STRUCTURAL AND BIOPHYSICAL CHARACTERISATION OF
PERK KINASE TOWARDS UNDERSTANDING ER STRESS SENSING
AND ACTIVATION OF THE UNFOLDED PROTEIN RESPONSE

by

MARTA CARRARA

A thesis submitted to Imperial College London
in candidature for the degree of Doctor of Philosophy

January 2014
Declaration of Originality

This PhD thesis describes the results of work performed solely by myself in the Centre for Structural Biology, Department of Molecular Biosciences, Imperial College London between April 2010 and January 2014. This thesis has been written in my own words. Any work which is not my own has been appropriately referenced. The results described in this PhD thesis have not been submitted for any other degree, diploma or qualification.

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives license. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the license terms of this work.

Marta Carrara

Imperial College London

January 2014
ABSTRACT

The unfolded protein response (UPR) is a cellular mechanism that detects the accumulation of misfolded proteins within the endoplasmic reticulum (ER). In mammalian cells, the UPR is mediated by three ER-transmembrane proteins: PERK, IRE1 and ATF6. Early studies in the field provided evidence for the role of BiP, the major ER Hsp70 chaperone, in UPR activation by binding to the luminal domains of PERK, IRE1 and ATF6 and maintaining them in an inactive state. However the underlying mechanism of ER stress sensing and UPR activation is not yet understood.

This thesis presents (i) the novel X-ray crystal structure of PERK luminal domain and (ii) a biochemical study of the unconventional interaction between BiP chaperone and the luminal domains of PERK and IRE1 in vitro. Firstly, the structure of PERK luminal domain was solved in two oligomeric states: dimers and tetramers. Compelling evidence is provided for a role of tetramer formation in directing downstream UPR signalling. Secondly, the unprecedented and unconventional direct binding of PERK and IRE1 luminal domains to BiP is demonstrated. The binding surface was mapped to the nucleotide-binding domain (NBD) of BiP. As such, this points away from a substrate-chaperone interaction and rather implies BiP as an explicit UPR signalling component. Upon binding to C41 unfolded protein, BiP is released from PERK and IRE1 luminal domains. BiP dissociation from PERK and IRE1 is known to lead to their activation.

Based on the work presented in this thesis a novel mechanism of ER stress sensing and UPR activation by PERK and IRE1 is proposed. BiP NBD normally interacts with the luminal domains of PERK and IRE1 and represses UPR signalling. During ER stress, binding of unfolded proteins to BiP’s substrate binding domain leads to the dissociation of BiP-luminal domain complexes. As such, the luminal domains are free to intertwine, mediated by an extended α-helix, and form active tetramers competent for cytoplasmic UPR signalling.
To my dad,

for your unquestioning faith in me and endless love
ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr Maruf Ali, for giving me the opportunity to work in his group, for always being supportive, encouraging and willing to help me whilst giving me the freedom to work in my own way. It has been an honour to be your first PhD student. Thank you to my co-supervisor, Dr Stuart Haslam, especially for his personal support at the start of my PhD. I owe a special thank you to Filippo for his help throughout these four years. You always had time to teach me, discuss my work and make sure I didn’t go astray with my experiments. I am indebted to Piotr for his daily support in the lab and for his collaboration with the in vivo experiments of my work.

I am grateful to Professor Xiaodong Zhang, Professor Paul Freemont and their groups for welcoming to the 5th floor. The group has been a source of many friendships as well as my main source of learning (and of lab equipment). A special mention to Ciaran without whom I would not have had a functioning lab to work in. To Andreas for transmitting me a passion for crystallography, for the many stimulating discussions, scientific and not, and for proofreading the crystallography chapters of this thesis – you have saved me from some tricky questions in my viva. To Raj for enduring my incessant questions, for his help with my structure, especially the initial threading of my chain through blobs of electron density – I may not have a structure today if it wasn’t for your patience with me. Thank you to all the beamline scientists at Diamond for their help with data collection, to all the tutors in the workshops I have attended for their invaluable scientific knowledge and to Professor Thomas Schwartz for letting me work in his lab at MIT.

I would have not survived my PhD without my friends. There are too many of you to address individually but a few do deserve a special mention here. Amalia, Myrsini and Adam – we didn’t give MIT enough credit for what a week making posters and playing blindfolded shepherds can achieve. I didn’t think that is how I would meet my best friends. Thank you for your love, your loyalty, your acceptance of me and my craziness. Adam for the science, with musical visions. Amalia and Myrsini, you know and understand me more than anyone else, I can’t imagine my life without you. A special thank you to Magda – miss chatty you are the main reason why I am writing this today. Stephen, it was an unexpected friendship but one that I truly treasure. There is no other person I would turn to when I need to escape
from everything in my life – your happiness is contagious. To all my Erasmus 2008 and 2011 friends, especially Andres, Ines, Giulia, Lukas, Stefan and Jordi. You made sure I had a life outside of the lab and reminded me to never take what London has to offer for granted. Thank you for inviting me to your homes around Europe. These five years would have not been the same without you. Gonzalo, for being able to go back home to a friend after late nights in the lab, for the countless times you took my mind off things. Emma and Lauryn, you are the reason I am still in London. Cara you were my first friend in London, you are like a sister to me. All these years you have always been there for me, even without me needing to say a thing. I love you so much and I can’t wait to keep growing up together.

Finally my most heartfelt thank you goes to my family. To my mum and dad for the decision you made almost 20 years ago – this would have been but just a distant dream if it wasn’t for the opportunities you gave me. You have always believed in me, supported me and been there for me. Thank you for your unconditional love. Mum, you are my rock. You are always the one I turn to when things are though, the one who picks up the pieces, the one who puts me back on the right track. I am sure I have given you a few grey hairs and wrinkles but I hope I have made you proud and made your hard work worthwhile. To my brothers, Andrea and Giovanni, for putting up with a sister like me. You are protective, caring and loyal. You two have the biggest hearts I have ever come across. I hope to grow up to be more like you, those genes seem to have overlooked me. Your sense of humour often drives me up the wall, but you wouldn’t be my brothers if you stopped being kids despite growing up and having amazing achievements of your own. I am very proud of both of you. I know it’s not always evident but I adore you. Never change, you are my heroes.
# TABLE OF CONTENTS

**ABSTRACT** ........................................................................................................................................ iii

**ACKNOWLEDGEMENTS** ...................................................................................................................... v

**TABLE OF CONTENTS** ....................................................................................................................... vii

**LIST OF FIGURES** ............................................................................................................................... x

**LIST OF TABLES** ................................................................................................................................. xiii

**ABBREVIATIONS** ................................................................................................................................. 14

**AMINO ACID ABBREVIATIONS** ........................................................................................................ 18

1 **Introduction** .................................................................................................................................... 19
   1.1 Protein folding, processing and quality control ................................................................. 20
   1.2 Molecular chaperones of the ER .......................................................................................... 24
   1.3 Activation of the Unfolded Protein Response .................................................................. 38
   1.4 UPR signalling ..................................................................................................................... 51
   1.5 UPR pathologies ................................................................................................................ 61

2 **X-ray crystallography theory** ......................................................................................................... 64
   2.1 Protein crystallisation ......................................................................................................... 65
   2.2 X-ray diffraction ............................................................................................................... 68
   2.3 The phase problem ............................................................................................................ 75
   2.4 Structure refinement ......................................................................................................... 78

3 **Perspective** .................................................................................................................................. 80

4 **Methods** ...................................................................................................................................... 82
   4.1 Cloning and expression ...................................................................................................... 83
   4.2 Site-directed mutagenesis ................................................................................................. 84
   4.3 PERK, IRE1 and BiP expression and purification ............................................................ 84
   4.4 C11 expression and purification ...................................................................................... 86
   4.5 Seleno-Met labeling of PERK .......................................................................................... 88
   4.6 Calibration of HiLoad 16/60 Superdex 200 column ............................................................ 88
   4.7 Multi-Angle Light Scattering ........................................................................................... 89
   4.8 Matrix-assisted laser desorption/ionization-time of flight ............................................... 89
   4.9 Native gel shift assays ...................................................................................................... 89
4.10 Chemical cross-linking ................................................................. 90
4.11 Western blotting ........................................................................ 90
4.12 Pull down experiments ................................................................ 91
4.13 Differential Scanning Fluorimetry .................................................. 91
4.14 Microscale Thermophoresis .......................................................... 92
4.15 Isothermal Titration Calorimetry .................................................... 93
4.16 Analytical ultracentrifugation ....................................................... 94
4.17 Crystallisation .............................................................................. 94
4.18 X-ray diffraction data ................................................................... 95
4.19 Phasing ...................................................................................... 96
4.20 Model building and refinement .................................................... 98
4.21 Bioinformatics programs ............................................................. 98

5 PERK, IRE1 and BiP purification ...................................................... 99
5.1 Summary .................................................................................... 100
5.2 Purification of PERK luminal domain for crystallisation ............... 102
5.3 Purification of hPERK and hIRE1 luminal domains for biophysical studies .... 113
5.4 Purification of BiP chaperone ....................................................... 115
5.5 Purification of Cq1 ..................................................................... 117
5.6 Discussion .................................................................................. 120

6 X-ray crystal structure of PERK luminal domain ............................. 122
6.1 Summary .................................................................................... 123
6.2 Introduction ................................................................................ 124
6.3 Crystallisation of PERK (II-IV) proteins ......................................... 125
6.4 PERK (II-IV) native X-ray diffraction data ..................................... 134
6.5 Phasing ...................................................................................... 141
6.6 Model building and structure refinement ...................................... 148
6.7 PERK luminal domain structure ................................................ 152
6.8 Discussion .................................................................................. 165

7 Biophysical characterisation of BiP interaction with H. sapiens PERK and IRE1 luminal domains ................................................................. 170
7.1 Summary .................................................................................... 171
7.2 Introduction ................................................................................ 172
7.3 Direct binding of BiP to the luminal domains of hPERK and hIRE1 .......... 173
LIST OF FIGURES

Figure 1-1. Protein misfolding within the ER ................................................................. 23
Figure 1-2. BiP chaperoning cycle .................................................................................. 29
Figure 1-3. Structure of BiP sub-domains ........................................................................ 31
Figure 1-4. AllostERIC regulation of BiP chaperone ....................................................... 33
Figure 1-5. Proposed mechanisms of ER stress sensing and UPR activation by PERK/IRE1 .... 40
Figure 1-6. X-ray crystal structure of IRE1 luminal domain .............................................. 48
Figure 1-7. Overview of the mammalian UPR .................................................................. 53
Figure 1-8. Translation attenuation by activated PERK ................................................... 55
Figure 1-9. Selective translation of atf4 mRNA ................................................................ 58
Figure 1-10. UPR-mediated apoptosis ............................................................................. 60
Figure 2-1. Crystallisation phase diagram ....................................................................... 67
Figure 2-2. Bragg’s law and Ewald sphere ...................................................................... 70
Figure 2-3. Experimental phasing .................................................................................... 77
Figure 5-1. hPERK luminal domain constructs for X-ray crystallography studies ........... 103
Figure 5-2. C-terminal cleavage of full-length hPERK luminal domain .............. 105
Figure 5-3. Identification of hPERK luminal domain core region .................................. 107
Figure 5-4. hPERK (II-IV) TALON affinity and ion-exchange chromatography ............ 109
Figure 5-5. hPERK (II-IV) SEC ...................................................................................... 110
Figure 5-6. Purification of PERK (II-IV) core luminal domain ......................................... 112
Figure 5-7. BiP FL purification ......................................................................................... 118
Figure 5-8. BiP and Cyt1 purification ............................................................................... 119
Figure 6-1. Optimization of hPERK (II-IV) crystals (Condition A) ............................... 128
Figure 6-2. hPERK (II-IV) microseeding matrix screen (Condition B) ..................... 130
Figure 6-3. mPERK (II-IV) crystals (Conditions C and D) ............................................ 132
Figure 6-4. Native hPERK (II-IV) Condition A crystal diffraction .................................. 136
Figure 6-5. SeMet mPERK (II-IV) Condition D anomalous signal ............................... 144
Figure 6-6. hPERK (II-IV) structure refinement ............................................................. 149
Figure 6-7. hPERK (II-IV) monomer .............................................................................. 153
Figure 6-8. hPERK (II-IV) dimers ............................................................................... 155
Figure 6-9. hPERK (II-IV) Interface 1 ........................................................................... 156
Figure 6-10. hPERK (II-IV) tetramerisation ................................................................. 158
Figure 6-11. hPERK (II-IV) Interface 2 ................................................................. 159
Figure 6-12. Disruption of hPERK (II-IV) Interface 2 ........................................ 161
Figure 6-13. PERK Lobe 2 electrostatic charge ................................................... 163
Figure 6-14. mPERK (II-IV) dimeric structure ....................................................... 164
Figure 6-15. Model for oligomeric state-dependent UPR signalling by PERK .......... 169
Figure 7-1. Typical MST binding experiment ......................................................... 174
Figure 7-2. Optimization of MST experiments ....................................................... 176
Figure 7-3. Interaction of hPERK and hIRE1 luminal domains with BiP sub-domains ... 178
Figure 7-4. Defining the core interaction of hPERK and hIRE1 with BiP NBD ............ 179
Figure 7-5. ITC of BiP FL binding to hPERK (II-IV) and hIRE1 (II-IV) .................... 181
Figure 7-6. Thermal stability of BiP FL upon nucleotide binding ............................ 183
Figure 7-7. Effect of nucleotide in BiP FL binding to hPERK and hIRE1 .................. 185
Figure 7-8. His6-BiP FL pull down of hPERK and hIRE1 luminal domains .............. 187
Figure 7-9. His6-BiP sub-domains pull down of hPERK (II-IV) and hIRE1 (II-IV) .... 188
Figure 7-10. SEC MALS of hPERK (II-IV)-BiP complexes .................................. 191
Figure 7-11. AUC of hPERK (II-IV)-BiP FL complex ........................................... 192
Figure 7-12. Titration of hPERK (II-IV) L200P with BiP FL .................................. 194
Figure 7-13. ΔEspP decreases the thermal stability of hPERK (II-IV) and hIRE1 (II-IV) 196
Figure 7-14. ΔEspP binding to hPERK (II-IV) and hIRE1 (II-IV) ............................ 197
Figure 7-15. Binding of C31 to BiP and luminal domain proteins ........................... 200
Figure 7-16. hPERK (II-IV) and hIRE1 (II-IV) do not bind BiP FL-C31 complexes .... 201
Figure 7-17. C31 promotes dissociation of BiP FL-hPERK (II-IV) and BiP FL-hIRE1 (II-IV) complexes .................................................................................................................. 202
Figure 7-18. Model for BiP-dependent ER stress sensing by PERK and IRE1 luminal domains ................................................................. 208
Figure 8-1. A proposed model for BiP-dependent initiation of UPR signalling by steric hindrance of active PERK and IRE1 tetramer formation .............................. 212
Figure 10-1. pTWO-E and pET-26b plasmid maps ............................................... 232
Figure 10-2. Calibration of HiLoad 16/60 Superdex 200 column ............................ 233
Figure 10-3. PERK and IRE1 luminal domain sequence and secondary structure alignment . 234
Figure 10-4. hPERK luminal domain disorder prediction ....................................... 235
Figure 10-5. Wilson Plot of hPERK (II-IV) Condition A crystals diffraction data ........ 236
Figure 10-6. hPERK (II-IV) and mPERK (II-IV) SeMet incorporation ..................... 237
Figure 10-7. hPERK (II-IV) heavy atom binding ................................................................. 238
Figure 10-8. Location of Na₂WO₄ heavy atom sites ........................................................... 239
Figure 10-9. Comparison of PERK and IRE1 luminal domain structures ......................... 240
Figure 10-10. PERK and IRE1 structure alignment .......................................................... 241
Figure 10-11. BiP and ΔEspP heat of dilution in ITC experiments ....................................... 242
Figure 10-12. Gel shift assays of hPERK (II-IV)-BiP FL complexes ................................. 243
Figure 10-13. ΔEspP binding to BiP FL ............................................................................. 244
Figure 10-14. Oligomeric state of hPERK (II-IV) and hIRE1 (II-IV) in complex with ΔEspP ...... 245
Figure 10-15. ΔEspP does not induce dissociation of BiP FL-hPERK (II-IV) and BiP FL-hIRE1 (II-IV) complexes ........................................................................................................ 246


**LIST OF TABLES**

Table 1-1. Molecular chaperones of the ER................................................................. 25
Table 2-1. Crystal space groups ..................................................................................... 74
Table 4-1. Purification buffers composition ..................................................................... 87
Table 5-1. PERK, IRE1, BiP and C\textsubscript{H}1 protein constructs .................................. 101
Table 5-2. hPERK and hIRE1 sub-regions I to V boundaries ........................................ 114
Table 6-1. PERK luminal domain crystallisation conditions ........................................... 126
Table 6-2. Native hPERK (II-IV) Condition A diffraction data ........................................ 137
Table 6-3. Native hPERK (II-IV) Condition B and mPERK (II-IV) Condition C and D diffraction data........................................................................................................ 140
Table 6-4. Na\textsubscript{2}WO\textsubscript{4} derivative hPERK (II-IV) Condition A MAD diffraction data............. 147
Table 6-5. hPERK (II-IV) and mPERK (II-IV) refinement ................................................ 151
Table 10-1. Cloning of PERK, IRE1, BiP and C\textsubscript{H}1 proteins ....................................... 227
Table 10-2. hPERK (II-IV) condition A crystals heavy atom datasets ............................ 229
Table 10-3. mPERK (II-IV) condition D crystals heavy atom datasets ......................... 230
Table 10-4. Hydrogen bonding of PERK luminal domain Interface 1 and 2 ................. 231
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>'</td>
<td>minute</td>
</tr>
<tr>
<td>&quot;</td>
<td>second</td>
</tr>
<tr>
<td>χ</td>
<td>chi</td>
</tr>
<tr>
<td>φ</td>
<td>phi</td>
</tr>
<tr>
<td>λ</td>
<td>wavelength</td>
</tr>
<tr>
<td>ρ</td>
<td>electron density</td>
</tr>
<tr>
<td>Å</td>
<td>Ångström</td>
</tr>
<tr>
<td>A260</td>
<td>absorbance at 260 nm</td>
</tr>
<tr>
<td>A280</td>
<td>absorbance at 280 nm</td>
</tr>
<tr>
<td>AAV</td>
<td>adeno-associated virus</td>
</tr>
<tr>
<td>ADP</td>
<td>adenosine diphosphate</td>
</tr>
<tr>
<td>AMP</td>
<td>adenosine monophosphate</td>
</tr>
<tr>
<td>Amp</td>
<td>ampicillin</td>
</tr>
<tr>
<td>AMPPNP</td>
<td>adenosine 5'-(b-g-imido)triphosphate</td>
</tr>
<tr>
<td>ASNS</td>
<td>asn synthetase</td>
</tr>
<tr>
<td>ATF3</td>
<td>activating transcription factor 3</td>
</tr>
<tr>
<td>ATF4</td>
<td>activating transcription factor 4</td>
</tr>
<tr>
<td>ATF6</td>
<td>activating transcription factor 6</td>
</tr>
<tr>
<td>ATP</td>
<td>adenosine triphosphate</td>
</tr>
<tr>
<td>ATPase</td>
<td>adenosine triphosphatase</td>
</tr>
<tr>
<td>Au</td>
<td>gold</td>
</tr>
<tr>
<td>AUC</td>
<td>analytical ultracentrifugation</td>
</tr>
<tr>
<td>BAG</td>
<td>Bcl2-associated athanogene</td>
</tr>
<tr>
<td>BAP</td>
<td>BiP-associated protein</td>
</tr>
<tr>
<td>BBF2</td>
<td>box B-binding factor 2</td>
</tr>
<tr>
<td>BBF2H7</td>
<td>BBF2 human homolog on chromosome 7</td>
</tr>
<tr>
<td>Bcl2</td>
<td>B-cell lymphoma 2</td>
</tr>
<tr>
<td>BiP</td>
<td>immunoglobulin binding protein</td>
</tr>
<tr>
<td>bZIP</td>
<td>basic leucine zipper</td>
</tr>
<tr>
<td>Ca</td>
<td>calcium</td>
</tr>
<tr>
<td>Cam</td>
<td>chloramphenicol</td>
</tr>
<tr>
<td>CCall</td>
<td>overall correlation coefficient</td>
</tr>
<tr>
<td>CCanom</td>
<td>anomalous correlation coefficient</td>
</tr>
<tr>
<td>CCD</td>
<td>charge-coupled detector</td>
</tr>
<tr>
<td>CD4</td>
<td>cluster of differentiation 4</td>
</tr>
<tr>
<td>cDNA</td>
<td>complementary DNA</td>
</tr>
<tr>
<td>C\text{H}\text{Im}</td>
<td>immunoglobulin heavy chains</td>
</tr>
<tr>
<td>C\text{H}1</td>
<td>first constant domain of immunoglobulin heavy chain</td>
</tr>
<tr>
<td>CHOP</td>
<td>CAAT/enhancer-binding protein homologous protein</td>
</tr>
<tr>
<td>C\text{L}</td>
<td>immunoglobulin constant light chain</td>
</tr>
<tr>
<td>CNX</td>
<td>calnexin</td>
</tr>
<tr>
<td>CPY</td>
<td>misfolded carboxypeptidase Y</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>CPY*</td>
<td>constitutively misfolded CPY</td>
</tr>
<tr>
<td>CREB-H</td>
<td>cyclic AMP responsive element-binding protein hepatocyte specific</td>
</tr>
<tr>
<td>CRT</td>
<td>calreticulin</td>
</tr>
<tr>
<td>DMSO</td>
<td>dimethyl sulphoxide</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>DNase</td>
<td>deoxyribonuclease</td>
</tr>
<tr>
<td>DSF</td>
<td>differential scanning fluorimetry</td>
</tr>
<tr>
<td>DTT</td>
<td>dithiothreitol</td>
</tr>
<tr>
<td>ECL</td>
<td>enhanced chemiluminescence</td>
</tr>
<tr>
<td>EDEM</td>
<td>ER-degradation-enhancing a-mannosidase-like protein</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>EGS</td>
<td>ethylene glycolbis(succinimidylsuccinate)</td>
</tr>
<tr>
<td>eIF2α</td>
<td>eukaryotic Initiation Factor 2α</td>
</tr>
<tr>
<td>ER</td>
<td>endoplasmic reticulum</td>
</tr>
<tr>
<td>ERAD</td>
<td>ER associated degradation</td>
</tr>
<tr>
<td>ERdj</td>
<td>ER localised DnaJ proteins</td>
</tr>
<tr>
<td>ERSE</td>
<td>ER stress response element</td>
</tr>
<tr>
<td>FEM</td>
<td>feature enhanced maps</td>
</tr>
<tr>
<td>FL</td>
<td>full-length</td>
</tr>
<tr>
<td>FOM</td>
<td>figure of merit</td>
</tr>
<tr>
<td>G1</td>
<td>gap 1</td>
</tr>
<tr>
<td>GADD34</td>
<td>growth arrest and DNA damage 34</td>
</tr>
<tr>
<td>GCN2</td>
<td>general control non-depressible-2</td>
</tr>
<tr>
<td>Grp78</td>
<td>glucose-regulated protein 78 kDa</td>
</tr>
<tr>
<td>GTP</td>
<td>guanosine triphosphate</td>
</tr>
<tr>
<td>h</td>
<td>hour</td>
</tr>
<tr>
<td>H2O2</td>
<td>hydrogen peroxide</td>
</tr>
<tr>
<td>HAC1</td>
<td>homologous to Atf/Creb1</td>
</tr>
<tr>
<td>HERP</td>
<td>homocysteine-responsive ER-resident protein</td>
</tr>
<tr>
<td>Hg</td>
<td>mercury</td>
</tr>
<tr>
<td>HRD1</td>
<td>ERAD-associated E3 ubiquitin-protein ligase</td>
</tr>
<tr>
<td>HRI</td>
<td>haem-regulated inhibitor</td>
</tr>
<tr>
<td>HRP</td>
<td>horseradish peroxidase</td>
</tr>
<tr>
<td>Hsp</td>
<td>heat shock protein</td>
</tr>
<tr>
<td>HspBP1</td>
<td>Hsp binding protein 1</td>
</tr>
<tr>
<td>I3C</td>
<td>5-amino-2,4,6-triiodoisophthalic acid</td>
</tr>
<tr>
<td>IPTG</td>
<td>isopropyl β-D-1-thiogalactopyranoside</td>
</tr>
<tr>
<td>IR</td>
<td>infrared</td>
</tr>
<tr>
<td>IRE1p</td>
<td>inositol requiring kinase 1</td>
</tr>
<tr>
<td>ITC</td>
<td>isothermal titration calorimetry</td>
</tr>
<tr>
<td>K</td>
<td>potassium</td>
</tr>
<tr>
<td>Kan</td>
<td>kanamycin</td>
</tr>
<tr>
<td>Kras2</td>
<td>kirsten rat sarcoma-2 viral</td>
</tr>
<tr>
<td>LB</td>
<td>luria-bertani</td>
</tr>
<tr>
<td>LED</td>
<td>light emitting diode</td>
</tr>
<tr>
<td>LLG</td>
<td>log-likelihood gain</td>
</tr>
</tbody>
</table>
SAD  single anomalous dispersion
SBD  substrate-binding domain
SDS-PAGE  sodium dodecyl sulphate polyacrylamide gel electrophoresis
Se  selenium
SE  standard error
SEC  size exclusion chromatography
SERp  surface entropy reduction prediction
SIR  single isomorphous replacement
SIRAS  single isomorphous replacement with anomalous scattering
spFRET  single-pair Förster resonance energy transfer
SPR  surface plasmon resonance
SSM  secondary-structure matching
T-coffee  tree based consistency objective function for alignment evaluation
TFZ  translation function Z-score
Tisp40  transcript induced in spermiogenesis-40
Tm  melting temperature
TRAF2  TNF receptor associated factor 2
U  units
UGT1  UDP-glucuronosyltransferase 1
uORF  upstream open reading frames
UPR  unfolded protein response
UPRE  UPR element
UV  ultraviolet
V0  void volume
Ve  elution volume
Vm  matthews coefficient
Vs  solvent content
VSVG  vesicular stomatitis virus G
W  tungsten
WRS  wolcott-rallison syndrome
XBP1  bZIP-containing X-box-binding protein 1
XBP1s  spliced XBP1
Zn  zinc
# AMINO ACID ABBREVIATIONS

