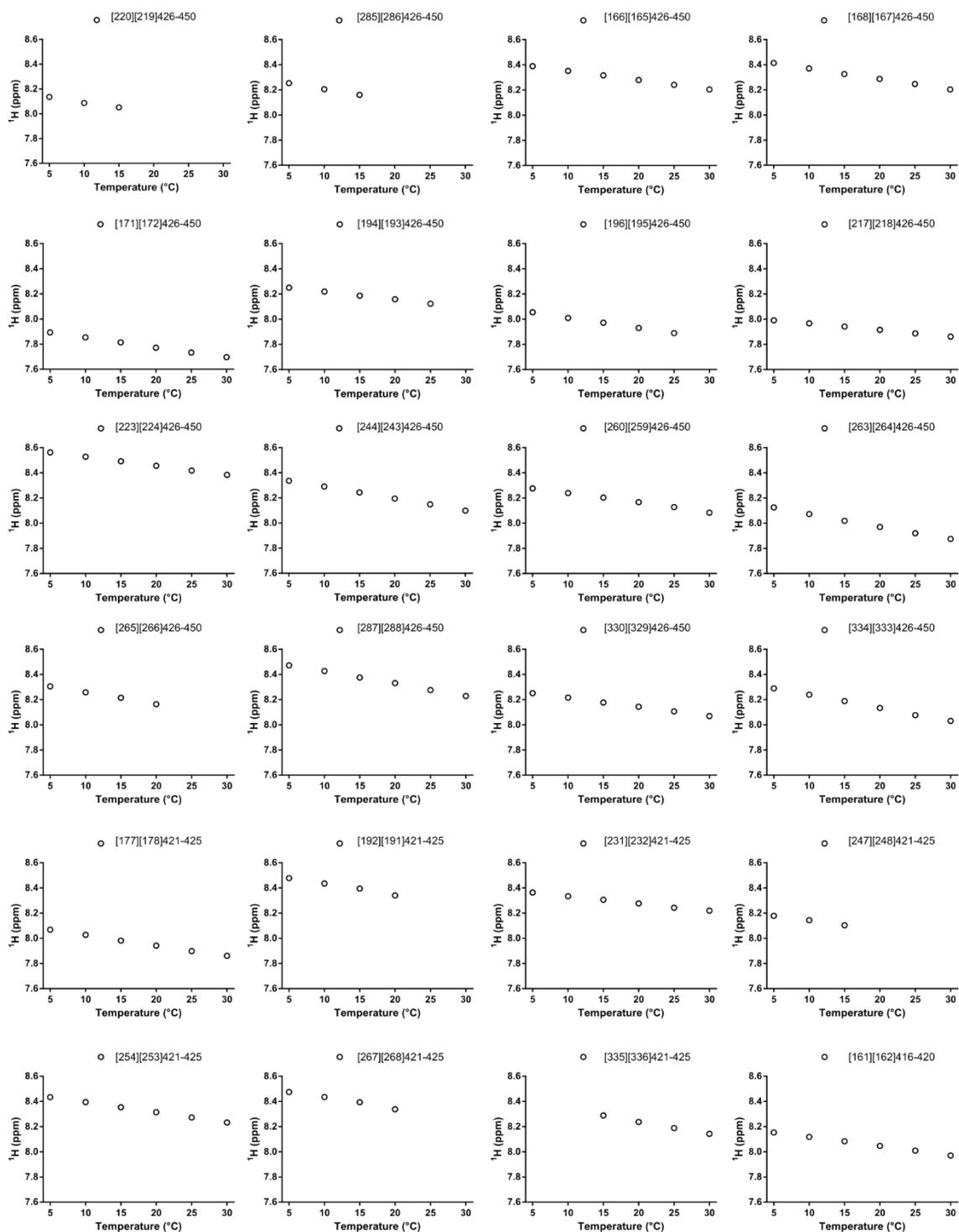
Sequencing Primer 2 result

CNNGCTTGTNAGCAGCCGGATCTCAGTGGTGGTGGTGGTGGTCTCGAGTGCGGCCGCAAGCTTGTGACGGCGCTCGAATTCGG
 ATCCGTTATCCACTTCCAATGTATTAGCTCAGGATGATGGTAGCCATGTCTTAAGCATGGAGTCCACGGGGCCTCGGGCCGGGC
 AGGGGATGGGCGGGCTTCTCTCCACATCCCAGACACGGTGGTAGGTGCGTTTTCTGAAGTCTGGTCAACCAGGTTGATAACTT
 CCTCAAAGCTCTTTTCTCTGAATCAGCAGGAATCACATTTTCCGATGTTTGGGTAGATTGTTGGGAAATCTCTCCAGCATCTG
 GTAGACGCAGACAGTGGGTTTCATGGTGTCTATCAGAAGCCAGTAATTCCTCAAGTAGTTGAGCTTGTCTTGTCTTTGAGTCC
 GGGATCAAATTTGACGTCCGTGCTGGGCGTGATCACACTCCCCCTTGTCCCGATGGTGACGCCATCAGTCTGGGGCCCTTCCA
 GCAAAGGAAGTGTGCTGCCGCGGGGCACGACAGCAACCCCTCGTGTACCATTTGAGGGAGAGGCATAGAGGCTGGTAGAGTATTT
 CCAACATACAGAGTGGGCGTCAGCTTGTCTTGGCCTCTGTCTCCTTGGGGAACGGGTACTTCCACTTTGTGATGCGCCCCACCTC
 CCCAGACATGAAGGTCAGATAGCGCAGGGTCTCCACAGCGACATTGATGTGCATCACCTTCTCAGACCCTCCCGCTGCCAGACAT
 AAAAGGCCACCACAGGGGAGGCGTAGTTTGGATCCACAGGACGTCCTCCAGATTCACCTGTCACAGTCAACCACAGCCCATACCA
 TTGGACAAAGTGGGACATCTTGTAGTCCACGTCGTCCTCAGGCAGTGAGGCCGATAGTCAAAGTAGNGGCATTCCACCGGAG
 CTCTCGGGTTTTGGTGTGTCATGGGNNAT

Figure 9.5

The luminal domain construct is shown with sequencing data from the forward and reverse primers used. Together the primers give complete coverage of the protein construct. The region of the protein construct covered by each sequencing result is represented by red for the forward primer and green for the reverse.





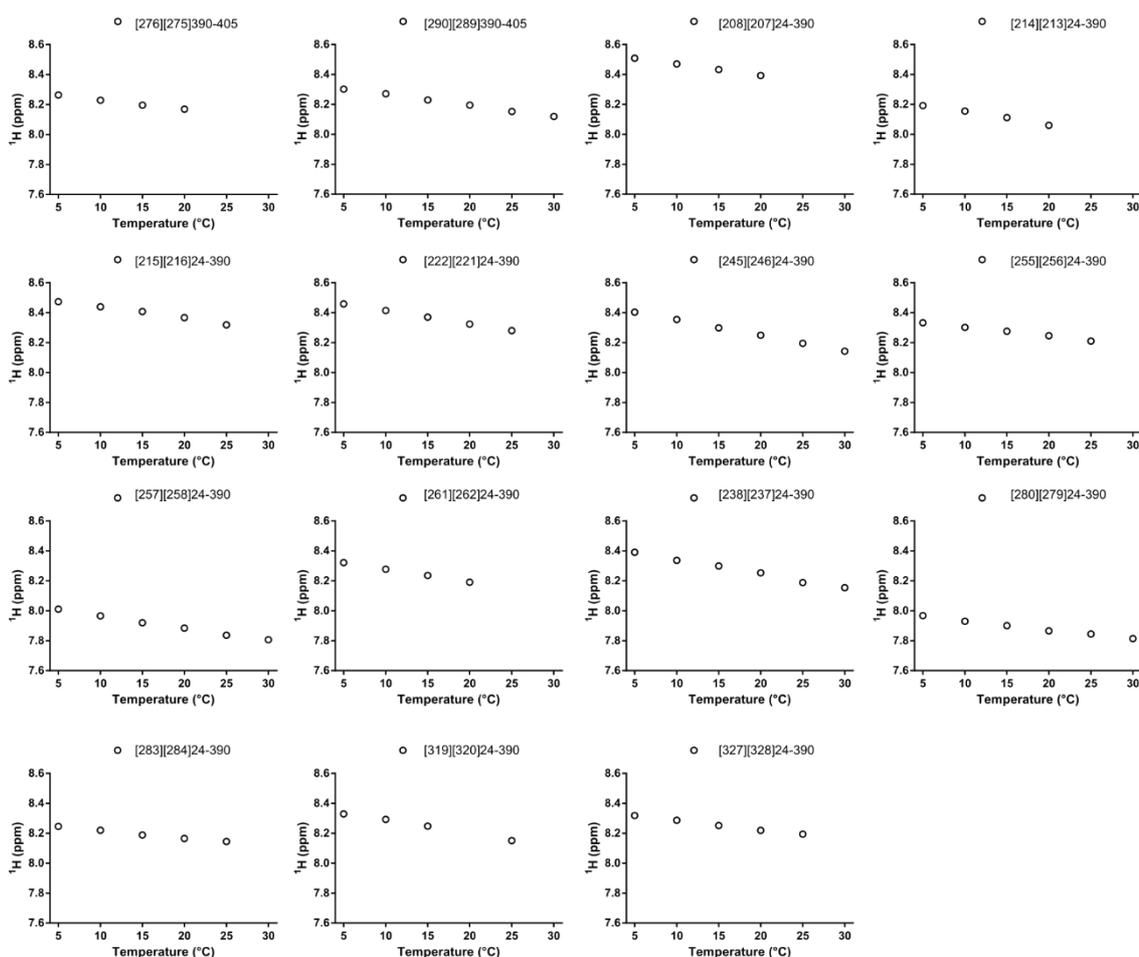


Figure 9.6

Temperature coefficient graphs for each peak in the spectra of wild-type luminal domain. The ^1H chemical shift is plotted against the temperature used. Each peak is given arbitrary coordinates and the area in the protein from which the residue is from (such as 24-390 is residues 24-390).