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Abbreviation</th>
<th>One Letter Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alanine</td>
<td>Ala</td>
<td>A</td>
</tr>
<tr>
<td>Arginine</td>
<td>Arg</td>
<td>R</td>
</tr>
<tr>
<td>Asparagine</td>
<td>Asn</td>
<td>N</td>
</tr>
<tr>
<td>Aspartate</td>
<td>Asp</td>
<td>D</td>
</tr>
<tr>
<td>Cysteine</td>
<td>Cys</td>
<td>C</td>
</tr>
<tr>
<td>Glutamate</td>
<td>Glu</td>
<td>E</td>
</tr>
<tr>
<td>Glutamine</td>
<td>Gln</td>
<td>Q</td>
</tr>
<tr>
<td>Glycine</td>
<td>Gly</td>
<td>G</td>
</tr>
<tr>
<td>Histidine</td>
<td>His</td>
<td>H</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>Ile</td>
<td>I</td>
</tr>
<tr>
<td>Leucine</td>
<td>Leu</td>
<td>L</td>
</tr>
<tr>
<td>Lysine</td>
<td>Lys</td>
<td>K</td>
</tr>
<tr>
<td>Methionine</td>
<td>Met</td>
<td>M</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>Phe</td>
<td>F</td>
</tr>
<tr>
<td>Proline</td>
<td>Pro</td>
<td>P</td>
</tr>
<tr>
<td>Serine</td>
<td>Ser</td>
<td>S</td>
</tr>
<tr>
<td>Threonine</td>
<td>Thr</td>
<td>T</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>Trp</td>
<td>W</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>Tyr</td>
<td>Y</td>
</tr>
<tr>
<td>Valine</td>
<td>Val</td>
<td>V</td>
</tr>
</tbody>
</table>
1 Introduction
1.1 Protein folding, processing and quality control

Eukaryotic secretory and membrane proteins attain their functional and assembled conformation in the endoplasmic reticulum (ER) prior to transit to the Golgi apparatus. Nascent polypeptide chains co-translationally transverse the ER membrane though the hydrophilic interior of the Sec61 translocon complex (SecY in prokaryotes), a protein conducting channel [1]. Once located in the ER lumen, a complex system of molecular chaperones, folding enzymes and co-factors assist in the folding, post-translational modification, quality control and trafficking of the newly synthesized polypeptide chains [2]. The cycle of a typical misfolded protein in the ER is discussed in this chapter and summarised in Figure 1-1.

1.1.1 Folding of nascent polypeptide chains

In vitro studies carried out in the early 1960s by Anfinsen and co-workers revealed that protein folding requires solely the information contained in the amino acid sequence [3]. The process of protein folding has been difficult to study. Most eukaryotic proteins fold on a millisecond or even microsecond time scale, which makes them hard to capture and examine kinetically and/or structurally. Chain folding is often started by the burial of non-polar residues or patches found in the core of proteins as a means of minimizing free surface energies [4]. Folding occurs in a heterogeneous and co-operative fashion. It is directed by several factors including predetermined folding units, or foldons, sequential stabilization of folding intermediates and misfolding errors.

Although nascent polypeptide chains sample several metastable folding intermediates, some defined factors limit the number of conformations available for them to trial. Constraints include (i) defined hydrophilic interactions (salt bridges and disulphide bonds), (ii) physical restrictions (vectorial folding from N- to C-terminus, tethering of the N-terminal signal sequence to the ER membrane and the width of the ribosomal tunnel and Sec61 channel which can accommodate limited secondary structures), and (iii) timing of events (usage of rare codons and ribosome stalling which affect translation speed, cleavage of the signal sequence and disulphide bond formation to control protein maturation and secretion) [5-7].
Since Anfinsen’s initial studies, a multitude of helper molecules and post-translational processes that edit and ensure the folding process have been uncovered. However, his thermodynamic theory stating that protein chains will always fold until the lowest energy (native) conformation is achieved remains undisputed. Some exceptions are known, such as amyloids, which can form even lower energy conformations as macromolecular aggregates [4].

1.1.2 Processing of folding polypeptides

1.1.2.1 Chain folding and post-translational modifications in the ER

In the ER lumen molecular chaperones, folding and glycosylation enzymes assist in the processing of the nascent polypeptide chains. Firstly, chaperones prevent aggregation of unfolded chains, facilitate protein maturation and retain folding proteins in the ER until they reach their functional and assembled conformation. Chaperones are discussed in more detail in Chapter 1.2. Secondly, folding enzymes, or foldases, accelerate the kinetics of protein folding. These mainly involve protein folding isomerases (PDI) and peptidyl-prolyl cis-trans isomerases (PPI) [2]. PDIs catalyse the formation of disulphide bonds between Cysteine residues of proteins. PPIs assist in the interconversion of cis and trans isomers of peptide bonds with Proline residues. Finally, protein glycosylation occurs in all three domains of life, with prokaryotic glycosylation being a relatively recent discovery [8]. It is estimated that approximately half of all mammalian proteins are glycoproteins [9]. Glycosylation involves the attachment of sugar molecules via glycosidic bonds to the nitrogen atom of Asparagine side chains (N-linked) or oxygen atom of Serine/Threonine side chains (O-linked). Glycosylation is fundamental for correct protein folding [9]. Firstly, the attached oligosaccharides occupy a large volume on the surface of proteins and can shield from surrounding proteins thus acting as chaperones. Secondly, they stabilize protein conformations by interacting with the peptide backbone or by increasing solubility due to their hydrophilic nature. Thirdly, the sequential trimming of the terminal glucose residues allows for the calnexin (CNX)/calreticulin (CRT) quality-control cycle [10]. The CNX/CRT cycle ensures that only proteins that reach their fully folded native structure and macromolecular assembly are exported to their respective cellular compartment. Misfolded proteins are directed towards degradation pathways.
1.1.2.2 ER associated degradation

Despite the numerous mechanisms that exist to ensure proper folding and processing, proteins may fail to reach their folded and functional conformation. Within the ER, terminally misfolded or unassembled N-linked glycoproteins are diverted from the CNX/CRT cycle and are targeted for ER associated degradation (ERAD) [2,11]. The main mechanism by which non-glycosylated misfolded proteins are directed towards ERAD is by interaction with the ER chaperone Immunoglobulin Binding Protein (BiP), which recognises their unfolded regions [12]. In addition, homocysteine-responsive ER-resident protein (HERP) and ER-degradation-enhancing α-mannosidase-like protein (EDEM) have been shown to divert non-glycosylated misfolded proteins towards ERAD pathways [13,14].

After recognition and targeting for ERAD, misfolded proteins are retrotranslocated, or dislocated, from the ER lumen to the cytosol. At the cytosolic face of the ER p97 protein directs the dislocation of the misfolded substrate [15]. p97 serves as a ratchet to pull misfolded polypeptide chains across the ER membrane, using adenosine triphosphate (ATP) hydrolysis as the driving force. Moreover, p97 also acts as a platform to which several ERAD-related enzymes are recruited. These include ubiquitin-chain modifying enzymes, chain elongation factors and deubiquitinases that all assist in the polyubiquitination of misfolded proteins [9]. Polyubiquitination consists in the successive addition of ubiquitin molecules onto Lysine residues by ubiquitin-activating, -conjugating and -ligating enzymes. Polyubiquitinated proteins are transferred to the 26S proteasome by shuttle proteins where peptidases proteolytically cleave the unfolded substrates into single amino acids that can be recycled for new protein biosynthesis [16].
After translocation of nascent polypeptide chains through the Sec61 channel, and possible modification by PPI and PDI enzymes, improperly folded proteins are recognised by BiP Hsp70 chaperone. BiP assists in the folding of misfolded proteins. Many nascent chains also undergo N-glycosylation (branched sugar moiety shown). N-glycosylation allows proteins to enter the CNX/CRT cycle. Removal of the terminal glucose by Glucosidase II allows exit from the CNX/CRT cycle. Misfolded glycoproteins can re-enter the CNX/CRT cycle through addition of the terminal glucose by UDP-glucuronosyltransferase 1 (UGT1). Terminally misfolded proteins are directed towards ERAD pathways through removal of a mannose residue by EDEM. Proteins are dislocated from the ER into the cytosol (putatively through Sec61, HRD1 and Derlin-1 channels). In the cytosol, p97 pulls the misfolded proteins across the ER membrane. Polyubiquitination of misfolded proteins targets them to the 26S proteasome where they are proteolytically cleaved into single amino acids. These are recycled and used in the synthesis of new polypeptide chains by the translating ribosome.
1.2 Molecular chaperones of the ER

Cellular compartments in which synthesis or translocation of proteins occurs, which include the ER, contain a high concentration of molecular chaperones. These have been highly conserved throughout evolution and mainly belong to the heat shock protein (Hsp) families but also include lectins and ribosome-associated chaperones [10]. Hsps are found in all living organisms from prokaryotes to higher eukaryotes and constitute up to 5% of cellular proteins. Their roles vary from intracellular chaperoning and housekeeping functions to mediating immune and stress responses. Hsps have been grouped into families according to their molecular weight and sequence similarities: small Hsp, Hsp40, Hsp60 (chaperonins), Hsp70, Hsp90, and Hsp100. Molecular chaperones of the ER are summarised in Table 1-1.

1.2.1 Hsp70 chaperones and BiP

Hsp70s are a highly conserved and ubiquitous class of ATP-regulated chaperones that play an important role in proteostasis [17]. Cellular functions of Hsp70s include chaperoning nascent protein chains, assisting in protein import into organelles and disassembly of macromolecular complexes and aggregates. Hsp70s are essential for survival during stress conditions. Their functions are mediated by Hsp70 interaction with the extended hydrophobic regions of substrate misfolded proteins. Sequence analysis of peptides identified as Hsp70 substrates revealed a hydrophobic binding motif composed of a heptapeptide sequence rich in Tryptophan, Phenylalanine and Leucine residues [18]. Protein folding defects are thus monitored through Hsp70 chaperone interaction with aromatic and hydrophobic residues that are normally found in the core of correctly folded and assembled proteins. Hsp70s recognise and transiently bind to these misfolded or non-native protein substrates. They assist in their folding mainly by preventing their irreversible aggregation thus allowing folding and assembly.
# Molecular Chaperones of the ER

The name and major functions of the yeast and mammalian orthologues of the ER chaperones are listed.

<table>
<thead>
<tr>
<th>Family</th>
<th>Protein ID</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hsp40</td>
<td>Mtj1p (yeast); ERdj1 (mammals)</td>
<td>ER transmembrane protein which inhibits entry of nascent polypeptides unless its ER luminal domain is bound to Kar2/BiP; serves as a translational check-point</td>
</tr>
<tr>
<td></td>
<td>Sec63p (yeast); ERdj2 (mammals)</td>
<td>ER transmembrane protein involved in SRP-independent translocation of nascent polypeptides into the ER</td>
</tr>
<tr>
<td></td>
<td>Scj1p, HEDJ (yeast); ERdj3 (mammals)</td>
<td>Promoting BiP chaperoning activity; delivery of substrates to Kar2/BiP; upregulated during ER stress</td>
</tr>
<tr>
<td></td>
<td>MDG-1p (yeast); ERdj4 (mammals)</td>
<td>ER-membrane anchored protein involved in ERAD; upregulated during ER stress</td>
</tr>
<tr>
<td></td>
<td>JPDip (yeast); ERdj5 (mammals)</td>
<td>Promotes ERAD of non-glycosylated misfolded proteins. Has four thioredoxin domains which break disulfide bonds in substrate proteins to facilitate their retrotranslocation.</td>
</tr>
<tr>
<td></td>
<td>ERdj6/pS8IPK (mammals)</td>
<td>Promoting BiP chaperoning activity; down-regulation of PERK phosphorylation</td>
</tr>
<tr>
<td></td>
<td>ERdj7 (mammals)</td>
<td>ER transmembrane protein; function is still unknown</td>
</tr>
<tr>
<td>Hsp70</td>
<td>Kar2p (yeast); BIP/Grp78/HspA (mammals)</td>
<td>Entry of nascent polypeptides into the ER; binding to unassembled/misfolded proteins, maintenance ER calcium level; regulation of the Unfolded Protein Response</td>
</tr>
<tr>
<td>Hsp90</td>
<td>Grp94/ERp99 (mammals)</td>
<td>Folding solely of immunoglobulin heavy chains, integrin, insulin-like growth factor I and II and toll-like receptors; only present in metazoans</td>
</tr>
<tr>
<td>Hsp100</td>
<td>Torsin A (mammals)</td>
<td>Function still unknown</td>
</tr>
<tr>
<td>Unclassified (Hsp110-homologs)</td>
<td>Lhs1 (yeast); HspH4 (mammals)</td>
<td>Nucleotide exchange factor for Kar2/BiP</td>
</tr>
<tr>
<td></td>
<td>Sil1 (yeast); BAP (mammals)</td>
<td>Nucleotide exchange factor for Kar2/BiP</td>
</tr>
<tr>
<td></td>
<td>Grp170 (mammals)</td>
<td>Nucleotide exchange factor for Kar2/BiP</td>
</tr>
</tbody>
</table>
DnaK is the canonical Hsp70 of *Escherica coli* and has been used for a large part of functional and structural studies of Hsp70s. Given the high sequence conservation of *E. coli* DnaK with mammalian Hsp70s (51% sequence identity between DnaK and *Homo sapiens* BiP), these studies are likely to also be relevant to the eukaryotic orthologues.

The major chaperone of the ER, often referred to as its master regulator, is the Hsp70 BiP (Kar2p in *Saccharomyces cerevisiae*) [12,19]. BiP is also known as the glucose-regulated protein 78 kilo Dalton (kDa) (Grp78) or HspA5. BiP was first identified in stable complexes with unassembled immunoglobulin heavy chains (Ci), hence its name [20]. More specifically, the first constant domain of immunoglobulin heavy chains (C1) represents the main BiP binding site. On its own, C1 is intrinsically unfolded and it becomes structured upon binding of its cognate partner, the immunoglobulin constant light chain (C1) [21]. Unfolded C1 stably binds to BiP in the presence of adenosine diphosphate (ADP) with an affinity in the low micromolar range [22-26]. In the assembled antibody, Ci and C1 are covalently linked via a disulphide bridge [25]. Once this disulphide bridge is formed, BiP can no longer associate with C1. Indeed, C1 oxidation inhibits its association with BiP. Additionally, BiP does not cycle on and off from unassembled C1 but requires the presence of Ci for release from BiP [23]. This has been shown *in vivo* by immunoprecipitation experiments, which used BiP adenosine trisphosphatase (ATPase) mutants as kinetic traps. Together these results demonstrate that C1 is a specific substrate for BiP chaperoning activity. C1 and C1-derived peptides have been used for many studies to characterise the interaction of BiP with unfolded substrates.

In addition to protein folding, BiP also plays a role in (i) entry of nascent chains into the ER lumen, (ii) homeostasis of ER calcium (Ca2+), (iii) delivery of misfolded substrates for ERAD and (iv) mediating stress responses. The latter is thoroughly discussed in Chapter 1.3. Firstly, BiP assists in the entry of newly synthesized polypeptides into the ER lumen. This is achieved by opening of the Sec61 channel, co- and post-translational insertion of the nascent polypeptide into the translocation complex [27]. BiP also acts as a molecular ratchet to pull the growing chain across the ER membrane. Secondly, BiP minimizes efflux of Ca2+ from the ER by occluding the Sec61 pore when it is not being used [28]. Finally, the role of BiP in ERAD remains somewhat elusive. In yeast expression of mutant Kar2p forms has been shown to reduce degradation of ERAD substrates [29,30]. Thus, it was first speculated that BiP plays a
role in the selection of ERAD substrates. Since, hypothesised roles for BiP in ERAD include misfolded protein ‘solubilisation’, or prevention of aggregation, which would allow them to fit through the retrotranslocation channel [27,31]. Alternatively BiP may deliver substrates to the Sec61 channel. However, the role for BiP in ERAD has always been an interpretation of results rather than a direct observation. In addition, since the same BiP regions are involved in binding both ERAD substrates and the Sec61 channel the accuracy of the proposed models has further been questioned [31].

BiP knockout mice exhibit severe developmental defects and are embryonically lethal at day 3.5 [32]. Abnormal BiP function has been linked to many disease states including infectious diseases, inherited syndromes, neurological disorders and tumours [33]. To name a few examples, BiP is fundamental for the initiation and progression of cancerous cells in fibrosarcoma and mammary tumour models [34,35]. Experimentally, BiP can protect breast cancer cells against ER stress-induced apoptosis, whereas inhibition of its function sensitizes glioma cells to chemotherapy treatments [36,37]. Clearly BiP is an essential ER component and plays a critical role in cellular homeostasis. Further characterising its mechanism of action is fundamental and will aid in its exploitation as a drug development target, in particular for chemotherapy.
1.2.2 Conformational cycling of Hsp70s

1.2.2.1 Structure of Hsp70 sub-domains

Hsp70 proteins are composed of an N-terminal nucleotide-binding domain (NBD) connected by a flexible hydrophobic linker to a C-terminal substrate-binding domain (SBD). The affinity and kinetics of substrate binding to the SBD is modulated by the catalytic ATPase activity of the NBD [38]. In their natural resting state, Hsp70s exist in an ATP-bound open form and have low affinity for unfolded substrates. Upon substrate binding to the SBD, hydrolysis of the bound ATP induces conformational changes to confer high affinity for the bound peptide.

The ATPase activity of Hsp70s is key for their function. This has been highlighted in a multitude of biochemical and cellular published studies, only some of which are mentioned here. Substrate binding stimulates ATP hydrolysis by DnaK approximately 10-fold; similar rates have been measured for BiP [22,39,40]. Indeed, BiP interaction with substrates is dependent on the presence of ADP. Addition of ATP decreases affinity for substrates by 10 to 50-fold and can stimulate the release of bound substrates [18,23,38,41]. Functional mutations in the SBD affect Hsp70 ATPase activity, whereas mutations of key catalytic NBD residues inhibit the release of bound substrates and secretion of assembled immunoglobulins [21,42-44]. Although the functions of Hsp70s are biochemically well defined, recent structural studies have been key to understanding Hsp70 allosteric cycling and how this affects their chaperoning functions. Figure 1-2 illustrates the chaperoning cycle of BiP discussed in detail in this chapter.
Figure 1-2. BiP chaperoning cycle

(1) In its ‘resting’ ATP-bound substrate-free state, BiP’s NBD (pink) and SBD (teal) are docked and confer low substrate affinity. (2) Upon binding of misfolded protein substrates (dark purple), hydrolysis of the bound ATP induces a conformational change in BiP. The delivery of misfolded proteins and ATP hydrolysis can be facilitated by Hsp40 co-chaperones (orange), which bind to BiP’s SBD. (3) In the ADP-bound state the SBD adopts a closed lid conformation, which traps the bound peptide and confers high substrate affinity. BiP assists in folding of the misfolded protein. (4) Once the native conformation is reached, the folded protein exits the cycle. (5) Nucleotide exchange factors (NEF) (purple) bind to BiP’s NBD and catalyse exchange of ADP for ATP so that BiP returns to its ‘resting’ state. In the diagram, grey shading of BiP refers to its conformation in the previous step.
In the past two decades, many structures of Hsp70 NBDs and SBDs have been solved by X-ray crystallography and Nuclear Magnetic Resonance (NMR). These have been captured in various functional conformations. Overall, these structures are consistent as to their structural arrangement and support the known mechanisms of ATP binding/hydrolysis and substrate capture/release of Hsp70 chaperones.

Hsp70 NBD consists of two large globular sub-domains (NBD-I and NBD-II) that are further subdivided into two smaller regions (A and B for each sub-domain) (Figure 1-3A) [17]. Nucleotides bind in complex with one magnesium (Mg$^{2+}$) and two potassium (K$^+$) ions in a crevice formed between NBD-I and NBD-II. The structures of the NBD in the apo, ATP, ADP, and adenosine 5’-(β-γ-imido)triphosphate (AMPPNP) (a non-hydrolysable ATP analogue)-bound forms have been solved [45-47]. Surprisingly, they are all very similar and they do not reveal any nucleotide-dependent conformational change. This may be due to the transiency of some of the intermediate states and/or their dependence on the adjacent SBD for the correct structural assembly. In fact, the NBD is joined to the SBD by a highly conserved linker. Interestingly, when the SBD was crystallised together with the linker, the existence of two structural variants was revealed [48]. This proposed a role for the SBD in the ATPase cycle and allosteric regulation of Hsp70s, and not exclusively in substrate binding, for the first time.

The SBD is composed of a compact β-sandwich domain (SBD-β) that contains a hydrophobic substrate-binding cleft and a C-terminal α-helical domain (SBD-α) termed the lid (Figure 1-3B) [17,48]. The unfolded substrates bind to two loops of SBD-β (L1,2 and L3,4) through extensive hydrogen bonding and van der Waals interactions. Hydrophobic interactions between two outer loops of SBD-β (L1,2 and L5,6) and the helices αA and αB of SBD-α allows for movement of the lid, which opens and closes the substrate-binding cavity.
BiP is a Hsp70 chaperone composed of an N-terminal NBD and a C-terminal SBD. A highly conserved linker connects the two domains.

(A) X-ray crystal structure of ATP-bound BiP NBD (PDB: 3LDL) [45]. The NBD consists of two sub-domains (I and II), which are both further divided into two regions (A and B). Nucleotides bind between NBD-I and NBD-II by coordination with one Mg$^{2+}$ and two K$^+$ ions (no density for these observed in this crystal structure).

(B) X-ray crystal structure of substrate-bound DnaK SBD (PDB: 1DKX) [48]. The SBD consists of a substrate binding β-sandwich (SBD-β) and an α-helical lid (SBD-α). Hydrophobic interactions between SBD-αB and two SBD-β loops (L3,4 and L5,6) mediate docking of the two sub-domains. The main residue side chains involved are shown. This closes the SBD-α lid to trap the substrate between two SBD-β loops (L1,2 and L3,4).

Figure 1-3. Structure of BiP sub-domains


1.2.2.2  Allosteric regulation

To address the question of how ATP-cycling in the NBD and substrate binding to the SBD confer changes to the adjacent SBD and NBD respectively, studying full-length Hsp70s is key. Hsp70s structures comprising both the NBD and SBD have been reported. These include DnaK from *Thermus thermophilus* and *Geobacillus kaustophilus*, *Bos taurus* Hsc70, and *S. cerevisiae* Hsp110 [49-51]. These structures are surprisingly incompatible amongst themselves and their validity has been somewhat disputed. Firstly, Hsp110s are Hsp70-related proteins locked in the ATP-bound state. Thus they are not necessarily representative of Hsp70s, despite their structure being interpreted as such. Secondly, the location where the NBD and SBD are linked varies by more than 10 Ångströms (Å), and in the case of *G. kaustophilus* DnaK they are not even docked. Thirdly, the oligomers formed by these proteins in the crystal lattice are non-biologically relevant. Finally, all the proteins used contained truncations, mutations or deletions in their sequences.

More recently, four NBD-SBD constructs of *E. coli* DnaK have been studied by crystallography and NMR [52-55]. These studies are more in agreement as to their findings and they elegantly demonstrate how the conformation and function of Hsp70 NBD and SBD are tightly coupled. In the ATP-bound substrate-free state the NBD and SBD are docked (Figure 1-4A). Both adopt an open conformation within their respective sub-domain structures. Binding of substrates to the SBD promotes ATP hydrolysis in the NBD via subtle movements of the inter-domain linker region. This orients residues in the nucleotide-binding pocket optimally for catalysis. Binding of substrates and hydrolysis of ATP reverses the docking of the two domains (Figure 1-4B). When independent from the NBD, the substrate-bound SBD adopts a compact conformation in which SBD-α and SBD-β are compactly sandwiched together closing the lid and conferring the high affinity for peptides. These results illustrate how the sub-domains of Hsp70s communicate by an allosteric mechanism during nucleotide cycling and substrate binding.
Figure 1-4. Allosteric regulation of BiP chaperone

(A) X-ray crystal structure of ATP-bound (PDB: 4B9Q) *E. coli* DnaK Hsp70 [54]. Binding of ATP to the NBD brings sub-domains IB and IIB closer so as to trap the bound substrate whereas sub-domains IA and IIA separate. This creates a hydrophobic cleft to which the linker (green) binds to. This induces a conformational change in the SBD so that the SBD-α lid is docked onto the NBD. In this conformation, the substrate-binding cleft of SBD-β is exposed and BiP has low affinity for misfolded substrates.

(B) X-ray crystal structure of ADP plus substrate-bound (PDB: 2KHO) *E. coli* DnaK Hsp70 [53]. Misfolded substrates bind to SBD-β, which stimulates ATP hydrolysis. In this conformation, NBD-IA and NBD-IIA move closer together so that the linker can no longer be accommodated. The ‘free’ linker extends allowing the two domains to move independently of each other. This results in the docking of SBD-α lid onto SBD-β, which traps the bound substrate and confers high affinity. Conformational changes are highlighted with dashed red arrows.
The conformational changes that BiP undergoes upon binding nucleotides and substrates have been elegantly illustrated in a recent single-pair Förster resonance energy transfer (spFRET) study [26]. In the AMPPNP-bound state, the separation between the NBD and SBD-β is short indicating the two domains are tightly coupled, as expected. The distance between SBD-β and SBD-α is broad indicative of opening of the lid. On the other hand, in the apo and ADP-bound states, the distance between the NBD and SBD-β widens and that between the SBD-α and SBD-β becomes narrower. This represents a closed lid and high substrate affinity conformation. These results are in agreement with known conformational changes in Hsp70 cycling, discussed more in detail below.

1.2.2.3 Hsp70 ATP-bound substrate-free state

In the ATP-bound state, binding of the inter-domain linker to a hydrophobic cleft between Hsp70 NBD-I and NBD-II compacts the distance between the NBD and SBD and docks the two domains (Figure 1-4A) [52,54,55]. Indeed, mutagenesis of the linker has highlighted its importance in allosteric coupling of the NBD and SBD [56]. Docking of the NBD and SBD leads to major conformational changes in both domains. Most importantly, at least in terms of biological function, is the opening of the SBD-α lid [52,55,57]. This confers a low substrate affinity; bound substrates are released. Substrate release and domain docking also induces a conformational change within the NBD. Early biochemistry work on full-length DnaK showed that nucleotide cycling affects its susceptibility to proteolysis and alters the fluorescence signature of a conserved Tryptophan residue (Trp102) [58,59]. These experiments mirror changes in the surrounding physical and chemical environment of the NBD respectively. In agreement, the NMR spectra for the ATP-bound isolated NBD comports in large chemical shifts when compared to those of the ATP-bound NBD-SBD construct [52].

However, the flexibility of the ATP-bound state of full-length Hsp70s has inhibited its structural characterisation until recently. By engineering disulphide bonds Kityk and co-workers were first able to lock DnaK in this closed ATP-bound conformation, without affecting its activity [54]. As such they solved the crystal structure of the intermediate state. Shortly after, and in agreement with Kityk’s structure, Qi and co-workers crystallised full-length wild-type DnaK in the presence of ATP [55]. These structures unveiled that ATP-binding induces rotation of NBD-II and results in the closure of the space between Lobes IB
and IIB, so as to trap the bound ligand, and separation of Lobes IA and IIA. The latter creates a crevice for binding of the linker, as well as a surface for binding of SBD-β. Thus SBD-α and SBD-β become separated and attain the open lid conformation capable of accommodating new substrates.

1.2.2.4 Hsp70 substrate binding, ATP hydrolysis and the ADP-bound state

Upon substrate binding, the ATP-bound compact conformation is disrupted and the ATPase activity of Hsp70s is increased by 2 to 10-fold (Figure 1-4B) [17]. Biochemical data has found the linker domain to be both necessary and sufficient to stimulate the ATPase activity of the NBD. Firstly, mutant proteins in which a short conserved sequence of the linker domain (VLLL) is either deleted or mutated (to VDDL) results in full-length DnaK proteins with defective ATPase activity [52]. Secondly, truncated proteins consisting only of the NBD plus the linker region show very similar ATPase activity to full-length DnaK [52]. Thirdly, mutation studies have highlighted key residues that affect NBD-SBD communication [53]. These data, together with NMR spectra of the ATP, ADP and substrate-bound domain conformations have led to a proposed mechanism in which substrate binding loosens the inter-domain contact. In this conformation, some element of the linker and/or SBD remains in contact with the NBD to stimulate its catalytic activity. Once again, the transiency of this substrate-bound state in which ATP has not yet been hydrolysed makes it a challenging target for crystallisation studies. Indeed, in the ADP-bound state the NBD and SBD move independently of each other in a highly flexible manner as measured using NMR techniques [52,53]. Here, the SBD exists in a closed lid conformation where the SBD-α domain is docked to the SBD-β binding cleft and confers high affinity for the substrate [53]. The NBD of the full-length DnaK structure also adopts a closed state. This conformation is identical to that observed in the NBD-only structures obtained irrespectively of whether ATP, ADP or no nucleotide is bound [60-63].
1.2.3 Hsp70 co-factors

The chaperone cycle of Hsp70s is regulated by co-factors that facilitate either ATP hydrolysis or the exchange of ADP for ATP. These belong to the Hsp40 chaperone family and nucleotide exchange factors (NEF) respectively.

1.2.3.1 Hsp40 co-chaperones

Hsp40s are more abundant and diverse than Hsp70s. A single Hsp70 can bind multiple Hsp40s, albeit not simultaneously, to drive distinct cellular processes [64]. The specificity of Hsp40s is dictated by factors such as their sub-domain composition, their localisation or the specific residues involved in binding to the Hsp70 partner [64,65]. All Hsp40s contain a highly conserved 75 amino acid N-terminal region termed the J-domain. The J-domain binds at the NBD-SBD interface of Hsp70s and couples ATP hydrolysis and substrate binding [65,66]. A conserved HPD motif within the J-domain is involved in promoting ATP hydrolysis of Hsp70s, even in the absence of substrate [66]. Although the J-domain is the predominant site of interaction with Hsp70s, Hsp40s may contain additional domains: a disordered Glycine/Phenylalanine rich region, a Cysteine-rich zinc ion (Zn$^{2+}$)-binding region and C-terminal regions involved in binding of substrates and/or mediating dimerisation [64]. Because of their ability to associate with unfolded substrates, some Hsp40s are considered to be molecular chaperones themselves. Nonetheless it is not clear whether these assist in folding or simply function to deliver misfolded proteins to Hsp70s.

DnaJ is the canonical Hsp40 in E. coli. It was first identified as a co-factor for DnaK that stimulates its ATPase activity and helps replicate lambda phage deoxyribonucleic acid (DNA) in host cells [67]. Consequently, this family of ER-resident Hsp40s was termed ER localised DnaJ proteins (ERdj). Mammalian cells contain at least seven ERdjs, named in order of their discovery [12]. They differ in their functions and have varied sub-ER localisation and affinity for BiP. ERdj1, 2, 3 and 6 have been linked to folding of nascent proteins whereas ERdj4 and 5 assist in ERAD pathways. ERdj3 has been of interest in studies of BiP chaperoning activity.

Initial characterisation identified that (i) ERdj3 can bind to unfolded proteins which are also BiP substrates even in the absence of BiP, (ii) BiP and ERdj3 have similar tissue distribution
with highest levels expressed in secretory tissues such as the placenta, liver and kidneys as seen in expression profiles, and (iii) both proteins are transcriptionally upregulated during conditions of ER stress [68]. ERdj3 stimulates the ATPase activity of BiP by approximately 2-fold and accelerates binding of C-terminal [26,68]. spFRET experiments have shown that whereas ATP normally induces C-terminal dissociation from BiP, this is inhibited by the addition of ERdj3 [26]. Upon recruitment of BiP to the substrate, ATP hydrolysis must occur in order for ERdj3 to be released [69]. Because unlike BiP, ERdj3 does not assist in the folding and assembly of misfolded proteins, it is speculated to play a role in delivering unfolded substrates to BiP.

Early experiments that used Trp fluorescence to follow the conformational cycling of DnaK found that DnaJ accelerates the conversion of DnaK from an ATP to ADP-bound state [70]. More recently, spFRET measurements indeed have demonstrated that BiP assumes the substrate-bound closed conformation in the presence of ATP nucleotide and C-terminal substrate only when ERdj3 is also present [26]. Together, these studies support a mechanism by which ERdj3 binds unfolded proteins, recruits ATP-bound BiP and accelerates nucleotide hydrolysis. This would reorient BiP so that the misfolded proteins bind to the SBD and ERdj3 is released. Structural studies will be essential to further appreciate this system.

1.2.3.2 Nucleotide Exchange Factors

The release of substrates from Hsp70s requires exchange of ADP to ATP. In the mammalian ER, nucleotide exchange is catalysed by BiP-associated protein (BAP) and Grp170 [12]. GrpE is the NEF that acts on DnaK in E. coli. NEFs bind to the NBD of Hsp70s and actively promote the release of ADP thus stimulating the release of bound substrates. BAP and Grp170 both bind to BiP in the ER, however their mechanism of action and precise roles remain poorly characterised. Structural studies of cytosolic Hsp Binding Protein 1 (HspBP1) and B-cell lymphoma 2 (Bcl2)-associated athanogene (BAG) proteins, that are homologs of BAP and Grp170 respectively, in complex with their partner Hsp70s have allowed for the elucidation of their role in promoting nucleotide exchange [57,71,72]. HspBP1 and BAG bind to Lobe IB of Hsp70s’ NBD and induce a rotation of Lobe IIB. This locks the NBD in an open conformation that no longer confers high affinity for ADP. Although not definitive, these studies can be useful to suggest a mechanism by which BAP and Grp170 may function to control BiP cycling.
1.3 Activation of the Unfolded Protein Response

Conditions that perturb the physiological state of the ER, or ER stresses, can compromise the processing of mature proteins leading to the accumulation of misfolded proteins in the ER lumen and initiation of the Unfolded Protein Response (UPR) [73]. ER stresses include nutrient and energy deprivation, hypoxia, alterations in cell redox status, depletion of ER Ca\(^{2+}\) stores, elevated protein trafficking through the ER, altered post-translational modifications and pathogen infection. The UPR is a cellular response which attempts to restore ER homeostasis by up-regulating specific target genes and by reducing protein translation to limit the influx of newly synthesized proteins in the already stressed ER lumen. If the cells are unable to restore ER homeostasis and protein folding ability, apoptotic pathways are initiated.

The UPR was first characterised in yeast where the response to ER stress is mediated uniquely by the inositol requiring kinase 1 (IRE1p) protein [74]. In mammals, UPR signalling is mediated by three ER-localised signal transducers: IRE1, double-stranded ribonucleic acid (RNA) activated protein kinase (PKR)-like ER kinase (PERK) and activating transcription factor 6 (ATF6) [73]. There are two isoforms of IRE1: \(\alpha\) and \(\beta\). IRE1\(\alpha\) is expressed in most cells and tissues whereas IRE1\(\beta\) expression is limited to intestinal epithelial cells [75]. Throughout this report IRE1 will refer to the \(\alpha\) isoform unless otherwise specified.

IRE1, PERK and ATF6 all have an ER luminal domain that senses conditions of ER stress, an ER transmembrane domain and a cytosolic domain that transmits signals to transcriptional and translational machineries [73]. In the stressed ER, detection of the accumulation of unfolded proteins by the ER luminal sensor domains of IRE1, PERK and ATF6 leads to initiation of UPR signalling. The underlying mechanism of ER stress sensing and UPR activation is not yet clearly understood. The three UPR sensor domains do not share high overall sequence similarity. However since experimental induction of ER stress can activate all three signalling pathways it is possible that they have similar mechanism of sensing the accumulation of unfolded proteins [76]. Based on numerous biochemical and structural studies that have addressed this key question, different mechanisms have been proposed. These can be subdivided into two main models: BiP-dependent and BiP-independent [77]. The proposed
mechanisms by which ER stress initiates UPR signalling are discussed in detail in this chapter and are summarised in Figure 1-5. Downstream pathways that result from active UPR signalling are discussed in Chapter 1.4.

1.3.1 BiP-dependent mechanisms

1.3.1.1 BiP as a negative regulator of the UPR

The most supported mechanism of ER stress sensing is a competition model in which BiP is normally bound to the luminal domains of IRE1, PERK and ATF6 to keep them in an inactive state [76,78]. Upon ER stress, the high concentrations of unfolded proteins compete for BiP-binding. This sequesters it from the luminal domains of the UPR mediators allowing for their oligomerisation, in the case of IRE1 and PERK, and their subsequent activation. Experiments in the early 1990s first observed a correlation between BiP overexpression and attenuation of the UPR [79,80]. Later, BiP was shown to co-immunoprecipitate with the UPR transducers in the absence of ER stress in transfected yeast and mammalian cell lines [76,81,82]. Upon induction of ER stress, the reversible loss of BiP binding to IRE1p, IRE1, PERK and ATF6 was observed. BiP dissociation leads to oligomerisation and activation of IRE1 and PERK and initiation of UPR signalling [76]. In the absence of BiP, ATF6 is also free to translocate to the Golgi for activation [82]. Interestingly, over-expression of unfolded proteins that do not bind to BiP does not result in activation of UPR signalling [83,84]. This illustrates the specificity of BiP interaction with the UPR luminal domains.

In an attempt to identify the BiP-binding region, the luminal domains have been subdivided into five regions. Scanning deletions of IRE1p, IRE1 and PERK luminal domains were carried out in three separate studies [85-87]. Deletion of the ER-membrane proximal regions, leads to loss of BiP-binding in co-immunoprecipitation experiments. This is termed region V and corresponds to residues 448-517 of IRE1p, 389-434 of H. sapiens IRE1 and 411-481 of H. sapiens PERK. Indeed, deletion of region V of mammalian IRE1 and PERK results in constitutively active proteins even in the absence of ER stress [86,87]. The failure of these truncated proteins to be repressed, putatively due to their inability to bind BiP, points towards a negative regulatory role for BiP binding to region V.
Figure 1-5. Proposed mechanisms of ER stress sensing and UPR activation by PERK/IRE1

(A) BiP-dependent mechanism (BiP SBD is shown in teal; BiP NBD is shown in magenta; the linker is shown in green). During physiological conditions, BiP is bound to the luminal domains of PERK/IRE1 and as such keeps them in an inactive (possibly monomeric) state. It is widely believed that it is BiP SBD that binds to the luminal domains. Since misfolded proteins are also substrates for BiP SBD, during ER stress, their high levels compete for binding to BiP SBD. As such, BiP dissociates from the luminal domains of PERK/IRE1 allowing for their activation (autophosphorylation of their cytoplasmic domains and oligomerisation). BiP may mediate PERK/IRE1 activation by physically blocking the dimerisation motifs of the luminal domains so that they cannot oligomerise and become active. Alternatively, BiP binding/release may cause a conformational change in PERK/IRE1, which alters their cytoplasmic activation state.