Table 9.1

The chemical shift temperature coefficient of the peaks shown in Figure 9.6.

Arbitrary Number	Partial assignment	CSTC ppm/K
[220] [219]	426-450	-0.008454
[285] [286]	426-450	-0.009293
[166] [165]	426-450	-0.0074032
[168] [167]	426-450	-0.0083381
[171] [172]	426-450	-0.0079363
[194] [193]	426-450	-0.0063078
[196] [195]	426-450	-0.008195
[217] [218]	426-450	-0.0052115
[223] [224]	426-450	-0.0071888

[244]	[243]	426-450	-0.0094987
[260]	[259]	426-450	-0.0076011
[263]	[264]	426-450	-0.0100438
[265]	[266]	426-450	-0.0093522
[287]	[288]	426-450	-0.0097882
[330]	[329]	426-450	-0.0072374
[334]	[333]	426-450	-0.010483
[177]	[178]	421-425	-0.0083684
[192]	[191]	421-425	-0.0090398
[231]	[232]	421-425	-0.0058606
[247]	[248]	421-425	-0.007564
[254]	[253]	421-425	-0.0080635
[267]	[268]	421-425	-0.0089934
[335]	[336]	421-425	-0.0097442
[161]	[162]	416-420	-0.0073345
[201]	[202]	416-420	-0.0086947
[331]	[332]	416-420	-0.0106284
[226]	[225]	411-415	-0.0091304
[250]	[249]	411-415	-0.0061108
[277]	[278]	411-415	-0.0058232
[281]	[282]	411-415	-0.0059338
[273]	[274]	411-415	-0.0083513
[174]	[173]	406-410	-0.0079165
[175]	[176]	406-410	-0.0099272
[170]	[169]	406-410	-0.0073723
[199]	[200]	406-410	-0.0066324
[229]	[230]	406-410	-0.010503
[179]	[180]	390-405	-0.007653
[189]	[190]	390-405	-0.008277
[197]	[198]	390-405	-0.0072282
[204]	[203]	390-405	-0.0054399
[210]	[209]	390-405	-0.0037494
[233]	[234]	390-405	-0.0110418
[239]	[240]	390-405	-0.0089566
[241]	[242]	390-405	-0.0079332

[276]	[275]	390-405	-0.006302
[290]	[289]	390-405	-0.007455

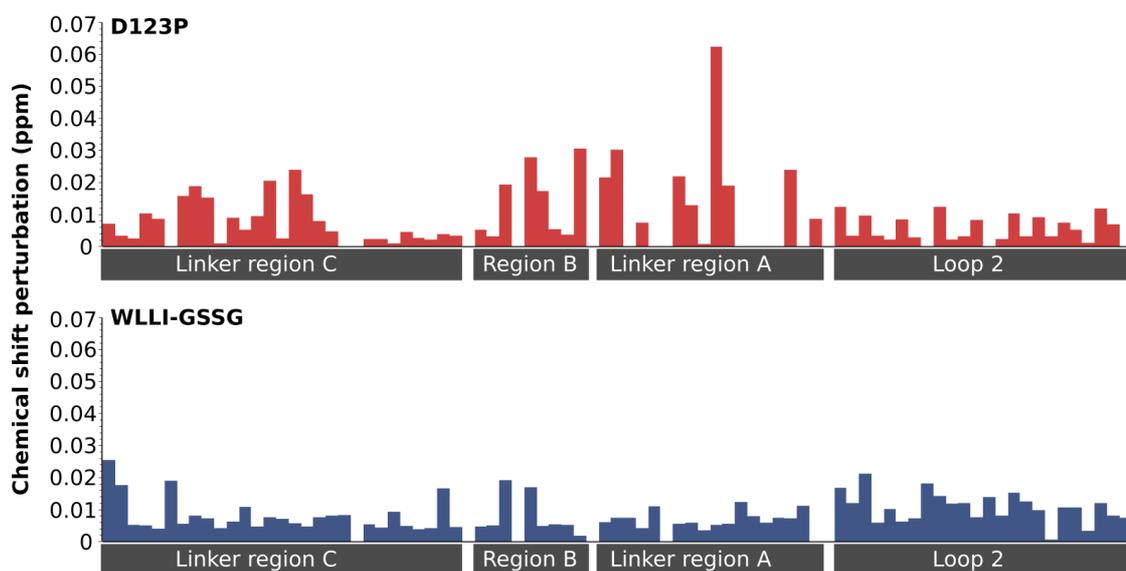


Figure 9.7

Chemical shift perturbations of the D123P and WLLI-GSSG mutants compared to wild-type protein (Section 2.7.1).

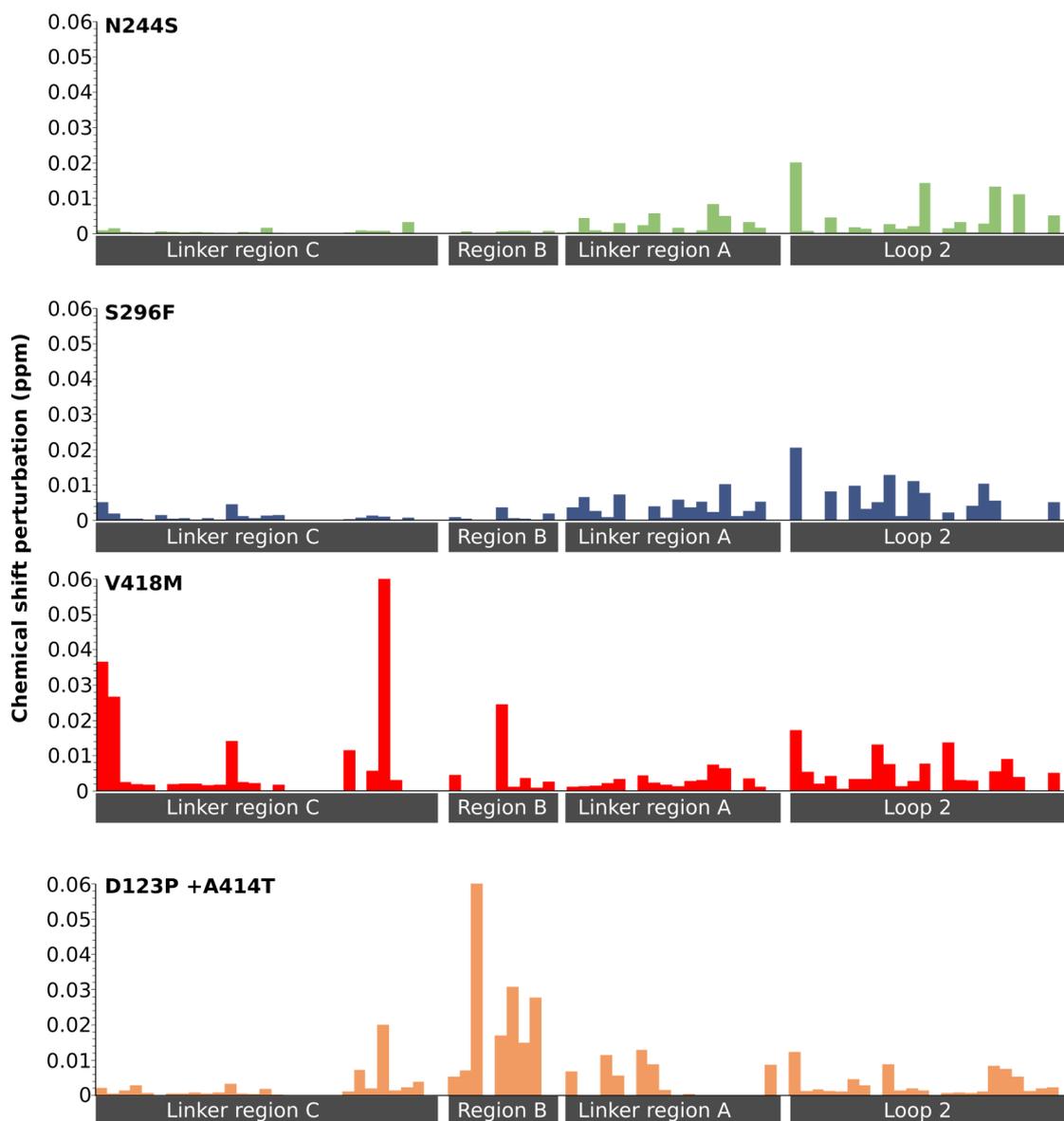


Figure 9.8

Chemical shift perturbations for the cancer-associated mutants. The N244S, S296F and V418M mutants' chemical shift perturbations are calculated in relation to wild-type protein (Section 2.7.1), the chemical shift perturbations for the D123P + A414T construct is compared to the D123P mutant.

10. Bibliography

Adams, B.M., Oster, M.E. and Hebert, D.N. 2019. Protein Quality Control in the Endoplasmic Reticulum. *Protein J.*

Ali, M.M.U., Bagratuni, T., Davenport, E.L., Nowak, P.R., Silva-Santisteban, M.C., Hardcastle, A., McAndrews, C., Rowlands, M.G., Morgan, G.J., Aherne, W., Collins, I., Davies, F.E. and Pearl, L.H. 2011. Structure of the Ire1 autophosphorylation complex and implications for the unfolded protein response. *EMBO J.* **30**(5), pp.894-905.