(B) BiP-independent mechanism. Misfolded proteins may bind directly to the luminal domains and as such promote their dimerisation and activation of their cytosolic domains. Prolonged ER stress may lead to clustering of the misfolded protein-bound luminal domains that allows for co-operativity of the cytosolic domains and more robust downstream signalling.
In contrast, two studies have independently reported the association of BiP to IRE1 constructs that lack the BiP-binding region V [88,89]. In addition, deletion of region V of IRE1p was reported to not be sufficient to activate signalling, and sensing of ER stress conditions required a core region that is independent of the BiP-binding site [85]. From this latter study it was suggested that BiP is not the primary sensor of ER stress but rather this relies on the core sensing region of IRE1p luminal domain (regions II-IV) that then in turn positively functions to release BiP. What is clear from these studies is that BiP can repress IRE1, PERK and ATF6 activation and UPR signalling. However the exact model of BiP binding and dissociation from the UPR sensors and how this leads to their activation is not yet understood. A number of mechanisms have been proposed.

1.3.1.2 BiP repression by steric hindrance of dimerisation motifs

BiP-binding and oligomerisation regions of IRE1 luminal domain were found to overlap and a ‘steric hindrance’ inactivation mechanism was proposed [88]. Based on this study it was suggested that in the absence of ER stress, BiP blocks the IRE1 oligomerisation motif and thus maintains IRE1 in its inactive monomeric state. Upon release of BiP, the oligomerisation motifs are unmasked and become free to interact allowing dimerisation and activation of IRE1 (Figure 1-5A).

This mode of repression by BiP resembles a mechanism suggested to occur in ATF6. An ER stress responsive region in the ATF6 luminal domain required for the active release of BiP has been identified [82]. This is termed ‘luminal domain region 2’ and corresponds to residues 431-475 of H. sapiens ATF6. ATF6 ‘luminal domain region 2’ also contains to two Golgi-localisation sequences (residues 468-475 and 468-500). It has been proposed that association of ATF6 with BiP masks these sequences thereby retaining ATF6 in the ER and impeding its transport to the Golgi [82]. BiP over-expression correlates with reduced ATF6 translocation whereas deletion of the BiP-binding region leads to constitutive activation of ATF6 signalling [82].
1.3.1.3  The BiP-release model

The ATPase activity of BiP is central to its function in substrate binding and release and therefore is a plausible factor involved in UPR activation. Indeed, Kar2p mutants deficient in their ATPase activity are unable to activate the UPR in yeast cells [78]. Temperature-sensitive Kar2p NBD (type A) or SBD (type S) mutations that abolish their relevant functions have been identified [78]. At restrictive temperatures, although the type A mutants are unable to activate IRE1p signalling, they remain bound to the IRE1p substrate. On the other hand, type S mutations completely abolish Kar2p-IRE1p interaction and lead to constitutive UPR signalling in vivo even in the absence of ER stress signals [78]. Kar2p appears to associate to substrates via its SBD and serves as a repressor of the UPR, whereas the NBD is responsible for providing an activation signal to the bound substrate. However, in a later study by the same group, it was concluded that the constitutive activation of type S mutants in the first paper was due to elevated stress levels rather than to their inability to negatively regulate IRE1p [85]. This second study concludes on a tuning role for BiP rather than it being the main switch for UPR activation as discussed previously.

An independent study has suggested that BiP actually interacts with the UPR sensors via its NBD [90]. Mutations that lock BiP in the ATP-bound open state, but not the ADP-bound state, were seen to mitigate UPR signalling consistent with type A mutants described earlier. In an attempt to identify BiP regions to which IRE1p binds, oligosaccharide-shielding experiments were then carried out. Glycosylation of Lobe IB of BiP NBD specifically disrupted IRE1p binding. Detailed analyses of BiP NBD sequence conservation and structural properties followed by a mutation study identified Gln88 as the key residue involved in this interaction. Indeed, IRE1p did not co-immunoprecipitate with Q88E mutant BiP protein, despite this mutation having no effect on BiP’s chaperoning activity. Based on these results, together with known conformational changes, which occur during BiP substrate binding and ATPase cycle, the authors put forward a BiP-release mechanism. During ER stress, unfolded proteins would bind to BiP SBD and such induce a movement of the Lobe IB on the NBD and which would trigger the release of the bound UPR sensor luminal domain. However no evidence is provided for such mechanism.
Other ambiguities include the role of different nucleotides in the interaction. It has been reported that BiP-binding to IRE1p, IRE1 and PERK is only observed in the presence of ATP and not ADP [89,90]. However other studies observe dissociation of BiP in the UPR sensors upon the addition of ATP [78,91]. ATF6 binding appears to be independent of BiP nucleotide cycling [89]. A T37G BiP mutant that is unable to switch from an ADP to ATP-bound conformation was shown to be able of binding to ATF6 similarly to wild-type BiP [91]. Differences in BiP binding to ATF6 and IRE1 or PERK are likely to be a consequence of the absence of homology in their luminal domains and a reflection of the different mechanism of action.

1.3.1.4 Uncertainties regarding the BiP-dependent models

So far, the majority of the mechanistic understanding about UPR activation and the role of BiP in this process have come from immunoprecipitation experiments from whole-cell extracts and activity reporter assays in cells. It is possible that bridging proteins play a role in the interaction and that a multitude of other factors affect the results observed in these experiments. Additionally, from the studies above it cannot be excluded that IRE1, PERK and ATF6 binding to BiP is actually only a client interaction due to the sticky nature of BiP. Finally, the decrease in UPR signalling when BiP is over-expressed could be due to a decrease in levels of unfolded proteins rather than repression of IRE1, PERK and ATF6. In fact, published papers have reported that when expression of IRE1 and PERK is controlled, as opposed to over-expressed as is the case in most of the biochemical studies mentioned above, full activation of the mutants that lack the BiP-binding domain requires induction of ER stress [86,92]. Over-expression of the mutants possibly itself acts as an inducer of ER stress and therefore the observations seen do not represent physiological conditions of ER stress. Contradictorily, it has been shown that in the presence of ER stress, BiP actively and specifically dissociates from ATF6 and IRE1 but not from two other BiP substrates, Cm of immunoglobulins and a temperature-sensitive variant of vesicular stomatitis virus G (VSVG) protein [91]. This supports an explicit repressive role for BiP in UPR signalling.

Finally, as discussed above, type S Kar2p mutations abolish IRE1p-Kar2p binding and leads to constitutive activation of IRE1p [78]. Although this may be attributed to the absence of the UPR-repressive IRE1p-Kar2p interaction it could also be due to a lack of interaction of Kar2p
with unfolded proteins in the stressed ER and improper re-folding which would lead to severe ER stress and IRE1p activation.

1.3.2 BiP-independent mechanism

Kimata and co-workers first suggested that BiP is not the principal adjustor for sensing ER stress by the luminal domains of IRE1, PERK and ATF6. IRE1p deletion mutants lacking the BiP-binding site did not affect their ability to respond to ER stress [78]. From this study it was proposed that the central core portion of the luminal domain of IRE1p detects the presence of unfolded proteins in the stressed ER that leads to its activation. In this scenario, BiP binding and release serves merely as an adjustor mechanism to control the level of IRE1p signalling. These observations have led to the BiP-independent model which points towards a direct role for unfolded proteins in UPR activation, at least in the yeast system (Figure 1-5B).

1.3.2.1 Direct binding of unfolded peptides to IRE1p luminal domain

Since Kimata’s study, the direct binding of IRE1p to unfolded proteins has been observed and directly linked to its oligomerisation and activation [93,94]. Mutations of the putative peptide-binding pocket (M229A_F285A_Y301A) and dimerisation interface (T226F_F247A) decreased IRE1p-mediated (via HAC1) activation of the downstream UPR element (UPRE) promoter in reporter assays in vitro [95]. A peptide array screen derived by sequential tiling along the sequence of a constitutively misfolded carboxypeptidase Y (CPY) mutant (CPY*) was designed [96]. This was used to further characterise peptide binding to IRE1p. IRE1p appears to preferentially bind to peptides containing basic and hydrophobic residues rather than recognising a specific consensus sequence. These residues are normally found in the core of folded proteins but become exposed if the peptide chain is not properly folded. Although this resembles to the Kar2p/BiP mode of recognition of unfolded proteins, IRE1p was found to bind overlapping but not identical peptides to Kar2p. This observation probably rules out a peptide-competition model in which saturation of Kar2p is required for IRE1p activation.
1.3.2.2 IRE1p clustering

Sedimentation experiments have shown that binding of a CPY*-derived peptide induces formation of higher oligomeric species and therefore may be the driving force for cluster formation [96]. Clustering of IRE1p has been suggested to occur by a helical oligomerisation of IRE1 dimeric building blocks that interact via a secondary interface [97,98]. However it is difficult to foresee this model occurring in cells given that IRE1 is a transmembrane protein. Mutations affecting either peptide-binding, dimerisation or at the secondary clustering interface disrupt formation of higher oligomers IRE1p in vitro [93,96]. Since, several studies have also demonstrated that higher oligomer formation is essential for IRE1p activation following ER stress in vivo, pointing towards a biological role of this process, at least in the yeast system [93,99]. Clustering could allow for co-operativity between cytoplasmic domains of active IRE1p and as such facilitates full activation of downstream signalling.

1.3.3 Two-step activation mechanism

A two-step mechanism for UPR activation in which, additionally to the release of BiP, direct binding of unfolded polypeptides to the ER sensor luminal domains is also required for full activation of UPR signalling has been proposed [93]. Mutations that render IRE1p constitutively clustered require additional extrinsic ER stress signals for activation of the UPRE reporter. This was the first suggestion that although higher oligomer formation is necessary, it is not sufficient to activate IRE1p signalling. Furthermore, mutation of IRE1p regions I and V, either in combination or separately, results in differing ER membrane localisation patterns in vivo. Whereas single deletion of either region (ΔI and ΔV constructs) does not affect oligomerisation, ΔIΔV constructs show constitutive clustering on the ER membrane. From these observations it was proposed that a two-step regulatory mechanism for ER stress sensing and IRE1p activation exists. In this novel and more complex mechanism, during ER stress BiP indeed recognises and binds to unfolded proteins. As such, it is released from the ER sensor luminal domains allowing for their dimerisation and formation of the dimerisation that is capable of accommodating unfolded polypeptides [95]. Following binding of peptides to the luminal domains, conformational changes would trigger the formation of higher oligomers [96,100]. As such, this two-step activation mechanism could
function to ensure the tight control of UPR signalling. However, further evidence is required to sustain and characterise this complex model.

1.3.4 *S. cerevisiae* and *H. sapiens* IRE1 luminal domain structures

The X-ray crystal structures of *S. cerevisiae* and *H. sapiens* IRE1 luminal domains have been solved [95,101]. Despite the structural and functional conservation between the two, they support distinct models of ER stress sensing. *S. cerevisiae* and *H. sapiens* IRE1 luminal domain monomers are composed of a triangular assembly of β-sheet clusters, which are interspaced by α-helices (Figure 1-6). Dimerisation occurs by symmetric packing of two monomers through polar and hydrophobic interactions between two solvent-exposed antiparallel β-strands. This creates an interface groove in the luminal domain that resembles the peptide binding domains of major histocompatibility complexes (MHCs). The structure of IRE1p luminal domain supports the BiP-independent mechanism of UPR activation (Figure 1-6B) [95]. Two mechanisms of unfolded peptide binding to the MHC-like groove were proposed. Firstly, hydrophobic amino acid stretches present on the unfolded peptides may directly bind at the dimerisation interface. This is in agreement with the previously described peptide array study [96]. Alternatively, the MHC-like groove, being narrow, would be sterically inaccessible to compactly folded proteins but not to misfolded linear polypeptide chains.

However, based on the structure of *H. sapiens* IRE luminal domain, it was argued that the direct binding of unfolded proteins is neither probable nor necessary for initiation of the UPR (Figure 1-6C) [101]. The MHC-like groove appears to be too narrow and structurally unfavourable to accommodate peptide binding. Firstly, in *H. sapiens* IRE1, the conserved Gln105 residue forms a hydrogen bond with its symmetry related mate to contribute to dimer stability, and in doing so blocks access to the proposed peptide-binding groove. Secondly, mutagenesis of Met229, Phe285 and Tyr301 residues located in the groove of *S. cerevisiae* IRE1p were shown to affect its ability to activate the UPRE reporter in yeast cells [95]. The side chains of these residues point into the groove, and thus would contribute to peptide binding, were concluded to be key for IRE1p activation. However these residues are either buried (Tyr161 and Tyr179) or not conserved (Methionine-to-Lysine replacement; Lys121) in the *H. sapiens* IRE1 groove [101]. Finally, projection of the cytoplasmic domain of IRE1 would suggest that the MHC-like groove faces the ER membrane thus it is unlikely that
Based on these structures it would appear that *S. cerevisiae* and *H. sapiens* IRE1 have evolved diverse mechanisms of sensing ER stress.

Nonetheless, ER stress has also been shown to promote clustering of mammalian IRE1α and IRE1β [98,102]. Disruption of IRE1α dimerisation interface inhibits both cluster formation and splicing of *Basic Leucine Zipper (bZIP)-containing X-box-binding protein 1 (XBP1)* mRNA. Additionally, IRE1β, but not hIRE1α, has been shown to co-immunoprecipitate with unfolded proteins in stressed cells [102]. It is possible that the two proteins used to solve IRE1 luminal domain structures were trapped in different conformations and represent different states of dimeric IRE1 luminal domains. In addition, both *S. cerevisiae* and *H. sapiens* IRE1 luminal domain crystal structures lack electron density for the ER-membrane proximal region that is speculated to be key for BiP-binding. Therefore it is challenging to postulate on the properties of BiP-binding and what effect this would have on the conformation and oligomeric state of IRE1.

PERK and IRE1 luminal domains are functionally interchangeable: PERK luminal domain can substitute for IRE1’s to signal the UPR *in vitro* and vice versa [76,103]. This supports the idea that the luminal domains function solely as sensors for ER stress sensing and that dimerisation is the key signal for the activation of the cytoplasmic effector domains, at least for IRE1 and PERK. Additionally, secondary structure prediction suggests that IRE1 and PERK have similar folds and share topological arrangements in their luminal domains. Therefore, given the apparent conservation in structure and function between IRE1 and PERK luminal domains, it is likely that they have a similar mechanism for sensing perturbations in ER protein-folding reactions. Obtaining the crystal structure of PERK luminal domain is key to elucidate the ER stress sensing mechanism in higher eukaryotes.
Figure 1-6. X-ray crystal structure of IRE1 luminal domain

The N-terminal ER luminal domains of S. cerevisiae (residues 114-449) (PDB: 2BE1) and H. sapiens (residues 24-390) (PDB: 2HZ6) IRE1 orthologues have a similar structure [95,101].

(A) IRE1 luminal domain monomers are composed of a triangular assembly of β-sheets interspaced by α-helices. Two monomers come together through hydrogen bonding between two anti-parallel β-strands to form biological dimers. Dimerisation creates an MHC-like groove characterised by a bed of β-sheets with two transverse α-helices (cyan).

(B) Based on the S. cerevisiae structure this was postulated to constitute the region for misfolded protein binding which backs the BiP-independent model of ER stress sensing.

(C) The H. sapiens structure does not fit with these observations since (i) the width of between the groove is too narrow (ii) a conserved Glutamate forms hydrogen bond with its symmetry related partner which blocks access to the peptide-binding site, (iii) residues essential for peptide binding in S. cerevisiae IRE1p are either not conserved (Methionine-to-Lysine mutation) or buried (aromatic Tyrosine).
1.3.5 Alternative mechanisms of UPR activation

At least three distinct systems involved in detection of ER stress and activation of the UPR have been proposed: (i) ADP-ribosylation, (ii) N-linked glycosylation of ATF6 and (iii) aberrations in the ER-membrane. These will be briefly discussed below.

1.3.5.1 PARP16-mediated ribosylation activates IRE1 and PERK

Poly ADP ribose polymerase (PARP) proteins are known to regulate key stress response pathways including DNA damage repair and the cytoplasmic stress response and have recently also been linked to activation of IRE1 and PERK [104]. PARP16-mediated ADP-ribose modification is sufficient for activation of IRE1 and PERK signalling in vivo, even in the absence of ER stress. In PARP16+/− cells, IRE1 and PERK remain bound to BiP following induction of ER stress. Interestingly, ATF6 activation is not affected in any way by PARP16. Together, these observations have led to the proposal of a mechanism in which PARP16-induced ribosylation could induce conformational changes in IRE1 and PERK and as such would actively assist in the release of BiP. For example, ribosylation could open the binding groove of IRE1 and PERK similarly to the arrangement observed in the IRE1p crystal structure, so that it can accommodate unfolded peptides [95]. Because ATF6 exists as a monomer and does not require dimerisation or formation of a binding pocket for activation, it reasons that PARP16 does not play a role in its activation. Although this is an attractive hypothesis and it is consistent with previous work, further experimental work is needed to support it.

1.3.5.2 Under-glycosylation acts as an activating signal for ATF6

Glycosylation of ATF6 luminal domain has been suggested to serve as a sensor for ER stress and mediate ATF6 activation [105]. An ATF6 mutant with aberrant glycosylation, T645I, has impaired interaction with the CRT chaperone, faster Golgi translocation and enhanced transcription of target genes. Since ER stress perturbs glycosylation pathways and improper glycosylation is an inducer of the UPR, it follows that it could represent a mechanism for activation or regulation of the sensor domains of IRE1, PERK and ATF6. To date though, glycosylation of IRE1 and PERK luminal domains have not been linked to ER stress sensing and the role of their N-glycosylation are not known. In addition, it is possible, that anomalous
glycosylation of the T654I mutant actually reduced ATF6 function rather than decreased its susceptibility to ER stress.

1.3.5.3 Distortion of the ER membrane can activate the UPR

Extrinsic factors, have recently been suggested to initiate at least the IRE1 branch of the UPR. The cytoplasmic domain of IRE1 is composed of a kinase domain and an endoribonuclease (RNase) domain. When activated, the latter is responsible for splicing of the xbp1 mRNA (Homologous to Atf/Creb1 (hac1) mRNA in S. cerevisiae), which leads to downstream signalling cascades [74].

In addition, aberrations of the ER membrane can activate IRE1p signalling and thus may also play a role in UPR activation [106]. Interestingly, deletion of IRE1p luminal domain region III (ΔIII) compromises its ability to bind unfolded peptides and activate hac1 mRNA in response to ER stress but has no effect on the formation of clusters. ΔIII IRE1p is actually activated similarly to wild-type IRE1p following modifications of the lipid content of the ER membrane suggesting there is more than one signal capable of activating UPR during ER stress. Substitution of IRE1p luminal domain regions I to V with a dimer-forming bZIP motif has been used to probe the role of the luminal domain in UPR signalling. Since bZIP-IRE1p chimeras are capable of splicing hac1 mRNA in response to tunicamycin treatment it has been suggested that there exists a mode of sensing ER stress which is independent of luminal domain function [103]. Indeed, bZIP-IRE1p and wild-type IRE1p are activated alike following inositol depletion of the ER membrane [106]. Together, these results indicate that distortions of the ER membrane can act as a UPR activating signal and that the transmembrane/cytosolic domains also play a role in sensing conditions of ER stress.
1.4 UPR signalling

An overview of the UPR is illustrated in Figure 1-7.

1.4.1 UPR target gene transcription by the IRE1 and ATF6 pathways

IRE1 is a type I transmembrane protein kinase whose cytoplasmic domain comprises of a Serine/Threonine kinase domain and a site-specific RNase domain. Luminal activation of IRE1 leads to its dimerisation/oligomerisation [76]. As such, its cytoplasmic domains are brought into close proximity, facilitating trans-autophosphorylation and activation of the RNase domain. The importance of IRE1 dimerisation in its activation has been highlighted in several studies. For instance, when IRE1 luminal region is substituted by a bZIP dimerisation domain, the chimera is constitutively active [103]. Additionally, mutation of conserved residues involved in dimerisation (Q105E, D123P, W125A) of IRE1 is sufficient for disrupting phosphorylation of its cytoplasmic domain and downstream signalling activation [101].

Active IRE1 RNase domain excises a 26 nucleotide intron from the xbp1 mRNA (hac1 mRNA in S. cerevisiae). This leads to a translational frameshift and conversion from a 267 amino acid unspliced XBP1 (XBPlu) form to a 371 amino acid spliced XBP1 (XBP1s) version [107]. XBP1s comprises of the original N-terminal DNA-binding domain plus an additional potent transactivation domain in its C-terminus [107,108]. XBP1s binds to gene promoters containing the ER stress response element (ERSE) (CCAAT-N9-CCACAG consensus sequence) and activates their transcription. ERSEs are found on genes involved in processes that assist in restoring ER homeostasis. These include protein entry into the ER (components of the Sec61 channel), protein folding (CNX, CRT, PDIs, glycosylation enzymes) and ERAD components (EDEM, ERAD-associated E3 ubiquitin-protein ligase (HRD1)). Indeed, BiP and ERdj3 genes are also highly upregulated by XBP1s (and the active ATF6 p50 fragment). XBP1s has also been shown to enhance phospholipid biosynthetic pathways and thus indirectly regulates biogenesis of the ER and Golgi.

ATF6 is a type II transmembrane protein whose cytosolic domain contains a bZIP transcription factor. Upon ER stress, ATF6 translocates to the Golgi apparatus where the site-1 protease (S1P) and site-2 protease (S2P) cleave a 50 kDa cytosolic region to yield the ATF6 p50 active
fragment [109]. ATF6 p50 is a bZIP transcription factor that translocates to the nucleus and functions mainly in parallel with IRE1 to activate transcription of ERSE-containing UPR target genes [110]. Recently, a number of ATF6-like proteins have been identified as possible tissue-specific ER stress sensors [111]. These are cyclic adenosine monophosphate (AMP) responsive element-binding protein hepatocyte specific (CREB-H), Luman, old astrocyte specifically induced substance (OASIS), transcript induced in spermiogenesis-40 (Tisp40) and box B-binding factor 2 (BBF2) human homolog on chromosome 7 (BBF2H7). All these ATF6-related bZIP factors are processed at the Golgi similarly to ATF6 but their role, if any, in the UPR is not yet clear.

The pattern of genes induced by XBP1s and ATF6 p50 is a highly complex and dynamic response, which is dictated by a multitude of factors including the nature of the ER stress and the duration of UPR activation. Tissue-specific variations have also been identified. For example, overexpression of ATF6, in contrast to IRE1, does not induce the transcription of ERAD-inducing edem, herp and hrd1 mRNAs [110,112]. Immunoprecipitation experiments have shown that ATF6 p50 can heterodimerize with XBP1s to bind to promoter elements of edem1, herp and hrd1 genes with 8-fold higher affinity than XBP1s homodimers [110]. Clearly, both ATF6 and IRE1-XBP1 pathways must be simultaneously activated for the full execution of UPR-dependent ERAD. This illustrates the level of control of UPR signalling that exists to fine tune outputs such as ERAD.
Figure 1-7. Overview of the mammalian UPR

The three UPR mediators, PERK, IRE1 and ATF6 are ER-transmembrane proteins with an ER luminal domain that senses conditions of ER stress, a transmembrane region and cytosolic effector domains. During ER stress, the UPR is activated. The cytosolic domains of PERK and IRE1 auto-phosphorylate and form higher oligomeric species. Active PERK phosphorylates eIF2α which leads to (i) a global reduction in protein translation to alleviate the folding burden on the ER and (ii) specific translation of ATF4 mRNA. Active IRE1 mediates splicing of xbp1 mRNA that leads to expression of XBP1s protein. Binding of TRAF2 to active IRE1 cytosolic domain activates the JNK-pathway. On the other hand, active ATF6 translocates to the Golgi where it is proteolytically cleaved into an active p50 fragment. ATF4, XBP1s and ATF6 p50 work in parallel to induce transcription of UPR target genes and, under prolonged ER stress, to induce apoptotic pathways. JNK also plays an important role in apoptosis.
1.4.2 Translation attenuation

PERK is a type I transmembrane protein with a cytoplasmic domain eukaryotic Initiation Factor 2α (eIF2α) kinase domain. Similarly to IRE1, PERK dimerisation is essential for cytoplasmic domain autophosphorylation and activation of downstream pathways. Fusion of PERK’s cytoplasmic domain to a cluster of differentiation 4 (CD4) T-lymphocyte receptor results in full signal response following chemical crosslinking of the receptors [76]. Three other mammalian eIF2α kinases have been identified and each sense distinct stress conditions [113]. General control non-depressible-2 (GCN2) is induced during amino acid starvation, ultraviolet (UV) irradiation and proteasome inhibition. Haem-regulated inhibitor (HRI) is regulated by heme-deficiency and oxidative stress in erythroid tissues. PKR participates in the interferon-mediated antiviral defence. When activated, these kinases phosphorylate Ser51 on eIF2α, which inhibits the formation of the pre-initiation complex (PIC), a fundamental step in translation initiation [113]. The common downstream effect of activation of these eIF2α kinase pathways is the attenuation of translation. The mechanism involved is summarised in Figure 1-8. In UPR signalling, the reversible and transient reduction of mRNA translation limits the influx of newly synthesized proteins in the already stressed ER lumen. Other physiological effects of PERK-mediated translation attenuation include inhibition of cyclin D1 translation leading to Gap 1 (G1) phase cell cycle arrest, and targeted ubiquitination and degradation of inhibitors of nuclear factor-like 2 (Nrf2) which leads to activation of genes involved in redox homeostasis [114,115].
Figure 1-8. Translation attenuation by activated PERK

eIF2-GTP (green) binding to Met-tRNAi leads to the recruitment of additional elFs (grey) and of the 40S ribosomal subunit (yellow) to form the PIC. PIC scans the mRNA to be translated until it recognises the AUG start codon. At this point the elFs are released via hydrolysis of the GTP-bound eIF2. This allows for the recruitment of the 60S ribosomal subunit (blue) and formation of the 80S elongation complex competent of translation. The guanine nucleotide exchange factor eIF2B (grey) is responsible for exchanging GDP for GTP in eIF2 for the cycle to start over. During ER stress, active PERK phosphorylates eIF2α that inhibits recycling of eIF2 from a GDP-inactive form to a GTP-active form. This reduces the levels of PIC formed and hence decreases global translation levels.
Two additional pathways for ER-selective reduction of substrate burden have been described: regulated IRE1-dependent decay (RIDD) and pre-emptive quality control (pQC). RIDD is a relatively novel and poorly characterised branch of IRE1 signalling in which mRNAs encoding proteins that transverse the secretory pathway are degraded [116-121]. Recent studies have suggested that at physiologically low levels of ER stress, IRE1’s RNase activity is very specific whereas RIDD may be activated during more intensive stress signalling [122,123]. One possibility is that RIDD may add a dimension to the tuning of the ER stress response as a consequence of different types or strengths of ER stress [122]. Alternatively it may be a driving force towards apoptotic pathways. The pQC pathway controls the efficiency of nascent chain translocation into the ER lumen by preferentially allowing entry of polypeptides containing a specific signal peptide. pQC is a thus a means of indirectly adapting a cell’s protein production to its needs and is especially important in stress responses [30]. During ER stress the pQC has been shown to selectively degrade nascent secretory or membrane proteins, but it directs fast and preferential translocation of BiP chains [30].

1.4.3 ATP4 translation

eIF2α phosphorylation does not simply result in a static change in the composition of proteins but rather leads to a coordinated change the translation of key regulatory proteins. Mainly, translation of mRNA encoding the bZIP activating transcription factor 4 (ATF4) is activated as a consequence of PERK-eIF2α signalling during ER stress in a mechanism involving ribosomal re-initiation (Figure 1-9). ATF4 synthesis is controlled by two upstream open reading frames (uORFs) on the 5’ untranslated region of its mRNA. uORF2 overlaps out of frame with atf4’s AUG codon. Under physiological conditions expression of uORF2 leads to the bypassing of that encoding ATF4. However, during stress, eIF2α phosphorylation reduces eIF2-guanosine triphosphate (GTP) levels which increases the time required for the scanning ribosomes to become competent to re-initiate translation. This delay means some of the scanning ribosomes bypass uORF2 AUG codon and instead re-initiate translation at atf4’s AUG codon [124,125]. Preferential translation of ATF4 is shared by the other eIF2α kinases in response to their relevant stress signals. These eIF2α-ATF4 pathways have been collectively referred to as the integrated stress response [126].
ATF4, similarly to XBP1s and ATF6 p50, directs transcription of UPR target genes and specifically targets genes important for amino acid biosynthesis and transport functions, anti-oxidative stress responses and apoptosis. These include Asparagine synthetase (ASNS), growth arrest and DNA damage 34 (GADD34), CAAT/enhancer-binding protein homologous protein (CHOP), activating transcription factor 3 (ATF3) and ATF6 [127,128]. The direct dependence of ATF6 activation on the PERK-eIF2α-ATF4 pathway during ER stress highlights the intricacy and cross-talk that exists between the various UPR branches [128].
Figure 1-9. Selective translation of atf4 mRNA

eIF2α-P increases translation of atf4 mRNA by a mechanism involving ribosomal re-initiation. atf4 mRNA contains two uORFs and an ATF4 ORF coding region. After translation of the 5’ uORF1, ribosomes resume scanning along the mRNA.

(A) During physiological conditions, eIF2-GTP (green) is readily available so that the 80S elongation complex is rapidly formed, translation quickly resumes and uORF2 is expressed. Since uORF2 overlaps out of frame with ATF4 ORF, ATF4 coding region is bypassed and the protein is not expressed.

(B) During ER stress, phosphorylation of eIF2α reduces the levels of eIF2-GTP, which increases the time required for the scanning 40S ribosome to acquire eIF2-GTP to form the active PIC. This results in the majority of scanning ribosomes to bypass uORF2. However in the interval between uORF2 and ATF4 ORF, eIF2-GTP may have sufficient time to bind to the 40S subunit, form the PIC, recruit the 60S subunit and allow for expression of ATF4.

ORFs are illustrated as boxes and shading indicates their encoding sequence is expressed. Active 40S-eIF2-GTP (PIC) is shown in yellow; inactive 40S alone is shown in white.
1.4.4 Apoptosis

Usually, all three branches of the UPR are activated by any given ER stress. However their timing can differ, which affects the physiological effect of UPR signalling (pro-survival versus apoptosis) [129]. Prolonged, chronic or high levels of ER stress sensitize cells to apoptosis. The regulation of this switch is mainly coordinated by IRE1 and PERK signalling.

Although PERK signalling appears to be constant throughout the UPR, translational repression is transient. 6 to 12 hours after initiation of the UPR, ATF4-dependent activation of the GADD34 promoter leads to dephosphorylation of eIF2α allowing for translational recovery [130]. GADD34 also results in increased levels of BiP expression, which inhibits PERK activation. In addition, GADD34 plays a role in driving cells towards apoptotic pathways [131]. GADD34 leads to translational recovery, by dephosphorylating eIF2α. If the cells are unable to cope with the increased protein synthesis, apoptotic pathways are initiated. In addition, P58 inhibitor of protein kinase (P58IPK) is activated during ER stress and has been shown to decrease PERK phosphorylation in the late ER stress response [132].

In contrast to PERK signalling, levels of active ire1 and xbp1s mRNA fall after prolonged ER stress [98]. This may serve as a mechanism to drive cells towards apoptosis by shutting down resolving pathways (such as IRE1’s). Indeed, ire1 -/- or xbp1 -/- cells show increased apoptosis as a result of ER stress [133]. On the other hand, experimentally prolonging IRE1 signalling pathways enhances the survival of cells [133]. During the early phase of the UPR, eIF2α phosphorylation attenuates translation of many mRNAs, including that encoding for XBP1s [134]. When translation resolves during the later phases of the ER stress response, the accumulated xbp1s mRNA is translated allowing transcription of target genes including its own creating a positive feedback loop [134].

It is clear that diversion from survival to apoptotic pathway is not a binary event but rather a carefully coordinated outcome of the three UPR pathways. Mechanisms by which apoptotic signals are generated following prolonged ER stress include CHOP, IRE1-mediated TNF receptor associated factor 2 (TRAF2) activation and caspase pathways and are summarised in Figure 1-10 [142].
Figure 1-10. UPR-mediated apoptosis

Prolonged conditions of ER stress can result in apoptosis. This can be mediated by PERK (ATF4-CHOP), IRE1 (XBP1s-CHOP; TRAF2-JNK/Bak/Bax; Jab1; RIDD; CD59) and ATF6 (CHOP) pathways. CHOP leads to inhibition of expression of anti-apoptotic Bcl2 proteins and increased expression of genes encoding pro-apoptotic factors. JNK leads to activation of Bim and Bad which also inhibit anti-apoptotic Bcl2 proteins. Release of \( \text{Ca}^{2+} \) from the ER (via Bak/Bax and generation of ROS) and mitochondria (due OMM permeabilization) leads to activation of the caspase pathways, which ultimately leads to apoptosis. Two \( \text{Ca}^{2+} \)-independent caspase pathways have been identified: release of mitochondrial Cytc and caspase 12-mediated in \( M. \text{musculus} \) cells. Transcription factors (purple) are shown as boxes; inactive (white) and active (blue) mediator proteins are shown as ovals.
1.5 UPR pathologies

In lower eukaryotes, ER stress leads to perturbations in cell wall assembly and function of the plasma and vacuolar membranes [73]. In higher eukaryotes, perturbations of ER function and UPR activation has been linked to several human diseases [135-137]. These include diabetes, cystic fibrosis, ischaemia, neurodegenerative conditions and cancer. The importance of mammalian UPR pathways in cellular functions, especially in that of secretory organs, has been highlighted in knockout animal models.

*perk* knockout mice develop type 1 diabetes and exocrine dysfunction after birth [138]. These mice died within 18 hours after birth due to hyperglycaemia. A similar phenotype was observed for mice containing a homozygous *elf2α S51A* mutation [139]. 39 PERK mutations have been reported in patients with Wolcott-Rallison Syndrome (WRS), a rare autosomal recessive disorder in humans [140]. Similarly to *perk*<sup>−/−</sup> and *elf2α S51A* mice, WRS causes defects in pancreatic function, metabolism and skeletal development.

*iire1* and *xbp1* knockout is embryonic lethal [141-143]. *iire1β* knockout allows for normal development provided the mice are not subjected to insults in the gut [144]. *atf6α* and *atfβ* single knockout mice are viable and develop normally whereas their combined deletion results in embryonic lethality [110]. This indicates that the two ATF6 isoforms functional in a combinatorial fashion.
1.5.1 PERK-mediated disease states

The number of studies implicating components of the UPR with human diseases is vast (recently reviewed by Yoshida and co-workers) [137]. Only the major discoveries regarding PERK will be outlined here.

1.5.1.1 Diabetes

PERK is essential for development of endocrine progenitor cells into insulin-expressing β-cells and glucose homeostasis during the early neonatal period and can cause type 1 diabetes [145]. Several factors play a role in the development of insulin resistance and β-cell death as a result of ER stress which ultimately causes type 2 diabetes [146]. Diabetes may develop due to PERK-mediated disruption of protein translation, including that of proinsulin, or from β-cell death as a result of apoptotic UPR pathways.

1.5.1.2 Ischaemia

Deletion of perk, but not the other elf2α kinases, blocked cerebral ischaemia-induced elf2α phosphorylation [147]. Activation of PERK can lead to stalling of PICs which are essential for the reconfiguration of stress granules during ischaemia [147]. When stress granules are sequestered into larger protein aggregates, they signal irreversible inhibition of protein synthesis and death of affected neuronal tissues [147].

1.5.1.3 Cancer

Cancer cells must be able to proliferate in harsh nutrient deprived and hypoxic environments. These conditions are often causative of ER stress. Activation of the UPR is a mechanism for tumour cells to be able to cope in such environments. Microarray analyses have identified several tumorigenic genes, which are preferentially translated in a PERK-dependent manner [148]. In fact, injection of perk−/− mouse embryonic fibroblasts (MEFs) transformed with the oncogenic kirsten rat sarcomaA-2 viral (kras2) gene into nude mice leads to substantially slower growing tumours and decreased microvessel formation [148]. Loss of perk in breast cancer mice models results in delayed metastasis and tumorgenesis [149].
### 1.5.2 Targeting the UPR for therapies

Given that the functionality of the UPR is critical for many disease states, its components represent excellent targets for therapy of these pathologies. As thoroughly reviewed by Hetz and co-workers these include compounds to attenuate ER stress levels (enhancers of eIF2α phosphorylation, use of chemical chaperones, chemical induction of BiP) or to inhibit UPR pro-survival effects/promote adaptation to ER stress (inhibitors of IRE1 and PERK cytosolic domains, modulators of chaperones and ERAD components) [150]. Gene therapy using recombinant viruses to deliver active UPR components to specific affected tissues has been successfully employed. For example use of adeno-associated virus (AAV) to incorporate BiP gene has been shown to restore visual function in mutant rhodopsin transgenic rats and reduce liver steatosis in obese mice [151]. Inhibition of PERK by using the small inhibitor molecule ISRIB has been shown to reduce the viability of cells subjected to chronic ER stress [152]. ISRIB blocks the activation of PERK, and therefore eIF2α phosphorylation and translation attenuation, so that cells are unable to cope with their protein synthesis. As such it is an important tool that can be employed to kill cancerous cells, especially those derived from secretory lineages and thus have increased protein synthesis and higher basal levels of ER stress. This includes myelomas, and pancreatic and breast cancers. Finally, Guanabenz has been used to increase the viability of cells subjected to ER stress [153]. Guanabenz selectively inhibits GADD34 which delays eIF2α dephosphorylation and translational recovery.

Structural insights of the UPR components are certainly fundamental for the efficacy and specificity of drug development. Indeed, since the resolution of the structure of IRE1’s cytoplasmic domain, several specific inhibitors of IRE1 RNase activity have been developed for anti-cancer therapies, especially in the context of multiple myeloma [136,154,155].
2 X-ray crystallography theory
2.1 Protein crystallisation

Obtaining crystals of a protein is a key step in determining its structure. For crystals to form, the protein molecules must separate from the surrounding solution and self assemble in a three dimensional periodic pattern [156]. This forms the crystal lattice, which is held together by weak intermolecular interactions. The simplest building block from which the lattice can be constituted, by successive translation along its edges, is the unit cell [157]. The asymmetric unit of the unit cell is its part that can be used to generate the complete unit cell by applying the symmetry operations of the crystal’s space group (Chapter 6).

Protein crystallisation can be elegantly illustrated by the phase diagram (Figure 2-1) [158]. The solubility curve of a protein represents the thermodynamic equilibrium between the liquid and solid phases. In the supersaturated metastable zone, the solid phase is more thermodynamically stable and hence the formation of protein crystals is favoured. In solution, protein molecules constantly tumble, randomly collide in different orientations and may form inter-molecular contacts. However, the kinetic barrier that exists for phase separation must be overcome to allow the self-assembly of protein molecules into crystals. This is achieved by a reduction in the net Gibbs free energy of the system resulting from the formation of favourable intermolecular contacts, accompanied by a gain in entropy.

As the proteins shift from the liquid to solid phase, there is a reduction in entropy due to the loss of conformational freedom. This unfavourable effect however, is more than compensated by the gain in entropy that occurs as a result of the release of the water molecules around the hydrophobic and polar solvent-exposed residues of the protein in solution. Overall this results in the reduction in the net Gibbs free energy of the system. This overcomes the kinetic barrier that exists for phase separation and allows the self-assembly of protein molecules into crystals. This process is known as nucleation and is driven by supersaturation in the labile zone. Nucleation is followed by crystal growth to expand the lattice (metastable zone). This involves the addition of protein molecules following the regular and repeating arrangement defined at the nucleation stage. The gain in enthalpy provided by the binding reactions overcomes the entropic loss during crystal growth so that the system can proceed. When supersaturation is exhausted, due to the concentration of free protein molecules in the solution being depleted, crystallisation is complete. The size
and rate of crystal formation can be in part controlled by varying factors such as the composition of the crystallisation buffer, protein concentration, pH or temperature. These affect a protein’s solubility and thus its behaviour in the phase diagram.

Controlled nucleation is frequently difficult to achieve. Initial hits often consist of either too many nucleation events, resulting in showers of microcrystals, or too few, yielding very poor crystal quantities. Seeding is a technique used to manipulate nucleation [159]. Crystal seeds are made from crushing pre-existing crystals, or sometimes crystalline or non-amorphous precipitates. These are introduced as heterogeneous nucleation sites in supersaturated protein solutions. When optimised this can lead to the growth of single large crystals of high quality.

Common methods employed to crystallise proteins include vapour diffusion, sandwich drop, microbatch, under oil, dialysis and free interface diffusion [156]. Vapour diffusion was used throughout this study. In this method, the protein is mixed with the crystallisation buffer and left to equilibrate against a reservoir containing the same crystallisation buffer, in a sealed environment. Owing to differences in the concentration of the precipitating agent between the crystallisation drop and the reservoir, water vapour diffuses from the drop to the reservoir. This raises both the protein and precipitant concentrations of the drop, which can result in supersaturation and thus spontaneous nucleation, followed by crystal growth.
Figure 2.1. Crystallisation phase diagram
The solubility of proteins can be represented by the phase diagram. This information can be used to understand protein crystallisation as a function of protein and precipitant concentration. In conditions below the solubility curve, in the undersaturated zone, proteins remain in solution. Spontaneous nucleation can occur only in the supersaturated labile zone. This causes the protein concentration in solution to decrease. Supersaturated metastable conditions then sustain growth of the microcrystals. If the concentration of protein or precipitant is too high, precipitation occurs.
2.2 X-ray diffraction

2.2.1 X-ray scattering by protein crystals

Protein X-ray crystallography is based on the diffraction of X-rays by the protein’s atoms, primarily by their electrons. The structure, or arrangement, of the scattering atoms within the crystal lattice is derived from their X-ray diffraction pattern [157].

In order for an object to diffract light, the wavelength must be in the order of the size of the object. The bonded atoms in protein molecules are roughly 1.5 Å apart. X-rays have a wavelength between 0.1 and 100 Å and therefore fall within the required range to record diffraction by atoms in a protein. Synchrotron radiation sources are used to accelerate electrons, to near the speed of light, through a magnetic field around the booster and storage rings [160]. Bending magnets are used to deflect the electrons passing through their magnetic field to give off electromagnetic radiation. This produces a beam of synchrotron light whose intensity is increased by insertion devices, commonly wigglers and undulators. Insertion devices are found in the straight sections of synchrotrons so that light in the X-ray range reaches the end station of beamlines. The wavelength of the light produced can be modified by changing the gap between the components of the insertion devices.

2.2.1.1 Bragg’s law and the Ewald sphere

In 1913 by William Lawrence Bragg and his father William Henry Bragg showed that a diffracted X-ray beam occurs when a specific set of conditions, defined by Equation 1, are satisfied [157,161].

Equation 1

\[ 2d \sin \theta = n\lambda \]

Here \( d \) is the spacing between two lattice planes in the atomic lattice, \( \lambda \) is the wavelength of the incident wave and \( \theta \) is the angle between the incident ray and the scattering plane. This is known as Bragg’s law and is the foundation of X-ray crystallography (Figure 2-2A). The intensity of the diffracted X-ray is dependent on how many atoms, or electrons, lie on the
planes upon which it impinges. The ‘space’ occupied by the reflections of the scattered X-ray beam is known as the reciprocal lattice and is defined by the coordinates \(hk\ell\) (Miller indices). The coordinates of the real space are \(xyz\) and define the locations of the protein’s atoms in the unit cell. The vertex of the unit cell is known as the origin, which has \(xyz\) coordinates 000.