Amin-Wetzel, N., Saunders, R.A., Kamphuis, M.J., Rato, C., Preissler, S., Harding, H.P. and Ron, D. 2017. A J-Protein Co-chaperone Recruits BiP to Monomerize IRE1 and Repress the Unfolded Protein Response. *Cell.* **171**(7), pp.1625-1637.e1613.

Aragon, T., van Anken, E., Pincus, D., Serafimova, I.M., Korennykh, A.V., Rubio, C.A. and Walter, P. 2009. Messenger RNA targeting to endoplasmic reticulum stress signalling sites. *Nature.* **457**(7230), pp.736-740.

Arakawa, T., Ejima, D., Tsumoto, K., Obeyama, N., Tanaka, Y., Kita, Y. and Timasheff, S.N. 2007. Suppression of protein interactions by arginine: a proposed mechanism of the arginine effects. *Biophys Chem.* **127**(1-2), pp.1-8.

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T. and Ben-Tal, N. 2016. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44**(Web Server issue), pp.W344-350.

Aslanidis, C. and de Jong, P.J. 1990. Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res.* **18**(20), pp.6069-6074.

Baynes, B.M., Wang, D.I. and Trout, B.L. 2005. Role of arginine in the stabilization of proteins against aggregation. *Biochemistry.* **44**(12), pp.4919-4925.

Bertolotti, A., Zhang, Y., Hendershot, L.M., Harding, H.P. and Ron, D. 2000. Dynamic interaction of BiP and ER stress transducers in the unfolded-protein response. *Nat Cell Biol.* **2**(6), pp.326-332.

Blattner, F.R., Plunkett, G., 3rd, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B. and Shao, Y. 1997. The complete genome sequence of Escherichia coli K-12. *Science.* **277**(5331), pp.1453-1462.

Bocking, T., Aguet, F., Harrison, S.C. and Kirchhausen, T. 2011. Single-molecule analysis of a molecular disassemblase reveals the mechanism of Hsc70-driven clathrin uncoating. *Nat Struct Mol Biol.* **18**(3), pp.295-301.

Borgia, M.B., Nickson, A.A., Clarke, J. and Hounslow, M.J. 2013. A Mechanistic Model for Amorphous Protein Aggregation of Immunoglobulin-like Domains. *Journal of the American Chemical Society.* **135**(17), pp.6456-6464.

Braun, P., Hu, Y., Shen, B., Halleck, A., Koundinya, M., Harlow, E. and LaBaer, J. 2002. Proteome-scale purification of human proteins from bacteria. *Proc Natl Acad Sci U S A*. **99**(5), pp.2654-2659.

Brutscher, B., Felli, I.C., Gil-Caballero, S., Hosek, T., Kummerle, R., Piai, A., Pierattelli, R. and Solyom, Z. 2015. NMR Methods for the Study of Intrinsically Disordered Proteins Structure, Dynamics, and Interactions: General Overview and Practical Guidelines. *Adv Exp Med Biol*. **870**, pp.49-122.

Burgess, R.R. 2009. Refolding solubilized inclusion body proteins. *Methods Enzymol*. **463**, pp.259-282.

Bustamante, C., Tinoco, I., Jr. and Maestre, M.F. 1983. Circular differential scattering can be an important part of the circular dichroism of macromolecules. *Proc Natl Acad Sci U S A*. **80**(12), pp.3568-3572.

Calfon, M., Zeng, H., Urano, F., Till, J.H., Hubbard, S.R., Harding, H.P., Clark, S.G. and Ron, D. 2002. IRE1 couples endoplasmic reticulum load to secretory capacity by processing the XBP-1 mRNA. *Nature*. **415**(6867), pp.92-96.

Carrara, M., Prischi, F., Nowak, P.R., Kopp, M.C. and Ali, M.M.U. 2015. Noncanonical binding of BiP ATPase domain to Ire1 and Perk is dissociated by unfolded protein CH1 to initiate ER stress signaling. *eLife*. **4**, pe03522.

Cescon, M., Chen, P., Castagnaro, S., Gregorio, I. and Bonaldo, P. 2016. Lack of collagen VI promotes neurodegeneration by impairing autophagy and inducing apoptosis during aging. *Aging (Albany NY)*. **8**(5), pp.1083-1101.

Chakrabarti, A., Chen, A.W. and Varner, J.D. 2011. A review of the mammalian unfolded protein response. *Biotechnol Bioeng*. **108**(12), pp.2777-2793.

Chalmers, F., Mogre, S., Son, J., Blazanin, N. and Glick, A.B. 2019. The multiple roles of the unfolded protein response regulator IRE1alpha in cancer. *Mol Carcinog*.

Chen, X., Iliopoulos, D., Zhang, Q., Tang, Q., Greenblatt, M.B., Hatziapostolou, M., Lim, E., Tam, W.L., Ni, M., Chen, Y., Mai, J., Shen, H., Hu, D.Z., Adoro, S., Hu, B., Song, M., Tan, C., Landis, M.D., Ferrari, M., Shin, S.J., Brown, M., Chang, J.C., Liu, X.S. and Glimcher, L.H. 2014. XBP1 promotes triple-negative breast cancer by controlling the HIF1alpha pathway. *Nature*. **508**(7494), pp.103-107.

Chen, X., Shen, J. and Prywes, R. 2002. The luminal domain of ATF6 senses endoplasmic reticulum (ER) stress and causes translocation of ATF6 from the ER to the Golgi. *J Biol Chem*. **277**(15), pp.13045-13052.

Choi, J.H., Keum, K.C. and Lee, S.Y. 2006. Production of recombinant proteins by high cell density culture of Escherichia coli. *Chemical Engineering Science*. **61**(3), pp.876-885.

Cierpicki, T. and Otlewski, J. 2001. Amide proton temperature coefficients as hydrogen bond indicators in proteins. *J Biomol NMR*. **21**(3), pp.249-261.

Clerico, E.M., Tilitsky, J.M., Meng, W. and Gierasch, L.M. 2015. How hsp70 molecular machines interact with their substrates to mediate diverse physiological functions. *J Mol Biol.* **427**(7), pp.1575-1588.

Coelho, D.S. and Domingos, P.M. 2014. Physiological roles of regulated Ire1 dependent decay. *Front Genet.* **5**, p76.

Colombano, G., Caldwell, J.J., Matthews, T.P., Bhatia, C., Joshi, A., McHardy, T., Mok, N.Y., Newbatt, Y., Pickard, L., Strover, J., Hedayat, S., Walton, M.I., Myers, S.M., Jones, A.M., Saville, H., McAndrew, C., Burke, R., Eccles, S.A., Davies, F.E., Bayliss, R. and Collins, I. 2019. Binding to an Unusual Inactive Kinase Conformation by Highly Selective Inhibitors of Inositol-Requiring Enzyme 1alpha Kinase-Endoribonuclease. *J Med Chem.* **62**(5), pp.2447-2465.

Concha, N.O., Smallwood, A., Bonnette, W., Totoritis, R., Zhang, G., Federowicz, K., Yang, J., Qi, H., Chen, S., Campobasso, N., Choudhry, A.E., Shuster, L.E., Evans, K.A., Ralph, J., Sweitzer, S., Heerding, D.A., Buser, C.A., Su, D.S. and DeYoung, M.P. 2015. Long-Range Inhibitor-Induced Conformational Regulation of Human IRE1alpha Endoribonuclease Activity. *Mol Pharmacol.* **88**(6), pp.1011-1023.

Craig, E.A., Huang, P., Aron, R. and Andrew, A. 2006. The diverse roles of J-proteins, the obligate Hsp70 co-chaperone. *Rev Physiol Biochem Pharmacol.* **156**, pp.1-21.

Credle, J.J., Finer-Moore, J.S., Papa, F.R., Stroud, R.M. and Walter, P. 2005. On the mechanism of sensing unfolded protein in the endoplasmic reticulum. *Proc Natl Acad Sci U S A.* **102**(52), pp.18773-18784.

Cuevas, E.P., Eraso, P., Mazon, M.J., Santos, V., Moreno-Bueno, G., Cano, A. and Portillo, F. 2017. LOXL2 drives epithelial-mesenchymal transition via activation of IRE1-XBP1 signalling pathway. *Sci Rep.* **7**, p44988.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. 1995. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR.* **6**(3), pp.277-293.

Denoyelle, C., Abou-Rjaily, G., Bezrookove, V., Verhaegen, M., Johnson, T.M., Fullen, D.R., Pointer, J.N., Gruber, S.B., Su, L.D., Nikiforov, M.A., Kaufman, R.J., Bastian, B.C. and Soengas, M.S. 2006. Anti-oncogenic role of the endoplasmic reticulum differentially activated by mutations in the MAPK pathway. *Nat Cell Biol.* **8**(10), pp.1053-1063.

Dobson, J., Kumar, A., Willis, L.F., Tuma, R., Higazi, D.R., Turner, R., Lowe, D.C., Ashcroft, A.E., Radford, S.E., Kapur, N. and Brockwell, D.J. 2017. Inducing protein aggregation by extensional flow. *Proc Natl Acad Sci U S A.* **114**(18), pp.4673-4678.

Dong, B., Niwa, M., Walter, P. and Silverman, R.H. 2001. Basis for regulated RNA cleavage by functional analysis of RNase L and Ire1p. *Rna.* **7**(3), pp.361-373.

Dong, X.Y., Huang, Y. and Sun, Y. 2004. Refolding kinetics of denatured-reduced lysozyme in the presence of folding aids. *J Biotechnol.* **114**(1-2), pp.135-142.