Another geometric construct which can be used to demonstrate the relationship between the planes of a crystal lattice and its reciprocal is the Ewald sphere (Figure 2-2B). The Ewald sphere, or sphere of reflection, introduced by Paul Peter Ewald in 1913, has a radius of \(1/\text{wavelength}(\lambda)\) and passes through the origin of the reciprocal lattice [157,161]. A reciprocal lattice point is recorded only when it lies on the Ewald sphere. Therefore at any one orientation only a small proportion of the reciprocal lattice points will contribute to the diffraction pattern. By rotating the crystal in the X-ray beam, different points of the reciprocal lattice lie along the edge of the Ewald sphere and are recorded. The degrees of rotation of the crystal during each image, phi (\(\phi\)) angle, and throughout the data collection, chi (\(\chi\)) angle, are determined for each crystal.
Figure 2-2. Bragg’s law and Ewald sphere

(A) The difference in path length travelled by two rays reflected from successive planes (Pn) of a crystal lattice (purple), R1 and R2, is equal to 2d_hkl sinθ. Bragg’s law states that if this distance, 2d sinθ, equals to an integral multiple of the wavelength (nλ), then the two rays interfere constructively to produce a strong diffracted beam.

(B) X-rays are diffracted by a protein crystal (purple) to produce a reciprocal lattice (grey), with planes Lm. The diffracted X-ray diverges from the beam by the angle 2θ. The sphere of reflection, or Ewald sphere, is an imaginary sphere with radius 1/λ and which passes through the origin of the reciprocal lattice (000). Only reciprocal lattice points which lie on the surface of this Ewald sphere (red) are recorded and produce diffraction spots on the detector.
2.2.1.2 Electron density as a Fourier sum

Sinusoidal waves can be described by a periodic function defined by Equation 2 and Equation 3 where \( f(x) \) specifies the vertical height of the wave at any position \( x \) along the wave, \( h \) is the frequency (which is proportional to the inverse of the wavelength), \( F \) is the amplitude and \( \alpha \) is the phase, or position of the wave with respect to the origin [157,161]. The variable \( x \) and the constant \( \alpha \) are fractions of the wavelength expressed in angles.

**Equation 2**
\[
f(x) = F \cos(\pi hx + \alpha)
\]

**Equation 3**
\[
f(x) = F \sin(\pi hx + \alpha)
\]

In 1822, Jean Baptiste Joseph Fourier showed that waves, or Fourier terms, can be added to give a complicated function, known as a Fourier sum. A general Fourier sum of waves can be described using complex numbers as in Equation 4. More simply, replacing the complex number in square brackets using Equation 5 (derived from complex number theory), where \( \theta = 2\pi(hx) \), gives Equation 6.

**Equation 4**
\[
f(x) = \sum_{h} F_h [\cos 2\pi(hx) + i \sin 2\pi(hx)]
\]

**Equation 5**
\[
\cos \theta + i \sin \theta = e^{i\theta}
\]

**Equation 6**
\[
f(x) = \sum_{h} F_h e^{2\pi i(hx)}
\]

In X-ray crystallography, the scattered X-rays waves give rise to the recorded diffraction pattern. The scattered X-rays are described by three-dimensional waves of the general form \( f(xyz) \). Each recorded reflection, or diffraction spot, is the sum of diffractive contributions.
from all atoms that lie on the crystal lattice plane of the incident X-ray. Diffraction spots are described by the structure factors $F_{\iota\kappa\ell}$. Each structure factor can be written as a Fourier sum in which each term gives the contribution of one atom to the reflection $hkl$. Therefore, Fourier transform of all $F_{\iota\kappa\ell}$ terms of a diffraction experiment can be used to derive the protein’s atomic structure. The structure factor equation (Equation 7) is used to calculate each $F_{\iota\kappa\ell}$ from the sum of the waves scattered by the $i$ atoms with coordinates $xyz$.

Equation 7

$$F_{h\kappa\ell} = \sum_i f_i e^{2\pi i (hx+ky+iz)}$$

Since it is the electrons in a protein that mainly scatter X-rays it is more accurate to refer to the Fourier transform of $F_{\iota\kappa\ell}$ as giving rise to the electron density ($\rho$) of a protein, not its atomic structure per se. The inverse Fourier transform of all $F_{\iota\kappa\ell}$ terms defines the electron density at point $xyz$. This is denoted as $\rho(xyz)$. $\rho(xyz)$ can be derived from the diffraction pattern using Equation 8, where $V$ is the volume of the unit cell.

Equation 8

$$\rho (xyz) = \frac{1}{V} \sum_n \sum_k \sum_l F_{h\kappa\ell} e^{-2\pi i (hx+ky+iz)}$$

Calculation of $\rho(xyz)$ gives rise to the electron density map of a protein. This is essentially the three-dimensional space occupied by the orbiting electrons of the ordered atoms of a protein within a crystal. Once the electron density map is obtained, one can proceed with model building and refinement.

2.2.1.3 Unit cell symmetry and space groups

Unit cell parameters are defined by vectors $a$, $b$, $c$ with angles $\alpha$, $\beta$, $\gamma$ and length Å [157,161]. The operations that can be used to generate the contents of all the unit cells of the lattice, known as crystal symmetry, are described by space groups. These can be derived from symmetry and systematic absences in the diffraction pattern. When solving an X-ray crystal structure, only the contents of the asymmetric unit are refined. Crystal symmetry operators are used to generate the remainder of the unit cell and crystal lattice contents.
In total, mathematically, there are 230 possible space groups. Since protein molecules are chiral and incompatible with mirror symmetry the number of relevant space groups is 65 (Table 2-1) [162]. Based on their minimal internal symmetry, crystals are classed into seven lattice systems. These are triclinic, monoclinic, orthorhombic, tetragonal, rhombohedral, hexagonal and cubic. Additionally, crystals have one of the six lattice centring types: primitive (P), body-centred or internal (I), face-centred (F) and A, B or C-face centred (A, B or C respectively). Coupling of a lattice system with a lattice centring can be used to describe the 14 Bravais lattices. Finally, point groups are a set of symmetry operations (rotations and reflections), which move all the positions of the unit cell around a fixed central point to superimpose them onto identical motifs in the lattice. There are 32 crystallographic point groups, which fit in with the 14 Bravais lattice symmetries.
<table>
<thead>
<tr>
<th>Crystal system</th>
<th>Lattice property</th>
<th>Minimum internal symmetry</th>
<th>Point group</th>
<th>Bravais type</th>
<th>Chiral space group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Triclinic</td>
<td>$a \neq b \neq c$ $\alpha \neq \beta \neq \gamma \neq 90^\circ$</td>
<td>None</td>
<td>1</td>
<td>P</td>
<td>P1</td>
</tr>
<tr>
<td>Monoclinic</td>
<td>$a \neq b \neq c$ $\alpha = \gamma = 90^\circ$ $\beta \neq 90^\circ$</td>
<td>2-fold rotation axis parallel to b axis</td>
<td>2</td>
<td>C</td>
<td>C2</td>
</tr>
<tr>
<td>Orthorombic</td>
<td>$a \neq b \neq c$ $\alpha = \beta = \gamma = 90^\circ$</td>
<td>3 perpendicular, non-intersecting 2-fold axes</td>
<td>222</td>
<td>I</td>
<td>I222, I2121, C222, C2221</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>F</td>
<td>F222</td>
</tr>
<tr>
<td>Tetragonal</td>
<td>$a = b \neq c$ $\alpha = \beta = \gamma = 90^\circ$</td>
<td>4-fold rotation axis parallel to c axis</td>
<td>422</td>
<td>P</td>
<td>P4, P41, P42, P43</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>I</td>
<td>I4, I41</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>F</td>
<td>F422</td>
</tr>
<tr>
<td>Trigonal</td>
<td>$a = b \neq c$ $\alpha = \gamma = 90^\circ$ $\beta = 120^\circ$</td>
<td>3-fold rotation parallel to c axis</td>
<td>32</td>
<td>P</td>
<td>P312, P321, P312, P321, P312, P321</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R</td>
<td>R32</td>
</tr>
<tr>
<td>Hexagonal</td>
<td>$a = b \neq c$ $\alpha = \beta = \gamma = 90^\circ$</td>
<td>6-fold rotation parallel to c axis</td>
<td>622</td>
<td>P</td>
<td>P6, P61, P62, P63, P65, P65</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>P</td>
<td>P62, P62, P65, P62, P65, P65</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>I</td>
<td>I432, I412</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>F</td>
<td>F432, F41</td>
</tr>
<tr>
<td>Cubic</td>
<td>$a = b = c$ $\alpha = \beta = \gamma = 90^\circ$</td>
<td>Four 3-fold axes along space diagonals</td>
<td>432</td>
<td>I</td>
<td>I432, I412</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>F</td>
<td>F432, F41</td>
</tr>
</tbody>
</table>

Table 2-1. Crystal space groups
The lattice properties, internal symmetry, point group and Bravais lattice translation of the 65 chiral space groups of protein crystals are shown.
2.3 The phase problem

Each structure factor $F_\alpha$ describes a diffracted ray recorded as reflection $hkl$. $F_\alpha$ terms are dependant on the frequency, amplitude ($|F_\alpha|$) and phase ($\alpha_\alpha$) of the scattered X-ray waves [157,161].

The $hkl$ indices of the set of parallel planes that produce the reflections are the frequency terms. The amplitude is proportional to the square root of the measured intensity, $I_\alpha$, for each reflection. However the phase cannot be derived solely from the diffraction pattern of a native crystal. Since there is no experimental technique to directly measure phases, this information is lost during the diffraction experiment. Several methods for obtaining the phase, or ‘phasing’, the diffraction data exist: molecular replacement or experimental phasing by isomorphous replacement or anomalous scattering [163].

2.3.1 Molecular replacement

Molecular replacement relies on the use of phases from the structure factors of a previously solved structure, the search ensemble. Programs compare Patterson maps obtained from the diffraction pattern of the unknown protein to that calculated from the search ensemble in different orientations [164,165]. These are obtained by carrying out rotation and translation functions of the search ensemble. Complex algorithms are used to compute the orientation, outputted in Euler angles, which results highest correlation between the two Patterson maps. As such, phase estimates for the unknown protein can be obtained. Having a search ensemble though is necessary for phasing by molecular replacement. Typically this is a homologous protein, usually with at least 20% sequence similarity, a protein with a similar fold or in silico models. If an appropriate search ensemble is not available, experimental phasing must be used.

2.3.2 Experimental phasing

The scattering factor of atoms is proportional to the number of electrons it has, and thus its atomic number. Experimental phasing by isomorphous replacement relies on the strong
scattering of X-rays by heavy atoms compared to the carbon, oxygen and nitrogen atoms of proteins (Figure 2-3) [166]. Single Isomorphous Replacement (SIR) and Multiple Isomorphous Replacement (MIR) methods, respectively, make use of one or multiple heavy atom derivative crystals. These derivative crystals must be isomorphous in terms of space group, cell dimensions and crystal packing for useful phasing information to be obtained.

Additionally, some heavy atoms absorb X-rays at the energy used in diffraction experiments, which leads to the breakdown of Friedel’s law (Figure 2-3) [166]. Friedel’s law, named after Georges Friedel, states that Bragg reflections related by inversion through the origin, termed \( +h+k+l \) and \( -h-k-l \), have an equal amplitude and opposite phase. The difference in Friedel pairs caused by X-ray absorbance by heavy atoms is known as anomalous scattering. The intensity of the individual reflections is also affected by a change in the wavelength of the impinging X-ray. The latter are known as dispersive differences [166].
Figure 2-3. Experimental phasing

Diagram showing structure factors from native proteins (F(P)), the heavy atom (F(H)) and heavy atom derivative proteins without (F(PH)) and with (F”(PH)) anomalous scattering. These are shown for a Friedel pair (+h+k+l and -h-k-l). F(P)+ is the reflection of F(P) in the real axis; no experimental phase information can be extracted from native protein scattering. F(PH) is the contribution of scattering from the native structure F(P) and the heavy atom F(H). This difference in the structure factors of native and heavy atom derivative crystals is exploited in SIR and MIR experiments. At wavelengths near the heavy atom’s absorption edge (λ”), an imaginary component known as anomalous scattering (F”) is added to the structure factor of the heavy atom derivative to give F”(PH). This causes Friedel’s law to break down. Anomalous differences are used for phasing in SAD and MAD experiments.

F(P)+

F(P)

F(H)

F(PH)+

λ F(PH)+

λ F”(PH)+

λ” F” (+)

-h-k-l

+h+k+l

real

imaginary

F(PH)-

λ” F” (–)

F(P)-

F(H).
Single Anomalous Dispersion (SAD) experiments are carried out by collecting data at a wavelength near the heavy atom’s absorption edge. At this peak wavelength, its anomalous scattering properties are at a maximum. Multiple Anomalous Dispersion (MAD) makes use of wavelength-dependent anomalous scattering differences for any given heavy atom. Typically MAD datasets are collected at the peak, remote and sometimes inflection energy points.

For experimental phasing, firstly heavy atoms are located by creating Patterson maps [166,167]. Patterson functions represent Fourier sums without the phases and thus depend solely on the intensity of the reflections. In difference Patterson functions, the difference between the native and derivative amplitude of each term is calculated. The vectors between atoms are plotted in Patterson maps. Heavy atoms will result in large peaks in difference Patterson maps and, together with the cell symmetry of the crystal, can be used to locate the position of the heavy atoms. Since the diffractive contributions of all atoms to a reflection are additive, Equation 9 can be used to phase the native dataset. $F_{PH}, F_P, F_H$ refer to protein plus heavy atom (derivative), protein (native) and heavy atom structure factors respectively.

Equation 9

$$F_{PH} = F_P + F_H$$

2.4 Structure refinement

The first maps obtained after phasing are often poorly informative and insufficiently resolved to obtain a reliable structure. This is due to errors in phase measurements. Refinement involves minimization of differences between the measured diffraction intensities and the intensities predicted by a model [168].

In real space, the atomic coordinates of a model are adjusted by manually fitting it into to the electron density maps. In reciprocal space, comparison of the structure factors derived from the model with the measured experimental intensities allow for phase improvement. This involves the use of complex algorithms which employ least-square fitting, energy refinement and/or Bayesian methods. The correlation between real and reciprocal space is Fourier
transformation. Adjustment of the model can improve the accuracy of the phases, which in turn leads to a better electron density map allowing for better model fitting and so on.

The observed ($|F_{\text{obs}}|$) and calculated ($|F_{\text{calc}}|$) structure factors give rise to the $F_o$ and $F_c$ maps respectively [168]. Difference maps are generated for model building and refinement. An $F_o$-$F_c$ map is used to visualize differences between density calculated from the measured intensities and the model. It clearly emphasizes errors in the model and points out regions for which the model contributes to too much or too little density compared to that which the asymmetric unit actually contains. A more easily interpreted map is the $2F_o-F_c$ difference map. Here the density is positive everywhere. The influence of the model is reduced but still present. Finally, figure of merit, or $\sigma A$, weighted maps are given by $mF_o-DF_c$ where $m$ is the figure of merit for each model phase and $D$ is the overall estimate of the atomic coordinate errors in the current model.

As the accuracy of the model improved, $|F_{\text{obs}}|$ and $|F_{\text{calc}}|$ should converge [168]. Their differences are reflected in residual factors (R factor) statistics. Since the refinement process improves the atomic model of a given structure to make it fit better to the experimental data the R factor gradually improves. However a decrease in the R factor may be due to overfitting of the model to the data rather than a genuine improvement in the accuracy of the model. R-free is used as a less biased way to judge the model’s accuracy. Before refinement begins, 5-10% of the experimental observations are removed from the dataset. Refinement is performed using the remaining 90-95% of the observations. The R-free value is then calculated by seeing how well the model predicts the 10% that were not used in refinement.

Several rounds of refinement are often required to obtain the final structure. Other methods such as solvent flattening, density modification, non-crystallographic symmetry averaging and use of a partial model are also often useful during refinement.
3 Perspective
Over the last thirty years, the UPR has been characterised in increasing detail. Advances in structural biology, and biophysics techniques, have played a crucial role in providing detailed insights of UPR ER stress sensing and signal transduction pathways. However, despite intense efforts, the precise modes by which the UPR mediators sense the accumulation of unfolded proteins and how this leads to their activation remains poorly understood and structural insights are limited.

In this thesis, the work involved in solving the novel X-ray crystal structure of PERK luminal domain in two different states is described. Furthermore, the results of a comprehensive biophysical study of the interaction between BiP chaperone and the luminal domains of PERK and IRE1, and how this is disrupted by Cn1 unfolded protein, are presented. Together these results provide a new paradigm for ER stress sensing and activation of UPR signalling.

Aberrant protein folding and UPR signalling are a major cause for disease. Obtaining a detailed understanding the mechanistic aspects of UPR activation will open novel avenues for targeting the UPR to develop drugs against many important human pathologies.
4 Methods
4.1 Cloning and expression

H. sapiens perk, ire1 and bip genes were amplified from commercial complementary DNA (cDNA) libraries (Clontech). *Mus musculus* perk, *Drosophila melanogaster* perk and *H. sapiens* Cn1 sequences were amplified from synthetic genes produced by GeneArt Gene Synthesis services (Life Technologies). All polymerase chain reaction (PCR) amplification reactions were carried out as following: 1 μl DNA template, 15 pmol forward primer, 15 pmol reverse primer, 2 μl dimethyl sulfoxide (DMSO), 25 μl 2X Phusion Flash II DNA polymerase (ThemoScientific) and made up to a total volume of 50 μl with water. The following PCR protocol was used: 2 minutes (‘) initial denaturation at 95 °C, 30 amplification cycles (30 seconds (“) denaturation at 95 °C, 30” annealing at 56 °C, 1’ amplification at 72 °C) followed by a final 5’ amplification at 72 °C. Amplified DNA was then purified using QIAquick PCR Purification Kit (Qiagen) and digested overnight at 37 °C with the appropriate restriction enzymes (NEB). Primer sequences and restriction enzymes used for cloning can be found in Appendix Table 10-1.

Genes were inserted into the appropriate expression vector. In this study pTWO-E, with Ampicillin (Amp) resistance gene, and pET-26b, with Kanamycin (Kan) resistance gene, were used (Appendix Figure 10-1). pTWO-E is a modified version of the pET-17b vector (Merck), which contains an additional PreScission Protease cleavable N-terminal His6-tag and Nhel restriction site. All perk, ire1 and bip genes were cloned into pTWO-E. pET-26b was used for the cloning of untagged Cn1 antibody segment only. Vectors were digested with the same restriction enzymes as the PCR product for 2 hours (h) at 37 °C. Subsequently the 5’ phosphate was removed from the digested vectors to avoid self-ligation by incubation with Calf Intestine Alkaline Phosphatase (NEB) for 1 h at 37 °C. The digested and purified gene inserts were ligated into cut vectors using T4 DNA ligase (NEB). Ligation reactions were incubated for 1 h at room temperature (RT) after which they were transformed into *E. coli* TOP10 chemically competent cells (Life Technologies).

The following protocol was used for all transformations of DNA into chemically competent cells. 1 μl of DNA was incubated with 50 μl of cells on ice for 20’ followed by heat shock at 42 °C for 30”. Cells were transferred to ice for a further 2’ after which 250 μl of Luria-Bertani (LB) media was added. Cells were shaken at 220 rpm at 37 °C for 45’ and then plated onto LB
agar plates containing the appropriate antibiotic for selection. All antibiotic concentrations used were as following: 100 µg/ml Amp, 30 µg/ml Kan, 35 µg/ml Chloramphenicol (Cam). Plates were incubated overnight at 37 °C.

The following day individual colonies were used to inoculate 5 ml of LB containing the appropriate antibiotic. After overnight growth at 37 °C with shaking at 220 rpm, the plasmid DNA was extracted using the QIAprep Miniprep Kit (Qiagen). Test restriction digests were carried out to verify the presence of the insert by cutting with the appropriate restriction enzymes for 2 h at 37 °C. Digestions products run on a 1% agarose gel for 1 h at 75 V and visualized using GelRed stain (Life Technologies). Plasmids containing DNA inserts of the expected size were sequence verified (Eurofins MWG Operon).

### 4.2 Site-directed mutagenesis

For site-directed mutagenesis, Phusion Flash II DNA polymerase was used according to the manufacturer’s protocol. The following PCR protocol was used: 5’ initial denaturation at 95 °C followed by 15 cycles of plasmid amplification (45” denaturation at 95 °C, 45” annealing at 56 °C, 5’ amplification at 72 °C). The methylated parental plasmid was digested by incubation with DpnI enzyme for 1 h at 37 °C. The DNA was purified using the QIAquick PCR Purification Kit (Qiagen). 1 µl of DNA was transformed into TOP10 E. coli chemically competent cells and the plasmid was isolated as previously described. All plasmids were sequence verified.

### 4.3 PERK, IRE1 and BiP expression and purification

The following protocol was used for expression and purification of all PERK, IRE1 and BiP proteins used in this study. All buffers used are summarised in . Plasmids were transformed into E. coli Rosetta2 (DE3) chemically competent cells (Merck) and plated onto LB (Amp/Cam) agar plates. The plates were incubated overnight at 37 °C and a single colony was used to inoculate 200 ml of LB (Amp/Cam) culture media which was incubated overnight at 37 °C with shaking at 220 rpm. The following day, each 1 L of LB (Amp/Cam) media was inoculated with 10 ml of starter culture and grown at 37 °C with shaking at 220 rpm until the optical density (OD) at 600 nm (OD600) reached 0.6. Cells were induced with 0.5 mM Isopropyl β-D-
1-thiogalactopyranoside (IPTG) (Life Technologies), grown overnight at 22 °C with shaking at 220 rpm after which they were harvested by centrifugation at 4000 g for 15’. All subsequent steps were carried out at 4 °C to avoid protein degradation.