Drozdetskiy, A., Cole, C., Procter, J. and Barton, G.J. 2015. JPred4: a protein secondary structure prediction server. *Nucleic Acids Res.* **43**(W1), pp.W389-394.

Dumon-Seignovert, L., Cariot, G. and Vuillard, L. 2004. The toxicity of recombinant proteins in *Escherichia coli*: a comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3). *Protein Expr Purif.* **37**(1), pp.203-206.

Duran-Aniotz, C., Cornejo, V.H., Espinoza, S., Ardiles, A.O., Medinas, D.B., Salazar, C., Foley, A., Gajardo, I., Thielen, P., Iwawaki, T., Scheper, W., Soto, C., Palacios, A.G., Hoozemans, J.J.M. and Hetz, C. 2017. IRE1 signaling exacerbates Alzheimer's disease pathogenesis. *Acta Neuropathol.* **134**(3), pp.489-506.

Eletto, D., Eletto, D., Dersh, D., Gidalevitz, T. and Argon, Y. 2014. Protein disulfide isomerase A6 controls the decay of IRE1alpha signaling via disulfide-dependent association. *Mol Cell.* **53**(4), pp.562-576.

Fahnert, B., Lilie, H. and Neubauer, P. 2004. Inclusion bodies: formation and utilisation. *Adv Biochem Eng Biotechnol.* **89**, pp.93-142.

Faustino, A.F., Barbosa, G.M., Silva, M., Castanho, M., Da Poian, A.T., Cabrita, E.J., Santos, N.C., Almeida, F.C.L. and Martins, I.C. 2019. Fast NMR method to probe solvent accessibility and disordered regions in proteins. *Sci Rep.* **9**(1), p1647.

Fekete, S., Beck, A., Veuthey, J.L. and Guillaume, D. 2014. Theory and practice of size exclusion chromatography for the analysis of protein aggregates. *J Pharm Biomed Anal.* **101**, pp.161-173.

Feldman, H.C., Tong, M., Wang, L., Meza-Acevedo, R., Gobillot, T.A., Lebedev, I., Gliedt, M.J., Hari, S.B., Mitra, A.K., Backes, B.J., Papa, F.R., Seeliger, M.A. and Maly, D.J. 2016. Structural and Functional Analysis of the Allosteric Inhibition of IRE1alpha with ATP-Competitive Ligands. *ACS Chem Biol.* **11**(8), pp.2195-2205.

Frueh, D.P. 2014. Practical aspects of NMR signal assignment in larger and challenging proteins. *Prog Nucl Magn Reson Spectrosc.* **78**, pp.47-75.

Gardner, B.M. and Walter, P. 2011. Unfolded proteins are Ire1-activating ligands that directly induce the unfolded protein response. *Science.* **333**(6051), pp.1891-1894.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D. and Bairoch, A. 2003. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **31**(13), pp.3784-3788.

Gebauer, F. and Hentze, M.W. 2004. Molecular mechanisms of translational control. *Nat Rev Mol Cell Biol.* **5**(10), pp.827-835.

Ghosh, R., Wang, L., Wang, E.S., Perera, B.G., Igbaria, A., Morita, S., Prado, K., Thamsen, M., Caswell, D., Macias, H., Weiberth, K.F., Gliedt, M.J., Alavi, M.V., Hari, S.B., Mitra, A.K., Bhatarai, B., Schurer, S.C., Snapp, E.L., Gould, D.B., German, M.S., Backes, B.J., Maly, D.J., Oakes, S.A. and Papa, F.R. 2014. Allosteric inhibition of the IRE1alpha RNase preserves cell viability and function during endoplasmic reticulum stress. *Cell.* **158**(3), pp.534-548.

Greenman, C., Stephens, P., Smith, R., Dalgliesh, G.L., Hunter, C., Bignell, G., Davies, H., Teague, J., Butler, A., Stevens, C., Edkins, S., O'Meara, S., Vastrik, I., Schmidt, E.E., Avis, T., Barthorpe, S., Bhamra, G., Buck, G., Choudhury, B., Clements, J., Cole, J., Dicks, E., Forbes, S., Gray, K., Halliday, K., Harrison, R., Hills, K., Hinton, J., Jenkinson, A., Jones, D., Menzies, A., Mironenko, T., Perry, J., Raine, K., Richardson, D., Shepherd, R., Small, A., Tofts, C., Varian, J., Webb, T., West, S., Widaa, S., Yates, A., Cahill, D.P., Louis, D.N., Goldstraw, P., Nicholson, A.G., Brasseur, F., Looijenga, L., Weber, B.L., Chiew, Y.E., DeFazio, A., Greaves, M.F., Green, A.R., Campbell, P., Birney, E., Easton, D.F., Chenevix-Trench, G., Tan, M.H., Khoo, S.K., Teh, B.T., Yuen, S.T., Leung, S.Y., Wooster, R., Futreal, P.A. and Stratton, M.R. 2007. Patterns of somatic mutation in human cancer genomes. *Nature*. **446**(7132), pp.153-158.

Greenshields, A.L., Knickle, L.C., Syvitski, R. and Douglas, S.E. 2008. Strategies for recombinant expression of small, highly disulphide-bonded, cationic antimicrobial peptides. *Protein Pept Lett*. **15**(9), pp.985-994.

Groenendyk, J., Peng, Z., Dudek, E., Fan, X., Mizianty, M.J., Dufey, E., Urrea, H., Sepulveda, D., Rojas-Rivera, D., Lim, Y., Kim, D.H., Baretta, K., Srikanth, S., Gwack, Y., Ahnn, J., Kaufman, R.J., Lee, S.K., Hetz, C., Kurgan, L. and Michalak, M. 2014. Interplay between the oxidoreductase PDIA6 and microRNA-322 controls the response to disrupted endoplasmic reticulum calcium homeostasis. *Sci Signal*. **7**(329), pra54.

Gupta, S. and Bhattacharjya, S. 2014. NMR characterization of the near native and unfolded states of the PTB domain of Dok1: alternate conformations and residual clusters. *PLoS One*. **9**(2), pe90557.

Haas, E. 2012. Ensemble FRET methods in studies of intrinsically disordered proteins. *Methods Mol Biol*. **895**, pp.467-498.

Halbleib, K., Pesek, K., Covino, R., Hofbauer, H.F., Wunnicke, D., Hanelt, I., Hummer, G. and Ernst, R. 2017. Activation of the Unfolded Protein Response by Lipid Bilayer Stress. *Mol Cell*. **67**(4), pp.673-684.e678.

Han, D., Lerner, A.G., Vande Walle, L., Upton, J.P., Xu, W., Hagen, A., Backes, B.J., Oakes, S.A. and Papa, F.R. 2009. IRE1alpha kinase activation modes control alternate endoribonuclease outputs to determine divergent cell fates. *Cell*. **138**(3), pp.562-575.

Harding, H.P., Zhang, Y. and Ron, D. 1999. Protein translation and folding are coupled by an endoplasmic-reticulum-resident kinase. *Nature*. **397**(6716), pp.271-274.

Harrington, P.E., Biswas, K., Malwitz, D., Tasker, A.S., Mohr, C., Andrews, K.L., Dellamaggiore, K., Kendall, R., Beckmann, H., Jaeckel, P., Materna-Reichelt, S., Allen, J.R. and Lipford, J.R. 2015. Unfolded Protein Response in Cancer: IRE1alpha Inhibition by Selective Kinase Ligands Does Not Impair Tumor Cell Viability. *ACS Med Chem Lett*. **6**(1), pp.68-72.

Hayashi, K. and Kojima, C. 2008. pCold-GST vector: a novel cold-shock vector containing GST tag for soluble protein production. *Protein Expr Purif*. **62**(1), pp.120-127.

Haze, K., Yoshida, H., Yanagi, H., Yura, T. and Mori, K. 1999. Mammalian transcription factor ATF6 is synthesized as a transmembrane protein and activated by proteolysis in response to endoplasmic reticulum stress. *Mol Biol Cell*. **10**(11), pp.3787-3799.