Cell pellets were resuspended in buffer A supplemented with Complete ethylenediaminetetraacetic acid (EDTA)-free Protease Inhibitor Cocktail (Roche) and 25 μg/ml DNase (Sigma-Aldrich) for lysis. Where specified, lysis buffer for BiP proteins also contained additional 0.1% polyethylenimine (PEI) (Sigma-Aldrich) or 5 mM ATP (Sigma-Aldrich) plus 10 mM MgCl₂. Lysis was carried out by 10’ sonication at 50% amplitude. Cell debris were pelleted by centrifugation at 40000 g for 1 h and soluble proteins were further purified by chromatography steps. The quality of protein samples was visualized after each step by Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE). Samples were supplemented with Laemmli 2X buffer (Sigma-Aldrich), denatured for 10’ at 100 °C and run on a 4-12% Bis-Tris gel for 1 h at 180 V visualized by Coomassie Blue staining. SeeBlue Plus2 Pre-Stained Standard or HiMark Pre-Stained Protein Standard (Life Technologies) was used; molecular weights (MW), in kDa, of marker bands are indicated for each gel shown.

For affinity purification, a pre-packed 5 ml HiTrap TALON crude columns (GE Healthcare), containing cobalt ions (Co²⁺), was equilibrated with buffer A. Soluble cell lysate fractions were loaded onto the TALON column after which it was washed with 100 ml of buffer A containing additional 5 mM imidazole. Bound proteins were eluted with 30 ml of buffer A plus additional 250 mM imidazole. When cleavage of the His₆-tag was desired, 10 units (U) of Prescission Protease (GE Healthcare) were added per 1 mg of purified protein. Samples were dialyzed overnight against 2 L of buffer B at 4 °C. Samples in which the His₆-tag was cleaved were passed through a pre-equilibrated TALON column as previously. The flow-through containing only untagged protein species was collected and used in further purification steps.

Proteins were then purified by anion-exchange chromatography using a 5 ml HiTrap Q HP column (GE Healthcare). The ion-exchange step was omitted for BiP protein samples. Briefly, protein samples were loaded on the column pre-equilibrated with buffer C, washed with 30 ml of buffer D, which contains additional 150 mM NaCl. A shallow salt gradient (between 200 mM and 500 mM NaCl) was applied so as to separate bound protein species according to their charge. This was achieved by running samples in buffer B and applying a gradient of 20-
50% buffer E. Peak fractions were pooled, concentrated using Amicon Ultra Centrifugal Filter Units (Merck) of the appropriate MW cut off (at least half of the protein size).

Samples were further purified by size exclusion chromatography (SEC). Samples were loaded onto a HiLoad 16/60 Superdex 200 column (GE Healthcare) pre-equilibrated with buffer F. Fractions containing pure protein samples were pooled, concentrated as appropriate and flash frozen in liquid nitrogen for subsequent studies. All ion-exchange and SEC profiles are shown as normalized curves to allow for comparison.

### 4.4 C_H1 expression and purification

C_H1 antibody domain was expressed in inclusion bodies and purified as previously published [25]. Briefly, the protein was expressed in BL21 (DE3) chemically competent cells (Merck) in LB (Kan) for overnight at 37 °C, similarly to the protocol described above for PERK, IRE1 and BiP proteins. The harvested cell pellet was resuspended in buffer G supplemented with Complete EDTA-free Protease Inhibitor Cocktail, 25 µg/ml deoxyribonuclease (DNase), 2% triton X-100 (Sigma-Aldrich), 20 mM EDTA for lysis. After sonication with a Q125 Sonicator (Qsonica, USA) for 10” at 50% amplitude, the sample was centrifuged at 40000 g for 1 h. The supernatant was discarded and the pellet, which also contains the inclusion bodies where the insoluble C_H1 is found, was resuspended in 50 ml of buffer H and incubated with shaking for 30’ at 25 °C. The insoluble cell fragments were removed by centrifugation at 40000 g for 1.5 h. The resulting supernatant was applied to a 5 ml HiTrap SP HP column (GE Healthcare) pre-equilibrated with buffer I. C_H1 protein did not bind to the column in these conditions. The flow through was collected. Refolding was carried out by dialyzing overnight against 2 L of buffer J. The following day, aggregates were removed by centrifugation at 40000 g for 1 h, the soluble protein was concentrated and loaded on a HiLoad 16/60 Superdex 75 column (GE Healthcare) pre-equilibrated with buffer K. Fractions containing pure C_H1 protein were pooled, concentrated and flash frozen in liquid nitrogen.
<table>
<thead>
<tr>
<th>Buffer</th>
<th>Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>50 mM Na-HEPES (pH 7.8), 400 mM NaCl, 10% glycerol</td>
</tr>
<tr>
<td>B</td>
<td>50 mM Na-HEPES (pH 7.8) and 10% glycerol</td>
</tr>
<tr>
<td>C</td>
<td>50 mM Na-HEPES (pH 7.8), 10% glycerol and 2 mM TCEP</td>
</tr>
<tr>
<td>D</td>
<td>50 mM Na-HEPES (pH 7.8), 150 mM NaCl, 10% glycerol and 2 mM TCEP</td>
</tr>
<tr>
<td>E</td>
<td>50 mM Na-HEPES (pH 7.8), 1M NaCl, 10% glycerol and 2 mM TCEP</td>
</tr>
<tr>
<td>F</td>
<td>50 mM Na-HEPES (pH 7.8), 200 mM NaCl, 10% glycerol and 2 mM TCEP</td>
</tr>
<tr>
<td>G</td>
<td>75 mM Tris-HCl (pH 7.0), 500 mM NaCl</td>
</tr>
<tr>
<td>H</td>
<td>50 mM Tris-HCl (pH 7.5), 8 M Urea, 10 mM β-mercaptoethanol, 10 mM EDTA</td>
</tr>
<tr>
<td>I</td>
<td>50 mM Tris-HCl (pH 7.5), 5 M Urea, 10 mM β-mercaptoethanol, 10 mM EDTA</td>
</tr>
<tr>
<td>J</td>
<td>50 mM Tris-HCl (pH 7.5), 100 mM NaCl, 1 mM β-mercaptoethanol, 1 mM EDTA</td>
</tr>
<tr>
<td>K</td>
<td>50 mM Na-HEPES (pH 7.8), 50 mM NaCl and 1 mM TCEP</td>
</tr>
</tbody>
</table>

Table 4-1. Purification buffers composition
All the buffers used in this study are listed.
4.5  Seleno-Met labeling of PERK

Incorporation of Seleno (Se)-Methionine (SeMet) instead of Methionine residues in PERK recombinant proteins was carried out using a modified version of the protein expression protocol described above. Briefly, DNA plasmids were transformed into the Methionine auxotroph *E. coli* B834 (DE3) chemically competent strain (Merck) (Amp). Overnight cultures were made by inoculating a single colony in M9 minimal media supplemented with a nutrient mix (containing glucose, vitamins and essential amino acids, with the exception of Methionine) plus 40 µg/ml Methionine (Molecular Dimensions) (Amp). The overnight culture were pelleted by centrifugation at 4000 g for 15’ and the cell pellet was gently washed by resuspending it with 100 ml of 1X phosphate buffered saline (PBS) (pH 7.5). This step was repeated five times to thoroughly remove any Met for subsequent steps. After the final centrifugation step, the cell pellet was resuspended in 10 ml 1X PBS (pH 7.5) and 1 ml was used to inoculate each 1 L of M9 minimal media supplemented with the nutrient mix plus 40 µg/ml SeMet (Molecular Dimensions) (Amp). The remainder of the protein expression and purification protocol was identical to that of wild-type proteins.

4.6  Calibration of HiLoad 16/60 Superdex 200 column

For approximation of protein MW from SEC profiles, a High MW Calibration Kit (GE Healthcare) was used. The calibration protein stocks provided in the kit were made up in buffer F as specified by the manufacturer. Samples were mixed and applied to the HiLoad 16/60 Superdex 200 column (GE Healthcare) pre-equilibrated with buffer F. The void volume (V0) was separately determined by measuring elution volume (Ve) of Blue Dextran 2000. A calibration curve was calculated by plotting V0 – Ve values against log MW and fitting a line of nonlinear regression. The function obtained was used to estimate MW of samples from their elution volumes on the HiLoad 16/60 Superdex 200 column (Appendix Figure 10-2).
4.7 Multi-Angle Light Scattering

Multi-Angle Light Scattering (MALS) was used to measure the absolute MW of protein species. An Agilent 1260 (Agilent Technologies) system equipped with a miniDAWN TREOS (Wyatt Technologies) Light Scattering (LS) detector and an Optilab T-rEX (Wyatt Technologies) Refractive Index (RI) detector was used. A Superdex 200 PC 3.2/30 column (GE Healthcare) was pre-equilibrated with the appropriate buffer until LS and RI readings were stable. Buffer F was used for analytical SEC MALS of single protein species; buffer K was used for SEC MALS of protein-protein and peptide-protein samples. 100 µl of protein at 100 µM, plus 500 µM ΔEspP where relevant, were injected and LS and RI values were recorded. Peaks of interest were manually selected and the data was analysed using the ASTRA software (Wyatt Technologies) to calculate MW values and the polydispersity of the sample.

4.8 Matrix-assisted laser desorption/ionization-time of flight

In gel trypsin digestion samples were separated on a 4-12% Bis-Tris gel for 1 h at 180 V and stained using CuCl₂ negative staining. The peptide bands of interest were excised from the gel under sterile conditions and sent for in gel trypsin digestion and matrix-assisted laser desorption/ionisation-time of flight (MALDI-TOF) analysis by Dr Paul Hitchens at the Mass Spectrometry facility, Imperial College London.

MALDI-TOF experiments of SeMet derivative proteins were similarly carried out by Dr Len Packman at the PNAC Facility, University of Cambridge (with the omission of trypsin digestion).

4.9 Native gel shift assays

Protein samples were prepared as described in the results as desired (proteins alone or complexes). Coomassie native 4X loading dye was added (1% Bromophenol Blue (Sigma-Aldrich), 40% glycerol, 250 mM Tris-HCl (pH 6.8)). Samples were run on a 3-8% Tris-Acetate gel for 2 h at 150 V at 4 °C in Tris-Glycine Native Running Buffer (Life Technologies). Gels were visualized by Coomassie Blue staining.
4.10 Chemical cross-linking

Cross-linking is used to capture non-covalent interactions that exist between proteins that come in close proximity in solution. The chemical cross-linker ethylene glycolbis(succinimidylsuccinate) (EGS) (ThermoScientific) was used. The spacing between the two NHS-ester moieties of EGS is 16.1 Å. A 2.5 mM EGS stock was made in DMSO immediately prior to use since its reactive N-hydrosuccinimide (NHS)-ester moiety is easily hydrolysed. 1 µl of 2.5 mM EGS was thoroughly mixed with 9 µl of protein in buffer K at 11 µM so as to have approximately a 25-fold molar excess of EGS. 1 µl of DMSO was added instead of EGS for control experiments. The reaction was allowed to proceed for 1 h at RT after which it was quenched with 1 µl of 1 M Tris-HCl (pH 7.5) for 15’ at RT. Samples were run on a 4-12% Bis-Tris gel for 1 h at 180 V and visualized by Coomassie Blue staining.

4.11 Western blotting

Samples were first separated by running on a 4-12% Bis-Tris gel for 1 h at and were then transferred onto enhanced chemilluminescence (ECL) Nitrocellulose Membrane (GE Healthcare) at 200 mA for 80’ using a Tris-glycine buffer (pH 8.5). Blots were blocked for 2 h at 4 °C with 40 ml of a solution containing 2% powdered milk in PBS/2% Tween-20. The volume was reduced to 10 ml and 1 µl of mouse-α-His6 primary antibody (Life Technologies) was added for 1 h at 4 °C. Blots were carefully washed three times with 40 ml of PBS/2% Tween-20. The blot was incubated for a further 1 h at 4 °C with 1 µl of horseradish peroxidase (HRP)-conjugated sheep-α-mouse-IgG secondary antibody (Life Technologies) in 10 ml PBS/2% Tween-20. The blot was washed again as previously. Amersham ECL Plus Western Blotting Detection Reagents (GE Healthcare) was used to generate the chemiluminescent signal from the HRP. HRP catalyses the oxidation of the Lumigen PS-3 Acridan substrate to produce acridinium ester intermediates. These react with peroxide in alkaline conditions to produce a chemiluminescent signal with emission at 450 nm. This signal is detected by exposing the blots to a Chemiluminesce ECL HyperFilm (GE Healthcare) for 30” and then developed manually on a X-OMAT Film Developer (Kodak, Japan).
4.12 Pull down experiments

All pull down experiments were carried out in 5 ml gravity flow columns (ThermoScientific). 50 µl of TALON resin (Clonetech) pre-equilibrated with buffer K was incubated with 50 µl of purified His6-BiP protein at 50 µM for 1 h at RT. The resin was washed with 1 ml of buffer K to remove any unbound His6-BiP. His6-BiP was replaced by buffer K in control experiments. 50 µl of purified untagged PERK, IRE1 or Cn1 proteins at 250 µM were then added and incubated for 1 h at RT. The resin was extensively washed with a total of 5 ml of buffer K in 500 µl volumes. For competition pull-downs, 50 µl of PERK, IRE1, Cn1 or ΔEspP as appropriate at 250 µM in buffer K, supplemented with 5 mM ADP plus 10 mM MgCl2, were then added. Reactions were incubated for a further 1 h at RT and washed as previously with buffer K plus 5 mM ADP and 10 mM MgCl2. Finally, the resin was resuspended with 50 µl of buffer, spun at 10000 g for 5’ and the resulting supernatant was analysed on a 4-12% Bis-Tris gel for 1 h at 180 V and visualized by Coomassie Blue staining.

4.13 Differential Scanning Fluorimetry

Differential Scanning Fluorimetry (DSF) is used to determine the thermal stability of proteins and thus is a useful method to screen for binding of low-MW ligands such as peptides or nucleotide. DSF experiments were carried out using the following general protocol, and adapted depending on the assay. These comprised of screening for peptide binding to PERK and IRE1 versus nucleotide binding to BiP protein. The 500X Sypro Orange stock was diluted to 10X in DMSO. Each reaction consisted of 40 µl of the protein plus co-factors (peptide or nucleotides) as appropriate plus 1 µl of 10X Sypro Orange, added immediately prior to the assay measurements as it is light-sensitive. PERK and IRE1 plus peptide reactions were made by mixing 200 µM protein plus 500 µM peptide in buffer K. BiP plus nucleotides reactions were made by mixing 200 µM protein in buffer F plus 5 mM of the appropriate nucleotide (ATP, ADP, AMPPNP or buffer F for control experiments) and 10 mM MgCl2. Samples were mixed thoroughly and placed in a 96-well PCR plate (ThermoScientific). Buffer was added instead of protein in control experiments to monitor Sypro Orange background fluorescence. Experiments were carried out by heating samples from 20 to 95 °C in increments of 2 °C/min in a Mastercycler Ep Realplex (Eppendorf, Germany). Fluorescence at 569 nm was measured.
The melting transition curves were fitted with a Boltzmann equation and the inflection point was calculated to obtain melting temperature (T_m) values. Each experiment was repeated in triplicate and average values were used.

4.14 Microscale Thermophoresis

Microscale Thermophoresis (MST) was used to measure interaction between molecules in solution. MST has the advantage of allowing quantitative measurement of interactions directly in solution without the need of immobilization (as is the case in Surface Plasmon Resonance (SPR)), requires relatively small amounts of protein (compared to Isothermal Titration Calorimetry (ITC)) and is useful particularly to efficiently screen many variables with consistency in the experimental setup.

MST experiments were carried in buffer K (and where specified additional 5 mM ATP, ADP or AMPPNP and 10 mM MgCl_2 was included) at 25 °C using a Monolith NT.115 instrument (NanoTemper Technologies). Proteins were labelled by primary amine coupling using the Monolith NT Protein labelling Kit Red-NHS (NanoTemper Technologies). The concentration of labelled protein was adjusted so as to obtain fluorescence readings between absolute values of 200 and 2000 (the range of the instrumentation). All subsequent experiments with that particular labelled protein were carried out using the corresponding dilution. 10 µl of the labelled protein was thoroughly mixed with 10 µl of 16 two-fold serial dilutions of the unlabelled binding partner. Buffer K was used for all experiments and supplemented with 5 mM nucleotides and 10 mM MgCl_2 where specified. All experiments were carried out in standard treated capillaries with 100% LED power and 80% infrared (IR)-laser unless otherwise stated.

With the Monolith NT.115 (NanoTemper Technologies, Germany) set up, a fluorescently labelled molecule is titrated with an unlabelled binding partner in a series of capillaries. The reaction is allowed to equilibrate (Fluorescence ‘cold’ (F_cold)), after which an IR laser is used to create a microscopic temperature gradient that initiates thermophoresis of the molecules in the capillaries until equilibration is reached (Fluorescence ‘hot’ (F_hot)). When the IR laser is switched off the back-diffusion of the molecules is observed. A Light Emitting Diode (LED) is used to excite the fluorescent molecule and the changes in fluorescence at a specific and
constant point in the capillary is monitored. Curves are normalized to take into account differences in initial fluorescence values. Normalised $F_{\text{hot}}/F_{\text{cold}}$ values were plotted against the concentration of the titrant, in the logarithmic scale. The NanoTemper Analysis 1.2.101 software was used to fit the data with a nonlinear solution of the law of mass action to determine dissociation constant ($K_D$) values. Each measurement was repeated in three independent experiments and $K_D$ values were averaged and Standard Error (SE) values were calculated. $F_{\text{hot}}/F_{\text{cold}}$ values were normalized for each experiment so as to compare them independently of absolute fluorescence value. These are shown as fraction of labelled protein A bound to unlabelled protein B (assuming this to be 0 and 1 at minimum and maximum equilibriums respectively).

### 4.15 Isothermal Titration Calorimetry

ITC was used to determine the affinity and stoichiometry of protein-protein and protein-peptide interactions. For ITC experiments, samples were dialyzed against 2 L of buffer K, supplemented with 5 mM ADP and 10 mM MgCl$_2$ where specified, overnight at 4 °C. Proteins were then concentrated or diluted to the appropriate concentration: syringe sample concentrations ranged from 10-40 µM and cell sample concentrations were 9-10-fold in molar excess. Precise sample concentrations were measured and are indicated in the results section for each ITC experiment. Buffer and protein samples were degassed under vacuum for 5’ with gentle stirring. Samples were loaded into the syringe and sample cell as appropriate in a VP-ITC MicroCalorimeter (GE Healthcare). Titrations consisted of 10 µl injections of syringe sample, for a total of 30, at 300” intervals at 25 °C. For each experiment, the heat of dilution, as determined by titrating the syringe sample, at the same concentration, into dialysis buffer was subtracted from the raw titration data before analysis. Data were fit by least-squares procedure assuming one-site binding models using MicroCal Origin (version 7.0). $K_D$ values were averaged over three measurements; SE values are indicated.
4.16 Analytical ultracentrifugation

Purified protein samples made up to a final concentration of 30 μM, dialyzed overnight in buffer K at 4 °C, spun for 10’ at 40000 g and analysed by analytical ultracentrifugation (AUC). AUC experiments were carried out by Dr Katherine Stott at the Biochemistry Department of the University of Cambridge.

4.17 Crystallisation

4.17.1 PERK luminal domain crystallisation

Sitting drop and hanging drop vapour diffusion setups were used for crystallisation of PERK luminal domain proteins. Initial sitting drop nanolitre scale crystallisation experiments were set up using an automated Mosquito crystal robot (TTP LabTech, UK). Initial protein concentrations used varied between 5 and 25 mg/ml depending on the protein and are specified in the results section. Drops were set up in 100 nl protein plus 100 nl reservoir, 200 nl protein plus 100 nl reservoir and 100 nl protein plus 200 nl reservoir. The following commercially available crystallisation screens were tested: Crystal Screen 1 and 2, Index, SaltRx 1 and 2 (Hampton Research, USA), PACT premier, JCSG+, MemGold, Proplex, Morpheus, PGA Screen (Molecular Dimensions, USA) and Wizard 1 and 2 (Emerald Biosystems, USA). Plates were incubated at 4 or 18 °C. Drops were screened at regular intervals from 2 days and up to a year after initial setup. Crystal hits were optimized in 48 and 24-well sitting drop MRC plates or 24-well hanging drop plates by varying the concentration of protein and components of the crystallisation solution, drop and reservoir volumes, temperature as described in the results Chapter 6. In some instances, the Additive Screen (Hampton Research) was also carried out on pre-optimized crystallisation conditions.

4.17.2 Micro-seeding

Micro-seeding was achieved by transferring a single crystal into a Seed Bead kit (Hampton Research) containing 50 μl of the reservoir crystallisation solution. The crystal was crushed by vortexing for 2’ and the original seed stock was diluted 10⁴-fold using the same
crystallisation solution. This diluted seed solution was used instead of the reservoir buffer to set up crystallisation trays identically as described above. For matrix-microseeding the concentrated seed stock was added to the protein solution (1 to 50 ratio) immediately prior to setting up nanolitre scale crystallisation screens as above. All seed stocks were freshly made and not stored.

4.17.3 Crystal dehydration

Dehydration of crystals was achieved by a serial transfer of the coverslip holding the crystallised drop over reservoirs containing increasing concentrations of polyethylene glycol (PEG) 3350. PEG3350 concentration was increased by 2% w/v and incubated 8-12 h at each step up to a final 40% w/v PEG3500 concentration.

4.17.4 Cryoprotection and crystal harvesting

Crystals grown in conditions which were not cryoprotectant were transferred to a cryoprotectant solution prior to freezing. The following cryoprotectant reagents were added to Cryoprotectant solutions were made up which contained the reservoir components and additional 20% v/v glycerol, 30% w/v sucrose, 30% v/v ethylene glycol, 30% v/v 2-Methyl-2,4-pentanediol (MPD) and 30% w/v PEG400. Gentle cryoprotection is important as crystal diffraction is often sensitive to extensive handling and changes in its surrounding solution. Therefore, 1 µl of the cryoprotectant was first added to the crystallisation drop and crystals were then briefly transferred to 1 µl of pure cryoprotectant and then flash frozen in liquid nitrogen. Crystals that did not require cryoprotection were directly flash frozen in liquid nitrogen.