- Hendershot, L.M. 2004. The ER function BiP is a master regulator of ER function. *Mt Sinai J Med.* **71**(5), pp.289-297.
- Hibino, T., Misawa, S., Wakiyama, M., Maeda, S., Yazaki, K., Kumagai, I., Ooi, T. and Miura, K. 1994. High-level expression of porcine muscle adenylate kinase in *Escherichia coli*: effects of the copy number of the gene and the translational initiation signals. *J Biotechnol.* **32**(2), pp.139-148.
- Hirel, P.H., Schmitter, M.J., Dessen, P., Fayat, G. and Blanquet, S. 1989. Extent of N-terminal methionine excision from *Escherichia coli* proteins is governed by the side-chain length of the penultimate amino acid. *Proc Natl Acad Sci U S A.* **86**(21), pp.8247-8251.
- Hochuli, E., Bannwarth, W., Döbeli, H., Gentz, R. and Stüber, D. 1988. Genetic Approach to Facilitate Purification of Recombinant Proteins with a Novel Metal Chelate Adsorbent. *Bio/Technology.* **6**(11), pp.1321-1325.
- Hoffmann, F. and Rinas, U. 2004. Stress induced by recombinant protein production in *Escherichia coli*. *Adv Biochem Eng Biotechnol.* **89**, pp.73-92.
- Hogg, P.J. 2003. Disulfide bonds as switches for protein function. *Trends Biochem Sci.* **28**(4), pp.210-214.
- Hollien, J., Lin, J.H., Li, H., Stevens, N., Walter, P. and Weissman, J.S. 2009. Regulated Ire1-dependent decay of messenger RNAs in mammalian cells. *J Cell Biol.* **186**(3), pp.323-331.
- Hong, J., Lee, A., Han, H. and Kim, J. 2009. Structural characterization of immunoglobulin G using time-dependent disulfide bond reduction. *Anal Biochem.* **384**(2), pp.368-370.
- Hou, F., Sun, L., Zheng, H., Skaug, B., Jiang, Q.X. and Chen, Z.J. 2011. MAVS forms functional prion-like aggregates to activate and propagate antiviral innate immune response. *Cell.* **146**(3), pp.448-461.
- Hsiao, K. 1993. Exonuclease III induced ligase-free directional subcloning of PCR products. *Nucleic Acids Res.* **21**(23), pp.5528-5529.
- Hur, K.Y., So, J.S., Ruda, V., Frank-Kamenetsky, M., Fitzgerald, K., Koteliansky, V., Iwawaki, T., Glimcher, L.H. and Lee, A.H. 2012. IRE1alpha activation protects mice against acetaminophen-induced hepatotoxicity. *J Exp Med.* **209**(2), pp.307-318.
- Iadanza, M.G., Silvers, R., Boardman, J., Smith, H.I., Karamanos, T.K., Debelouchina, G.T., Su, Y., Griffin, R.G., Ranson, N.A. and Radford, S.E. 2018. The structure of a β (2)-microglobulin fibril suggests a molecular basis for its amyloid polymorphism. *Nat Commun.* **9**(1), p4517.
- Ignatova, Z. and Gierasch, L.M. 2006. Inhibition of protein aggregation in vitro and in vivo by a natural osmoprotectant. *Proc Natl Acad Sci U S A.* **103**(36), pp.13357-13361.
- Imagawa, Y., Hosoda, A., Sasaka, S., Tsuru, A. and Kohno, K. 2008. RNase domains determine the functional difference between IRE1alpha and IRE1beta. *FEBS Lett.* **582**(5), pp.656-660.

Ishiwata-Kimata, Y., Promlek, T., Kohno, K. and Kimata, Y. 2013. BiP-bound and nonclustered mode of Ire1 evokes a weak but sustained unfolded protein response. *Genes Cells*. **18**(4), pp.288-301.

Itzhak, D., Bright, M., McAndrew, P., Mirza, A., Newbatt, Y., Strover, J., Widya, M., Thompson, A., Morgan, G., Collins, I. and Davies, F. 2014. Multiple autophosphorylations significantly enhance the endoribonuclease activity of human inositol requiring enzyme 1alpha. *BMC Biochem*. **15**, p3.

Iyer, K.S. and Klee, W.A. 1973. Direct spectrophotometric measurement of the rate of reduction of disulfide bonds. The reactivity of the disulfide bonds of bovine γ -lactalbumin. *J Biol Chem*. **248**(2), pp.707-710.

Jameson, D.M. and Ross, J.A. 2010. Fluorescence polarization/anisotropy in diagnostics and imaging. *Chem Rev*. **110**(5), pp.2685-2708.

Jeggo, P.A., Pearl, L.H. and Carr, A.M. 2016. DNA repair, genome stability and cancer: a historical perspective. *Nat Rev Cancer*. **16**(1), pp.35-42.

Jerabek-Willemsen, M., André, T., Wanner, R., Roth, H.M., Duhr, S., Baaske, P. and Breitsprecher, D. 2014. MicroScale Thermophoresis: Interaction analysis and beyond. *Journal of Molecular Structure*. **1077**, pp.101-113.

Joshi, A., Newbatt, Y., McAndrew, P.C., Stubbs, M., Burke, R., Richards, M.W., Bhatia, C., Caldwell, J.J., McHardy, T., Collins, I. and Bayliss, R. 2015. Molecular mechanisms of human IRE1 activation through dimerization and ligand binding. *Oncotarget*. **6**(15), pp.13019-13035.

Kapust, R.B. and Waugh, D.S. 1999. Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci*. **8**(8), pp.1668-1674.

Karagoz, G.E., Acosta-Alvear, D., Nguyen, H.T., Lee, C.P., Chu, F. and Walter, P. 2017. An unfolded protein-induced conformational switch activates mammalian IRE1. *Elife*. **6**.

Kaser, A., Lee, A.H., Franke, A., Glickman, J.N., Zeissig, S., Tilg, H., Nieuwenhuis, E.E., Higgins, D.E., Schreiber, S., Glimcher, L.H. and Blumberg, R.S. 2008. XBP1 links ER stress to intestinal inflammation and confers genetic risk for human inflammatory bowel disease. *Cell*. **134**(5), pp.743-756.

Kaufman, R.J., Scheuner, D., Schroder, M., Shen, X., Lee, K., Liu, C.Y. and Arnold, S.M. 2002. The unfolded protein response in nutrient sensing and differentiation. *Nat Rev Mol Cell Biol*. **3**(6), pp.411-421.

Kebache, S., Cardin, E., Nguyen, D.T., Chevet, E. and Larose, L. 2004. Nck-1 antagonizes the endoplasmic reticulum stress-induced inhibition of translation. *J Biol Chem*. **279**(10), pp.9662-9671.

Keestra-Gounder, A.M., Byndloss, M.X., Seyffert, N., Young, B.M., Chavez-Arroyo, A., Tsai, A.Y., Cevallos, S.A., Winter, M.G., Pham, O.H., Tiffany, C.R., de Jong, M.F., Kerrinnes, T.,

Ravindran, R., Luciw, P.A., McSorley, S.J., Baumler, A.J. and Tsolis, R.M. 2016. NOD1 and NOD2 signalling links ER stress with inflammation. *Nature*. **532**(7599), pp.394-397.

Kimata, Y., Ishiwata-Kimata, Y., Ito, T., Hirata, A., Suzuki, T., Oikawa, D., Takeuchi, M. and Kohno, K. 2007. Two regulatory steps of ER-stress sensor Ire1 involving its cluster formation and interaction with unfolded proteins. *J Cell Biol.* **179**(1), pp.75-86.

Kimata, Y., Oikawa, D., Shimizu, Y., Ishiwata-Kimata, Y. and Kohno, K. 2004. A role for BiP as an adjutor for the endoplasmic reticulum stress-sensing protein Ire1. *J Cell Biol.* **167**(3), pp.445-456.

Kitai, Y., Ariyama, H., Kono, N., Oikawa, D., Iwawaki, T. and Arai, H. 2013. Membrane lipid saturation activates IRE1alpha without inducing clustering. *Genes Cells.* **18**(9), pp.798-809.

Kleckner, I.R. and Foster, M.P. 2011. An introduction to NMR-based approaches for measuring protein dynamics. *Biochim Biophys Acta.* **1814**(8), pp.942-968.

Kokame, K., Kato, H. and Miyata, T. 2001. Identification of ERSE-II, a new cis-acting element responsible for the ATF6-dependent mammalian unfolded protein response. *J Biol Chem.* **276**(12), pp.9199-9205.

Kono, N., Amin-Wetzel, N. and Ron, D. 2017. Generic membrane-spanning features endow IRE1alpha with responsiveness to membrane aberrancy. *Mol Biol Cell.* **28**(17), pp.2318-2332.

Konrat, R. 2014. NMR contributions to structural dynamics studies of intrinsically disordered proteins. *J Magn Reson.* **241**, pp.74-85.

Kopp, M.C., Nowak, P.R., Larburu, N., Adams, C.J. and Ali, M.M. 2018. In vitro FRET analysis of IRE1 and BiP association and dissociation upon endoplasmic reticulum stress. *Elife.* **7**.

Korennykh, A. and Walter, P. 2012. Structural basis of the unfolded protein response. *Annu Rev Cell Dev Biol.* **28**, pp.251-277.

Korennykh, A.V., Egea, P.F., Korostelev, A.A., Finer-Moore, J., Stroud, R.M., Zhang, C., Shokat, K.M. and Walter, P. 2011a. Cofactor-mediated conformational control in the bifunctional kinase/RNase Ire1. *BMC Biol.* **9**, p48.

Korennykh, A.V., Egea, P.F., Korostelev, A.A., Finer-Moore, J., Zhang, C., Shokat, K.M., Stroud, R.M. and Walter, P. 2009. The unfolded protein response signals through high-order assembly of Ire1. *Nature.* **457**(7230), pp.687-693.

Korennykh, A.V., Korostelev, A.A., Egea, P.F., Finer-Moore, J., Stroud, R.M., Zhang, C., Shokat, K.M. and Walter, P. 2011b. Structural and functional basis for RNA cleavage by Ire1. *BMC Biol.* **9**, p47.

Kornev, A.P., Haste, N.M., Taylor, S.S. and Eyck, L.F. 2006. Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism. *Proc Natl Acad Sci U S A.* **103**(47), pp.17783-17788.

Kosol, S., Contreras-Martos, S., Cedeno, C. and Tompa, P. 2013. Structural characterization of intrinsically disordered proteins by NMR spectroscopy. *Molecules*. **18**(9), pp.10802-10828.

Koumenis, C., Naczki, C., Koritzinsky, M., Rastani, S., Diehl, A., Sonenberg, N., Koromilas, A. and Wouters, B.G. 2002. Regulation of protein synthesis by hypoxia via activation of the endoplasmic reticulum kinase PERK and phosphorylation of the translation initiation factor eIF2 α . *Mol Cell Biol*. **22**(21), pp.7405-7416.

LeBlanc, S.J., Kulkarni, P. and Weninger, K.R. 2018. Single Molecule FRET: A Powerful Tool to Study Intrinsically Disordered Proteins. *Biomolecules*. **8**(4).

Lee, A.H., Heidtman, K., Hotamisligil, G.S. and Glimcher, L.H. 2011. Dual and opposing roles of the unfolded protein response regulated by IRE1 α and XBP1 in proinsulin processing and insulin secretion. *Proc Natl Acad Sci U S A*. **108**(21), pp.8885-8890.

Lee, A.H., Iwakoshi, N.N. and Glimcher, L.H. 2003. XBP-1 regulates a subset of endoplasmic reticulum resident chaperone genes in the unfolded protein response. *Mol Cell Biol*. **23**(21), pp.7448-7459.

Lee, D., Hilty, C., Wider, G. and Wuthrich, K. 2006. Effective rotational correlation times of proteins from NMR relaxation interference. *J Magn Reson*. **178**(1), pp.72-76.

Lee, K.P., Dey, M., Neculai, D., Cao, C., Dever, T.E. and Sicheri, F. 2008. Structure of the dual enzyme Ire1 reveals the basis for catalysis and regulation in nonconventional RNA splicing. *Cell*. **132**(1), pp.89-100.

Lhomond, S., Avril, T., Dejeans, N., Voutetakis, K., Doultinos, D., McMahon, M., Pineau, R., Obacz, J., Papadodima, O., Jouan, F., Bourien, H., Logotheti, M., Jegou, G., Pallares-Lupon, N., Schmit, K., Le Reste, P.J., Etcheverry, A., Mosser, J., Barroso, K., Vauleon, E., Maurel, M., Samali, A., Patterson, J.B., Pluquet, O., Hetz, C., Quillien, V., Chatziioannou, A. and Chevet, E. 2018. Dual IRE1 RNase functions dictate glioblastoma development. *EMBO Mol Med*.

Li, B., Gao, B., Ye, L., Han, X., Wang, W., Kong, L., Fang, X., Zeng, Y., Zheng, H., Li, S., Wu, Z. and Ye, L. 2007. Hepatitis B virus X protein (HBx) activates ATF6 and IRE1-XBP1 pathways of unfolded protein response. *Virus Res*. **124**(1-2), pp.44-49.

Li, H., Korennykh, A.V., Behrman, S.L. and Walter, P. 2010. Mammalian endoplasmic reticulum stress sensor IRE1 signals by dynamic clustering. *Proc Natl Acad Sci U S A*. **107**(37), pp.16113-16118.

Lin, J.H. and Lavail, M.M. 2010. Misfolded proteins and retinal dystrophies. *Adv Exp Med Biol*. **664**, pp.115-121.

Lin, J.H., Li, H., Yasumura, D., Cohen, H.R., Zhang, C., Panning, B., Shokat, K.M., Lavail, M.M. and Walter, P. 2007. IRE1 signaling affects cell fate during the unfolded protein response. *Science*. **318**(5852), pp.944-949.

Liu, C.Y., Wong, H.N., Schauerte, J.A. and Kaufman, R.J. 2002. The protein kinase/endoribonuclease IRE1 α that signals the unfolded protein response has a luminal N-terminal ligand-independent dimerization domain. *J Biol Chem*. **277**(21), pp.18346-18356.

- Liu, C.Y., Xu, Z. and Kaufman, R.J. 2003. Structure and intermolecular interactions of the luminal dimerization domain of human IRE1 α . *J Biol Chem.* **278**(20), pp.17680-17687.
- Lu, Y., Liang, F.X. and Wang, X. 2014. A synthetic biology approach identifies the mammalian UPR RNA ligase RtcB. *Mol Cell.* **55**(5), pp.758-770.
- Marciniak, S.J., Yun, C.Y., Oyadomari, S., Novoa, I., Zhang, Y., Jungreis, R., Nagata, K., Harding, H.P. and Ron, D. 2004. CHOP induces death by promoting protein synthesis and oxidation in the stressed endoplasmic reticulum. *Genes Dev.* **18**(24), pp.3066-3077.
- Mauro, C., Crescenzi, E., De Mattia, R., Pacifico, F., Mellone, S., Salzano, S., de Luca, C., D'Adamio, L., Palumbo, G., Formisano, S., Vito, P. and Leonardi, A. 2006. Central role of the scaffold protein tumor necrosis factor receptor-associated factor 2 in regulating endoplasmic reticulum stress-induced apoptosis. *J Biol Chem.* **281**(5), pp.2631-2638.
- Mavridis, L. and Janes, R.W. 2017. PDB2CD: a web-based application for the generation of circular dichroism spectra from protein atomic coordinates. *Bioinformatics.* **33**(1), pp.56-63.
- Mayer, M.P. 2013. Hsp70 chaperone dynamics and molecular mechanism. *Trends Biochem Sci.* **38**(10), pp.507-514.
- Merrick, W.C. 2004. Cap-dependent and cap-independent translation in eukaryotic systems. *Gene.* **332**, pp.1-11.
- Mittermaier, A.K. and Kay, L.E. 2009. Observing biological dynamics at atomic resolution using NMR. *Trends Biochem Sci.* **34**(12), pp.601-611.
- Mori, K. 2009. Signalling pathways in the unfolded protein response: development from yeast to mammals. *J Biochem.* **146**(6), pp.743-750.
- Morris, G.A. and Freeman, R. 1979. Enhancement of nuclear magnetic resonance signals by polarization transfer. *Journal of the American Chemical Society.* **101**(3), pp.760-762.
- Mullis, K.B. 1990. The unusual origin of the polymerase chain reaction. *Sci Am.* **262**(4), pp.56-61, 64-55.
- Nallamsetty, S. and Waugh, D.S. 2006. Solubility-enhancing proteins MBP and NusA play a passive role in the folding of their fusion partners. *Protein Expr Purif.* **45**(1), pp.175-182.
- Nguyen Sle, T., Quyen, D.T. and Vu, H.D. 2014. Highly effective renaturation of a streptokinase from *Streptococcus pyogenes* DT7 as inclusion bodies overexpressed in *Escherichia coli*. *Biomed Res Int.* **2014**, p324705.
- Nishitoh, H., Matsuzawa, A., Tobiume, K., Saegusa, K., Takeda, K., Inoue, K., Hori, S., Kakizuka, A. and Ichijo, H. 2002. ASK1 is essential for endoplasmic reticulum stress-induced neuronal cell death triggered by expanded polyglutamine repeats. *Genes Dev.* **16**(11), pp.1345-1355.

Ohri, R., Bhakta, S., Fourie-O'Donohue, A., Dela Cruz-Chuh, J., Tsai, S.P., Cook, R., Wei, B., Ng, C., Wong, A.W., Bos, A.B., Farahi, F., Bhakta, J., Pillow, T.H., Raab, H., Vandlen, R., Polakis, P., Liu, Y., Erickson, H., Junutula, J.R. and Kozak, K.R. 2018. High-Throughput Cysteine Scanning To Identify Stable Antibody Conjugation Sites for Maleimide- and Disulfide-Based Linkers. *Bioconjug Chem.* **29**(2), pp.473-485.

Oikawa, D., Kimata, Y., Kohno, K. and Iwawaki, T. 2009. Activation of mammalian IRE1 alpha upon ER stress depends on dissociation of BiP rather than on direct interaction with unfolded proteins. *Exp Cell Res.* **315**(15), pp.2496-2504.

Oikawa, D., Tokuda, M., Hosoda, A. and Iwawaki, T. 2010. Identification of a consensus element recognized and cleaved by IRE1 alpha. *Nucleic Acids Res.* **38**(18), pp.6265-6273.

Okazaki, H., Matsuo, N., Tenno, T., Goda, N., Shigemitsu, Y., Ota, M. and Hiroaki, H. 2018. Using (1) H(N) amide temperature coefficients to define intrinsically disordered regions: An alternative NMR method. *Protein Sci.* **27**(10), pp.1821-1830.

Osorio, F., Tavernier, S.J., Hoffmann, E., Saeys, Y., Martens, L., Veters, J., Delrue, I., De Rycke, R., Parthoens, E., Pouliot, P., Iwawaki, T., Janssens, S. and Lambrecht, B.N. 2014. The unfolded-protein-response sensor IRE-1alpha regulates the function of CD8alpha+ dendritic cells. *Nat Immunol.* **15**(3), pp.248-257.

Ozcan, U., Cao, Q., Yilmaz, E., Lee, A.H., Iwakoshi, N.N., Ozdelen, E., Tuncman, G., Gorgun, C., Glimcher, L.H. and Hotamisligil, G.S. 2004. Endoplasmic reticulum stress links obesity, insulin action, and type 2 diabetes. *Science.* **306**(5695), pp.457-461.

Palmer, I. and Wingfield, P.T. 2004. Preparation and extraction of insoluble (inclusion-body) proteins from Escherichia coli. *Curr Protoc Protein Sci.* **Chapter 6**, pUnit 6.3.

Pandya, C., Uzilov, A.V., Bellizzi, J., Lau, C.Y., Moe, A.S., Strahl, M., Hamou, W., Newman, L.C., Fink, M.Y., Antipin, Y., Yu, W., Stevenson, M., Cavaco, B.M., Teh, B.T., Thakker, R.V., Morreau, H., Schadt, E.E., Sebra, R., Li, S.D., Arnold, A. and Chen, R. 2017. Genomic profiling reveals mutational landscape in parathyroid carcinomas. *JCI Insight.* **2**(6), pe92061.

Parsons, D.W., Jones, S., Zhang, X., Lin, J.C.H., Leary, R.J., Angenendt, P., Mankoo, P., Carter, H., Siu, I.M., Gallia, G.L., Olivi, A., McLendon, R., Rasheed, B.A., Keir, S., Nikolskaya, T., Nikolsky, Y., Busam, D.A., Tekleab, H., Diaz, L.A., Hartigan, J., Smith, D.R., Strausberg, R.L., Marie, S.K.N., Shinjo, S.M.O., Yan, H., Riggins, G.J., Bigner, D.D., Karchin, R., Papadopoulos, N., Parmigiani, G., Vogelstein, B., Velculescu, V.E. and Kinzler, K.W. 2008. An Integrated Genomic Analysis of Human Glioblastoma Multiforme. *Science.* **321**(5897), p1807.

Pervushin, K., Riek, R., Wider, G. and Wuthrich, K. 1997. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci U S A.* **94**(23), pp.12366-12371.

Peti, W. and Page, R. 2007. Strategies to maximize heterologous protein expression in Escherichia coli with minimal cost. *Protein Expr Purif.* **51**(1), pp.1-10.

Pickering, C.R., Zhou, J.H., Lee, J.J., Drummond, J.A., Peng, S.A., Saade, R.E., Tsai, K.Y., Curry, J.L., Tetzlaff, M.T., Lai, S.Y., Yu, J., Muzny, D.M., Doddapaneni, H., Shinbrot, E., Covington, K.R., Zhang, J., Seth, S., Caulin, C., Clayman, G.L., El-Naggar, A.K., Gibbs, R.A., Weber, R.S., Myers, J.N., Wheeler, D.A. and Frederick, M.J. 2014. Mutational landscape of aggressive cutaneous squamous cell carcinoma. *Clin Cancer Res.* **20**(24), pp.6582-6592.

Pincus, D., Chevalier, M.W., Aragon, T., van Anken, E., Vidal, S.E., El-Samad, H. and Walter, P. 2010. BiP binding to the ER-stress sensor Ire1 tunes the homeostatic behavior of the unfolded protein response. *PLoS Biol.* **8**(7), pe1000415.

Plumb, R., Zhang, Z.R., Appathurai, S. and Mariappan, M. 2015. A functional link between the co-translational protein translocation pathway and the UPR. *Elife.* **4**.

Poothong, J., Tirasophon, W. and Kaufman, R.J. 2017. Functional analysis of the mammalian RNA ligase for IRE1 in the unfolded protein response. *Biosci Rep.* **37**(2).

Preissler, S. and Ron, D. 2019. Early Events in the Endoplasmic Reticulum Unfolded Protein Response. *Cold Spring Harb Perspect Biol.* **11**(4).

Prischi, F., Nowak, P.R., Carrara, M. and Ali, M.M.U. 2014. Phosphoregulation of Ire1 RNase. *Nat Commun.* **5**.

Reimold, A.M., Iwakoshi, N.N., Manis, J., Vallabhajosyula, P., Szomolanyi-Tsuda, E., Gravallesse, E.M., Friend, D., Grusby, M.J., Alt, F. and Glimcher, L.H. 2001. Plasma cell differentiation requires the transcription factor XBP-1. *Nature.* **412**(6844), pp.300-307.

Ribeiro, C.M. and Lubamba, B.A. 2017. Role of IRE1 α /XBP-1 in Cystic Fibrosis Airway Inflammation. *Int J Mol Sci.* **18**(1).

Romero-Ramirez, L., Cao, H., Regalado, M.P., Kambham, N., Siemann, D., Kim, J.J., Le, Q.T. and Koong, A.C. 2009. X box-binding protein 1 regulates angiogenesis in human pancreatic adenocarcinomas. *Transl Oncol.* **2**(1), pp.31-38.

Romero, P., Obradovic, Z., Li, X., Garner, E.C., Brown, C.J. and Dunker, A.K. 2001. Sequence complexity of disordered protein. *Proteins.* **42**(1), pp.38-48.

Ron, D. and Walter, P. 2007. Signal integration in the endoplasmic reticulum unfolded protein response. *Nat Rev Mol Cell Biol.* **8**(7), pp.519-529.

Rosano, G.L. and Ceccarelli, E.A. 2014. Recombinant protein expression in Escherichia coli: advances and challenges. *Front Microbiol.* **5**, p172.

Ruan, H., Sun, Q., Zhang, W., Liu, Y. and Lai, L. 2019. Targeting intrinsically disordered proteins at the edge of chaos. *Drug Discov Today.* **24**(1), pp.217-227.

Sanborn, J.Z., Chung, J., Purdom, E., Wang, N.J., Kakavand, H., Wilmott, J.S., Butler, T., Thompson, J.F., Mann, G.J., Haydu, L.E., Saw, R.P., Busam, K.J., Lo, R.S., Collisson, E.A., Hur, J.S., Spellman, P.T., Cleaver, J.E., Gray, J.W., Huh, N., Murali, R., Scolyer, R.A., Bastian, B.C.

and Cho, R.J. 2015. Phylogenetic analyses of melanoma reveal complex patterns of metastatic dissemination. *Proc Natl Acad Sci U S A.* **112**(35), pp.10995-11000.

Sanches, M., Duffy, N.M., Talukdar, M., Thevakumaran, N., Chiovitti, D., Canny, M.D., Lee, K., Kurinov, I., Uehling, D., Al-awar, R., Poda, G., Prakesch, M., Wilson, B., Tam, V., Schweitzer, C., Toro, A., Lucas, J.L., Vuga, D., Lehmann, L., Durocher, D., Zeng, Q., Patterson, J.B. and Sicheri, F. 2014. Structure and mechanism of action of the hydroxy-aryl-aldehyde class of IRE1 endoribonuclease inhibitors. *Nat Commun.* **5**, p4202.

Saunders, M., Wishnia, A. and Kirkwood, J.G. 1957. THE NUCLEAR MAGNETIC RESONANCE SPECTRUM OF RIBONUCLEASE1. *Journal of the American Chemical Society.* **79**(12), pp.3289-3290.

Schein, C.H. 1989. Production of Soluble Recombinant Proteins in Bacteria. *Bio/Technology.* **7**(11), pp.1141-1149.

Scheuner, D. and Kaufman, R.J. 2008. The unfolded protein response: a pathway that links insulin demand with beta-cell failure and diabetes. *Endocr Rev.* **29**(3), pp.317-333.

Schmid, D., Baici, A., Gehring, H. and Christen, P. 1994. Kinetics of molecular chaperone action. *Science.* **263**(5149), pp.971-973.

Schuck, S., Prinz, W.A., Thorn, K.S., Voss, C. and Walter, P. 2009. Membrane expansion alleviates endoplasmic reticulum stress independently of the unfolded protein response. *J Cell Biol.* **187**(4), pp.525-536.

Sepulveda, D., Rojas-Rivera, D., Rodriguez, D.A., Groenendyk, J., Kohler, A., Lebeaupin, C., Ito, S., Urra, H., Carreras-Sureda, A., Hazari, Y., Vasseur-Cognet, M., Ali, M.M.U., Chevet, E., Campos, G., Godoy, P., Vaisar, T., Bailly-Maitre, B., Nagata, K., Michalak, M., Sierralta, J. and Hetz, C. 2018. Interactome Screening Identifies the ER Luminal Chaperone Hsp47 as a Regulator of the Unfolded Protein Response Transducer IRE1alpha. *Mol Cell.* **69**(2), pp.238-252.e237.

Shamu, C.E. and Walter, P. 1996. Oligomerization and phosphorylation of the Ire1p kinase during intracellular signaling from the endoplasmic reticulum to the nucleus. *Embo j.* **15**(12), pp.3028-3039.

Sharma, R.B., O'Donnell, A.C., Stamateris, R.E., Ha, B., McCloskey, K.M., Reynolds, P.R., Arvan, P. and Alonso, L.C. 2015. Insulin demand regulates beta cell number via the unfolded protein response. *J Clin Invest.* **125**(10), pp.3831-3846.

Shen, J., Chen, X., Hendershot, L. and Prywes, R. 2002. ER stress regulation of ATF6 localization by dissociation of BiP/GRP78 binding and unmasking of Golgi localization signals. *Dev Cell.* **3**(1), pp.99-111.

Shen, J. and Prywes, R. 2004. Dependence of site-2 protease cleavage of ATF6 on prior site-1 protease digestion is determined by the size of the luminal domain of ATF6. *J Biol Chem.* **279**(41), pp.43046-43051.

Shen, Y., Maupetit, J., Derreumaux, P. and Tuffery, P. 2014. Improved PEP-FOLD Approach for Peptide and Mini-protein Structure Prediction. *J Chem Theory Comput.* **10**(10), pp.4745-4758.

Siivari, K., Zhang, M., Palmer, A.G., 3rd and Vogel, H.J. 1995. NMR studies of the methionine methyl groups in calmodulin. *FEBS Lett.* **366**(2-3), pp.104-108.

Silvestre-Ryan, J., Bertocini, C.W., Fenwick, R.B., Esteban-Martin, S. and Salvatella, X. 2013. Average conformations determined from PRE data provide high-resolution maps of transient tertiary interactions in disordered proteins. *Biophys J.* **104**(8), pp.1740-1751.

Stevens, F.J. and Argon, Y. 1999. Protein folding in the ER. *Semin Cell Dev Biol.* **10**(5), pp.443-454.

Sun, C., Todorovic, A., Querol-Audi, J., Bai, Y., Villa, N., Snyder, M., Ashchyan, J., Lewis, C.S., Hartland, A., Gradia, S., Fraser, C.S., Doudna, J.A., Nogales, E. and Cate, J.H. 2011. Functional reconstitution of human eukaryotic translation initiation factor 3 (eIF3). *Proc Natl Acad Sci U S A.* **108**(51), pp.20473-20478.

Sun, S., Han, Y., Paramasivam, S., Yan, S., Siglin, A.E., Williams, J.C., Byeon, I.J., Ahn, J., Gronenborn, A.M. and Polenova, T. 2012. Solid-state NMR spectroscopy of protein complexes. *Methods Mol Biol.* **831**, pp.303-331.

Sundaram, A., Plumb, R., Appathurai, S. and Mariappan, M. 2017. The Sec61 translocon limits IRE1 α signaling during the unfolded protein response. *Elife.* **6**.

Tam, A.B., Koong, A.C. and Niwa, M. 2014. Ire1 has distinct catalytic mechanisms for XBP1/HAC1 splicing and RIDD. *Cell Rep.* **9**(3), pp.850-858.

Tameire, F., Verginadis, II and Koumenis, C. 2015. Cell intrinsic and extrinsic activators of the unfolded protein response in cancer: Mechanisms and targets for therapy. *Semin Cancer Biol.* **33**, pp.3-15.

Thevenet, P., Shen, Y., Maupetit, J., Guyon, F., Derreumaux, P. and Tuffery, P. 2012. PEP-FOLD: an updated de novo structure prediction server for both linear and disulfide bonded cyclic peptides. *Nucleic Acids Res.* **40**(Web Server issue), pp.W288-293.

Tirasophon, W., Lee, K., Callaghan, B., Welihinda, A. and Kaufman, R.J. 2000. The endoribonuclease activity of mammalian IRE1 autoregulates its mRNA and is required for the unfolded protein response. *Genes Dev.* **14**(21), pp.2725-2736.

Tomlinson, J.H. and Williamson, M.P. 2012. Amide temperature coefficients in the protein G B1 domain. *J Biomol NMR.* **52**(1), pp.57-64.

Travers, K.J., Patil, C.K., Wodicka, L., Lockhart, D.J., Weissman, J.S. and Walter, P. 2000. Functional and genomic analyses reveal an essential coordination between the unfolded protein response and ER-associated degradation. *Cell.* **101**(3), pp.249-258.

Tugarinov, V., Kanelis, V. and Kay, L.E. 2006. Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy. *Nat Protoc.* **1**(2), pp.749-754.

Uhlen, M., Fagerberg, L., Hallstrom, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, A., Kampf, C., Sjostedt, E., Asplund, A., Olsson, I., Edlund, K., Lundberg, E., Navani,

S., Szgyarto, C.A., Odeberg, J., Djureinovic, D., Takanen, J.O., Hober, S., Alm, T., Edqvist, P.H., Berling, H., Tegel, H., Mulder, J., Rockberg, J., Nilsson, P., Schwenk, J.M., Hamsten, M., von Feilitzen, K., Forsberg, M., Persson, L., Johansson, F., Zwahlen, M., von Heijne, G., Nielsen, J. and Ponten, F. 2015. Proteomics. Tissue-based map of the human proteome. *Science*. **347**(6220), p1260419.

Upton, J.P., Wang, L., Han, D., Wang, E.S., Huskey, N.E., Lim, L., Truitt, M., McManus, M.T., Ruggero, D., Goga, A., Papa, F.R. and Oakes, S.A. 2012. IRE1alpha cleaves select microRNAs during ER stress to derepress translation of proapoptotic Caspase-2. *Science*. **338**(6108), pp.818-822.

Urano, F., Wang, X., Bertolotti, A., Zhang, Y., Chung, P., Harding, H.P. and Ron, D. 2000. Coupling of stress in the ER to activation of JNK protein kinases by transmembrane protein kinase IRE1. *Science*. **287**(5453), pp.664-666.

Uversky, V.N., Oldfield, C.J. and Dunker, A.K. 2008. Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu Rev Biophys*. **37**, pp.215-246.

Volmer, R., van der Ploeg, K. and Ron, D. 2013. Membrane lipid saturation activates endoplasmic reticulum unfolded protein response transducers through their transmembrane domains. *Proc Natl Acad Sci U S A*. **110**(12), pp.4628-4633.

Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J. and Laue, E.D. 2005. The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins*. **59**(4), pp.687-696.

Walter, P. and Ron, D. 2011. The unfolded protein response: from stress pathway to homeostatic regulation. *Science*. **334**(6059), pp.1081-1086.

Wang, L., Perera, B.G., Hari, S.B., Bhatarai, B., Backes, B.J., Seeliger, M.A., Schurer, S.C., Oakes, S.A., Papa, F.R. and Maly, D.J. 2012. Divergent allosteric control of the IRE1alpha endoribonuclease using kinase inhibitors. *Nat Chem Biol*. **8**(12), pp.982-989.

Wang, P., Li, J., Tao, J. and Sha, B. 2018. The luminal domain of the ER stress sensor protein PERK binds misfolded proteins and thereby triggers PERK oligomerization. *J Biol Chem*. **293**(11), pp.4110-4121.

Wang, X.Z., Harding, H.P., Zhang, Y., Jolicoeur, E.M., Kuroda, M. and Ron, D. 1998. Cloning of mammalian Ire1 reveals diversity in the ER stress responses. *Embo j*. **17**(19), pp.5708-5717.

Wieteska, L., Shahidi, S. and Zhuravleva, A. 2017. Allosteric fine-tuning of the conformational equilibrium poises the chaperone BiP for post-translational regulation. *Elife*. **6**.

Wilkinson, B. and Gilbert, H.F. 2004. Protein disulfide isomerase. *Biochim Biophys Acta*. **1699**(1-2), pp.35-44.

Williamson, M.P. 2013. Using chemical shift perturbation to characterise ligand binding. *Prog Nucl Magn Reson Spectrosc*. **73**, pp.1-16.

- Wiseman, R.L., Zhang, Y., Lee, K.P., Harding, H.P., Haynes, C.M., Price, J., Sicheri, F. and Ron, D. 2010. Flavonol activation defines an unanticipated ligand-binding site in the kinase-RNase domain of IRE1. *Mol Cell*. **38**(2), pp.291-304.
- Wu, J., Boghigian, B.A., Myint, M., Zhang, H., Zhang, S. and Pfeifer, B.A. 2010. Construction and performance of heterologous polyketide-producing K-12- and B-derived *Escherichia coli*. *Lett Appl Microbiol*. **51**(2), pp.196-204.
- Xue, Z., He, Y., Ye, K., Gu, Z., Mao, Y. and Qi, L. 2011. A conserved structural determinant located at the interdomain region of mammalian inositol-requiring enzyme 1 α . *J Biol Chem*. **286**(35), pp.30859-30866.
- Yamaguchi, H. and Miyazaki, M. 2014. Refolding techniques for recovering biologically active recombinant proteins from inclusion bodies. *Biomolecules*. **4**(1), pp.235-251.
- Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J. and Zhang, Y. 2015. The I-TASSER Suite: protein structure and function prediction. *Nat Methods*. **12**(1), pp.7-8.
- Ye, J., Rawson, R.B., Komuro, R., Chen, X., Dave, U.P., Prywes, R., Brown, M.S. and Goldstein, J.L. 2000. ER stress induces cleavage of membrane-bound ATF6 by the same proteases that process SREBPs. *Mol Cell*. **6**(6), pp.1355-1364.
- Yorimitsu, T., Nair, U., Yang, Z. and Klionsky, D.J. 2006. Endoplasmic reticulum stress triggers autophagy. *J Biol Chem*. **281**(40), pp.30299-30304.
- Yoshida, H. 2007. ER stress and diseases. *FEBS J*. **274**(3), pp.630-658.
- Yoshida, H., Haze, K., Yanagi, H., Yura, T. and Mori, K. 1998. Identification of the cis-acting endoplasmic reticulum stress response element responsible for transcriptional induction of mammalian glucose-regulated proteins. Involvement of basic leucine zipper transcription factors. *J Biol Chem*. **273**(50), pp.33741-33749.
- Yoshida, H., Matsui, T., Yamamoto, A., Okada, T. and Mori, K. 2001. XBP1 mRNA is induced by ATF6 and spliced by IRE1 in response to ER stress to produce a highly active transcription factor. *Cell*. **107**(7), pp.881-891.
- Zhang, D., Fourie-O'Donohue, A., Dragovich, P.S., Pillow, T.H., Sadowsky, J.D., Kozak, K.R., Cass, R.T., Liu, L., Deng, Y., Liu, Y., Hop, C.E. and Khojasteh, S.C. 2019. Catalytic cleavage of disulfide bonds in small molecules and linkers of antibody- drug conjugates. *Drug Metab Dispos*.
- Zhang, Z., Kuipers, G., Niemiec, L., Baumgarten, T., Slotboom, D.J., de Gier, J.W. and Hjelm, A. 2015. High-level production of membrane proteins in *E. coli* BL21(DE3) by omitting the inducer IPTG. *Microb Cell Fact*. **14**, p142.
- Zhao, R., So, M., Maat, H., Ray, N.J., Arisaka, F., Goto, Y., Carver, J.A. and Hall, D. 2016. Measurement of amyloid formation by turbidity assay-seeing through the cloud. *Biophys Rev*. **8**(4), pp.445-471.

Zhou, J., Liu, C.Y., Back, S.H., Clark, R.L., Peisach, D., Xu, Z. and Kaufman, R.J. 2006. The crystal structure of human IRE1 luminal domain reveals a conserved dimerization interface required for activation of the unfolded protein response. *Proc Natl Acad Sci U S A.* **103**(39), pp.14343-14348.

Zuiderweg, E.R., Bertelsen, E.B., Rousaki, A., Mayer, M.P., Gestwicki, J.E. and Ahmad, A. 2013. Allostery in the Hsp70 chaperone proteins. *Top Curr Chem.* **328**, pp.99-153.