4.18 X-ray diffraction data

All X-ray diffraction data were collected at Diamond Light Source (Didcot, UK) on Macromolecular Crystallography beamlines (I-03, I04 and I04-1). Both charge-coupled detector (CCD) and Pilatus detectors were used. Native datasets were collected at the typical
working wavelength of 0.9795 Å. Diffraction data images were integrated using iMosflm and then merged and scaled using Scala (CCP4) [165].

4.19 Phasing

4.19.1 Heavy atom native gel shifts

For phasing by heavy atom derivatization, Heavy Atom Screens (Hampton Research) were used. To screen for heavy atom binding, a native gel shift assay was used. Fresh heavy atom 100 mM stocks were made up in water. The heavy atoms tested for each crystal are described in the results. Reactions containing 2 µl purified protein at 75 µM, 2 µl native loading buffer, 5 µl water and 1 µl of heavy atom stock at 100 mM were mixed and incubated at RT for 20’. Samples were then run on a pre-equilibrated 3-8% Tris-Acetate gel for 2 h at 100 V at 4 °C in Tris-Glycine Native Running Buffer (Life Technologies). Gels were visualized by Coomassie Blue staining.

4.19.2 Heavy atom soaking

For phasing by heavy atom derivatization, a soaking technique was used. The heavy atoms used are listed in the results section. For soaking, 200 mM stock solutions were made in water and then diluted to 50 mM in the reservoir buffer. The appropriate volume was added to crystallisation drops to reach the desired final heavy atom concentration. Final heavy atom concentrations used were 0.5, 1, 2, 5 and 10 mM. Crystals were left to soak for 30’, 2 h, 5 h or overnight after which they were harvested.

A slightly different protocol was used for soaking with 5-amino-2,4,6-triiodoisophthalic acid (I3C) Phasing Kit (Hampton Research) and Tantalum Cluster Derivatization Kit (Jena Bioscience). 1 M I3C stock was made in 2 M Lithium hydroxide so as to fully deprotonate the I3C carboxyl groups. This was diluted to 0.5 M I3C with the crystallisation buffer and added to the crystallisation drop as appropriate. Final I3C concentrations tested varied between 10 to 250 mM. 10 mM I3C was used for testing crystal diffraction. The Tantalum Bromide Cluster was either added directly to the crystallisation drop as powder or diluted to 2 mM.
with the crystallisation solution and crystals were transferred to the solution and soaked for 2 h. The latter method was used for soaking crystals whose diffraction was tested.

All crystals soaked in heavy atoms were either back-soaked to remove unspecifically bound heavy atoms or harvested directly. Back soaking was carried out by briefly transferring a crystal to a 1 μl drop containing the reservoir solution prior to freezing.

4.19.3 SeMet crystal oxidation and reduction

Oxidation and reduction of SeMet was achieved by soaking crystals for 1 h in the reservoir solution containing additional 0.1% hydrogen peroxide (H₂O₂) or 2 mM TCEP respectively. Crystals were then flash frozen in liquid nitrogen.

4.19.4 Heavy atom derivative data collection

SeMet and heavy-atom derivative SAD datasets were collected at the peak wavelength experimentally derived from fluorescence scans. For MAD experiments, additional datasets were collected at the inflection and lower remote wavelength, also obtained from the fluorescence scan. The same crystal, crystal orientation and strategy as the peak dataset were used to collect inflection and remote datasets. This served to minimize systematic errors that arise from differences within a crystal lattice and between crystals and thus more accurately measure anomalous signal.

4.19.5 Derivative data processing and phasing

Molecular replacement was carried out using Molrep (CCP4), MR Rosetta and Phaser MR (Phenix) programs [164,165]. Experimental phasing was carried out using AutoSol (Phenix), Crank and Shelx C/D/E, using the AutoSharp pipeline, (CCP4) [164,165,167].
4.20 Model building and refinement

Structure refinement was carried out using Phenix Refine and Feature Enhanced Maps (FEM) (Phenix) [164]. Model building was carried out manually using Coot [169].

4.21 Bioinformatics programs

Multiple sequence alignment was carried out using Tree based Consistency Objective Function For Alignment Evaluation (T-coffee) [170]. Secondary structure prediction was carried out using the Jpred3 server [171]. Disorder prediction was carried out using Regional Order Neural Network (RONN) [172]. Surface Entropy Reduction prediction (SERp) server was used to identify flexible or solvent exposed residues of PERK luminal domain to be replaced with residues with lower conformational entropy in an attempt to increase the efficiency of crystallisation [173]. The Phyre server was used to generate PERK luminal domain models to be used as molecular replacement ensembles [174]. DALI server was used to search for known fold motifs based on PERK luminal structure [175].
5 PERK, IRE1 and BiP purification
5.1 Summary

In this chapter, the cloning, expression and purification of wild-type PERK and IRE1 ER stress sensing luminal domains, BiP chaperone and C\textsubscript{H}1 unfolded protein is presented. Purification of PERK derivatives and mutant proteins is discussed in Chapters 0 and 7 as appropriate. All proteins used are summarised in Table 5-1. Hereon, \textit{H. sapiens}, \textit{M. musculus} and \textit{D. melanogaster} orthologues will be identified by ‘h’, ‘m’ and ‘d’ prefixes respectively. The ‘p’ suffix refers to \textit{S. cerevisiae} proteins as previously used.

For crystallisation studies, a stable luminal domain core region corresponding to residues 105-403 of hPERK, 101-399 of mPERK and 182–384 of dPERK was identified. Proteins were expressed in \textit{E. coli}, as for all other proteins used in this work, and dimeric oligomers were purified to near homogeneity. For biophysical characterisation experiments, hPERK and hIRE1 luminal domains were divided into five sub-regions based on boundaries used in previous studies. Full-length luminal domain construct (regions I-V) and truncated proteins (containing regions I-IV, II-V, II-IV) were purified.

A protocol was established for the purification of BiP chaperone to overcome initial problems of unspecific DNA contamination that rendered it prone to aggregation. Full-length BiP as well as its NBD and SBD in isolation were purified. The immunoglobulin constant heavy chain domain C\textsubscript{H}1 was purified from inclusion bodies following a published protocol.
<table>
<thead>
<tr>
<th>Name</th>
<th>Species</th>
<th>N-</th>
<th>C-</th>
<th>Description</th>
<th>MW</th>
</tr>
</thead>
<tbody>
<tr>
<td>hPERK 41-510</td>
<td>H.s.</td>
<td>41</td>
<td>510</td>
<td>Full-length luminal domain</td>
<td>52.1</td>
</tr>
<tr>
<td>hPERK (I-V); hPERK 54-510</td>
<td>H.s.</td>
<td>54</td>
<td>510</td>
<td>Regions I-V</td>
<td>50.9</td>
</tr>
<tr>
<td>hPERK (I-IV); hPERK 54-403</td>
<td>H.s.</td>
<td>54</td>
<td>403</td>
<td>Regions I-IV</td>
<td>38.6</td>
</tr>
<tr>
<td>hPERK (II-V); hPERK 105-510</td>
<td>H.s.</td>
<td>105</td>
<td>510</td>
<td>Regions II-V</td>
<td>45.9</td>
</tr>
<tr>
<td>hPERK (II-IV); hPERK 105-403</td>
<td>H.s.</td>
<td>105</td>
<td>403</td>
<td>Regions II-V; core luminal domain</td>
<td>33.5</td>
</tr>
<tr>
<td>hPERK (II-IV) L200P</td>
<td>H.s.</td>
<td>105</td>
<td>403</td>
<td>Dimerisation-defective</td>
<td>33.5</td>
</tr>
<tr>
<td>hPERK (II-IV) Δα2</td>
<td>H.s.</td>
<td>105</td>
<td>384</td>
<td>Tetramerisation-defective</td>
<td>31.3</td>
</tr>
<tr>
<td>hPERK (II-IV) W165A</td>
<td>H.s.</td>
<td>105</td>
<td>403</td>
<td>Tetramerisation-defective</td>
<td>33.5</td>
</tr>
<tr>
<td>hPERK (II-IV) L388N</td>
<td>H.s.</td>
<td>105</td>
<td>403</td>
<td>Tetramerisation-defective</td>
<td>33.5</td>
</tr>
<tr>
<td>hPERK (II-IV) W165A_L388N</td>
<td>H.s.</td>
<td>105</td>
<td>403</td>
<td>Tetramerisation-defective</td>
<td>33.5</td>
</tr>
<tr>
<td>hPERK (II-IV) E230A_E232A_E233A</td>
<td>H.s.</td>
<td>105</td>
<td>403</td>
<td>Surface-entropy reduction</td>
<td>33.5</td>
</tr>
<tr>
<td>mPERK (II-IV); mPERK 101-399</td>
<td>M.m.</td>
<td>101</td>
<td>399</td>
<td>Regions II-V; core luminal domain</td>
<td>33.3</td>
</tr>
<tr>
<td>dPERK (II-IV); dPERK 182-384</td>
<td>D.m.</td>
<td>182</td>
<td>384</td>
<td>Regions II-V; core luminal domain</td>
<td>34.8</td>
</tr>
<tr>
<td>hiRE1 (I-V)</td>
<td>H.s.</td>
<td>24</td>
<td>440</td>
<td>Regions I-V; full-length luminal domain</td>
<td>46.4</td>
</tr>
<tr>
<td>hiRE1 (I-IV)</td>
<td>H.s.</td>
<td>24</td>
<td>390</td>
<td>Regions I-IV</td>
<td>40.9</td>
</tr>
<tr>
<td>hiRE1 (II-V)</td>
<td>H.s.</td>
<td>32</td>
<td>440</td>
<td>Regions II-V</td>
<td>45.5</td>
</tr>
<tr>
<td>hiRE1 (II-IV)</td>
<td>H.s.</td>
<td>32</td>
<td>390</td>
<td>Regions II-V; core luminal domain</td>
<td>40.1</td>
</tr>
<tr>
<td>BiP FL</td>
<td>H.s.</td>
<td>28</td>
<td>654</td>
<td>Full-length</td>
<td>69.3</td>
</tr>
<tr>
<td>BiP NBD</td>
<td>H.s.</td>
<td>28</td>
<td>405</td>
<td>NBD</td>
<td>41.7</td>
</tr>
<tr>
<td>BiP SBD</td>
<td>H.s.</td>
<td>422</td>
<td>654</td>
<td>SBD</td>
<td>25.9</td>
</tr>
<tr>
<td>C$_{41}$</td>
<td>H.s.</td>
<td>1</td>
<td>88</td>
<td>-</td>
<td>10.3</td>
</tr>
</tbody>
</table>

**Table 5-1. PERK, IRE1, BiP and C$_{41}$ protein constructs**

The constructs of PERK, IRE1, BiP and C$_{41}$ proteins used in this study are summarised. The nomenclature used throughout this report as well as a brief description of each construct are shown. H.s., M.m. and D.m. abbreviations are used to refer to *H. sapiens*, *M. musculus* and *D. melanogaster* species respectively. N- and C- refer to the 5’ N-terminal starting residue and 3’ C-terminal ending residue of the construct respectively. The given MW (kDa) corresponds to the size of untagged species; tagged proteins have an additional 3 kDa mass due to the His$_6$-tag and PreScission Protease cleavage site.
5.2 Purification of PERK luminal domain for crystallisation

5.2.1 Identification of PERK luminal domain core region

The luminal domain of hPERK, without its signal sequence, spans residues Ala41 to Ile510. PERK luminal domain has limited sequence identity to hIRE1 luminal domain (18% and below depending on the species). However, secondary structure prediction of hPERK luminal domain suggests that their overall secondary structure composition is similar. Similarly to hIRE1 and IRE1p, hPERK luminal domain is suggested to be composed mainly of β-strands interspaced by several α-helices. Multiple sequence alignment, PERK luminal domain secondary structure prediction and the known structures of the hIRE1 (PDB: 2HZ6) and IRE1p (PDB: 2BE1) luminal domain orthologues were used to guide PERK luminal domain construct design for crystallisation studies (Appendix Figure 10-3) [95,101]. Figure 5-1 summarises hPERK luminal domain constructs expressed and purified for X-ray crystallography studies.
PERK and IRE1 are composed of a N-terminal ER luminal domain, a membrane-spanning region and cytosolic C-terminal domain. The constructs of hPERK luminal domain (purple) that were purified for structural characterisation and discussed in Chapter 6 are illustrated. Initially all protein were purified with a N-terminal His$_6$-tag (magenta). A 10 kDa fragment spontaneously degraded from the C-terminus of hPERK 41-510 and 54-510. hPERK 54-403, which does not contain the cleavable fragment, was stable but eluted as heterogeneous dimers. hPERK 105-403 was successfully purified as a stable, clean, homogeneous and monodispersed sample. Crystals were obtained only when using the untagged hPERK 105-403. hIRE1 luminal domain (blue) (PDB: 2HZ6) was used to guide hPERK construct design [95]. The positions of residues of particular interest are shown. Red lines indicate flexible or unstructured regions.
The hIRE1 luminal domain construct used in published crystallisation studies spans residues Ser24 to Val390 [101]. This corresponds to residues Glu95 to Ile510 of hPERK. Secondary structure prediction of hPERK luminal domain suggests there is an extended α-helix from residues Tyr54 to Glu76. This region is not conserved in hIRE1 and was included in initial hPERK luminal domain constructs as it could provide additional structural, and hence functional, insights. The initial hPERK luminal domain constructs designed comprised residues Ala41 to Ile510 (hPERK 41-510) and Ala54 to Ile510 (hPERK 54-510).

Recombinant proteins were expressed in E. coli as N-terminal His₆-tagged fusion proteins and initially purified by TALON affinity. SDS-PAGE analysis of the elution fractions revealed the presence of proteins of the expected size (Figure 5-2A). Two additional bands of lower MW suggestive of protein degradation were also present. The larger ∼70 kDa band, which is removed in subsequent purification steps, is probably a contaminant E. coli protein. α-His₆ western blotting was carried out on the TALON elution fractions. Both samples contained two His₆-containing species: (i) the intact 58 kDa His₆-hPERK 41-510 and 56 kDa His₆-hPERK 54-510, and (ii) a species ∼15 kDa smaller (Figure 5-2B). Ion-exchange and SEC were used in an attempt to separate the two proteins in His₆-hPERK 41-510 samples (Figure 5-2C-D). The 58 kDa intact protein and the 43 kDa degradation product co-elute a single homogenous species during both ion-exchange and SEC. This indicates the two proteins have similar chemical and physical properties, and possibly interact. Taken together these results suggest the smaller band is a C-terminally cleaved fragment of the expressed hPERK 41-510 and 54-510 luminal domain proteins.
Figure 5-2. C-terminal cleavage of full-length hPERK luminal domain

(A) hPERK 41-510 and 54-510 TALON affinity elution fractions analysed by SDS-PAGE. The -70 kDa band is an E. coli contaminant; the -58 and -56 kDa products are His₆-hPERK 41-510 and 54-510 respectively; all the other bands of lower MW are His₆-hPERK degradation products.

(B) α-His₆ western blot of samples in ‘A’. In each sample, two His₆-containing species are detected of -56-58 kDa (intact His₆-hPERK 41-510 and 54-510 respectively) and His₆-tagged species -15kDa smaller.

(C) His₆-hPERK 41-510 SEC (HiLoad 16/60 Superdex 200). His₆-hPERK 41-510 elutes a single a single homogenous peak at 58 ml.

(D) Peak fractions (52 – 64 ml elution) were analysed by SDS-PAGE; ‘ml’ numbers refer to starting volume of the fraction collected. The 58 kDa (intact His₆-hPERK 41-510) and the 43 kDa species co-elute during SEC.
The ~43 kDa bands were excised from the SDS-PAGE gel and analysed by in gel trypsin digestion and MALDI-TOF. The resulting peptides confirmed the identity of the bands as truncated fragments of hPERK 41-510 and 61-510 (Figure 5-3A). The most C-terminal hPERK residue identified was Arg392. Trypsin cuts at the C-terminus of Arginine and Lysine residues, unless they are adjacent to a Proline. Therefore, the cleaved species must terminate between Arg392 and Arg402, the next trypsin cleavage site downstream of Arg392. This region corresponds to β18 of hIRE1 luminal domain, its penultimate secondary structure. β19 of hIRE1 is a small three-residue feature located after a highly flexible region spanning over 50 residues for which no electron density was observed in the X-ray diffraction data used to solve the structure [101]. hPERK Ile403, which corresponds to the terminal residue of hIRE1 β18, was thus chosen as the optimized C-terminal residue for subsequent constructs.

hPERK 54-403 was expressed and purified by TALON affinity, ion-exchange and SEC. Although this new construct was stable and did not degrade, it eluted as two peaks during SEC MALS indicates that both species have an identical MW of 79.3 kDa, which correspond to dimers of hPERK 54-403 (Figure 5-3B). This suggests that hPERK 54-403 forms two distinct dimer conformations. Due to its heterogeneity, this sample was deemed unsuitable for crystallisation studies.

The N-terminal region of hPERK (Ala41 to Leu105) is predicted to be highly flexible and disordered based on bioinformatics analysis (Appendix Figure 10-4). Furthermore, this ~100 residue region is not conserved in IRE1 sequences. Leu105 of hPERK actually corresponds to Leu32 of hIRE1, its first ordered residue [101]. Leu105 was chosen as a new N-terminal residue for the optimized hPERK luminal domain construct to be used in crystallisation studies. The optimized hPERK construct thus spans Leu105 to Ile403. Identical boundaries were used for cloning of mPERK (Leu101 to Val399) and dPERK (Leu182 to Leu384) luminal domains for crystallisation studies.

hPERK 105-403, mPERK 101-399 and dPERK 182-384 optimized constructs represent the core region of PERK luminal domain. Since this core luminal domain region corresponds sub-regions II-IV, as described in Chapter 7, hereon hPERK 105-403, mPERK 101-399 and dPERK 182-384 will be referred to as hPERK (II-IV), mPERK (II-IV) and dPERK (II-IV) respectively for consistency.
Figure 5-3. Identification of hPERK luminal domain core region

(A) Results from in gel trypsin digestion and MALDI-TOF analysis of the 43 kDa cleaved species from His<sub>6</sub>-hPERK 41-510 and 54-510 purifications. The matched peptides are highlighted in cyan. Trypsin cleavage sites are shown in bold. The chosen Ile403 optimized C-terminal residue is shown in red. hIRE1 β18 is aligned to hPERK (PDB: 2HZ6) [95]. Numbers at the top refer to hPERK sequence.

(B) hPERK 54-403 and hPERK 105-403 SEC-MALS analysis (Superdex 200 PC 3.2/30). hPERK 54-403 (purple) elutes as two separate dimer species of identical 79.3 kDa size. hPERK 105-403 (blue) elutes as a single dimer peak of 67.2 kDa.
5.2.2 Purification of *H. sapiens*, *M. musculus* and *D. melanogaster* orthologues of PERK (II-IV)

Proteins were expressed as described in the materials and methods section. Initially, the purification protocol was optimized for hPERK (II-IV). A large proportion of *E. coli* protein contaminants were removed in the flow through and wash steps of TALON affinity purification (Figure 5-3A). The major species of the TALON elution fraction was His$_6$-hPERK (II-IV), which migrates with an apparent MW of 40 kDa. The elution fraction was further purified by ion-exchange (Figure 5-3B). The peak containing hPERK (II-IV), eluting from 400 to 500 mM NaCl concentration, was analysed by SDS-PAGE (Figure 5-3C). Fractions containing purest hPERK (II-IV) (420 – 490 mM NaCl) were pooled and run on a SEC column.

Initially, hPERK (II-IV) was purified in the absence of reducing agent. However the SEC elution profile indicates hPERK (II-IV) elutes as multiple overlapping peaks (Figure 5-5A). The heterogeneity of this sample renders it unsuitable for crystallisation studies. Addition of 2 mM TCEP reducing agent shifted the profile to a single homogeneous peak eluting at 70 ml on a HiLoad 16/60 Superdex 200 column. Fractions eluting from 68 to 76 ml contained pure hPERK (II-IV) protein and were pooled (Figure 5-5B). 2 mM TCEP was used in all purification steps after the initial TALON affinity step to prevent unspecific disulphide bond formation. TCEP was chosen as the reducing agent as it is less easily oxidized and is considered to break disulphide bonds irreversibly. 10% glycerol was included in all buffers as it increases protein solubility, helps prevent aggregation and assists in controlled nucleation during crystallisation. Identical TALON affinity, ion-exchange and SEC purification steps were carried out for mPERK (II-IV) and dPERK (II-IV).
Figure 5-4. hPERK (II-IV) TALON affinity and ion-exchange chromatography

(A) hPERK (II-IV) TALON affinity purification. The flow-through (FT), first and last 5 mM imidazole washes (W1 and W2) and elution (E) fractions were analysed by SDS-PAGE. hPERK (II-IV) is soluble, stable and migrates as ~40 kDa species.

(B) hPERK (II-IV) ion-exchange chromatography (HiTrap Q HP). hPERK (II-IV) elutes as a single peak at 430 mM NaCl.

(C) Fractions eluting from 400 to 480 mM NaCl (34 – 50 ml elution volume) were analysed by SDS-PAGE. ‘ml’ numbers refer to starting volume of the fraction collected. Fractions containing purest hPERK (II-IV) (38 – 50 ml elution volume corresponding to 420 – 490 mM NaCl) were pooled.
**Figure 5-5. hPERK (II-IV) SEC**

(A) hPERK (II-IV) SEC (HiLoad 16/60 Superdex 200). In the absence of reducing agent, hPERK (II-IV) elutes as multiple overlapping peaks from 45 to 80 ml. Upon addition of 2 mM TCEP, hPERK (II-IV) elutes as a single species at 70 ml.

(B) hPERK (II-IV) (+2 mM TCEP) fractions from 68 to 76 ml were analysed on SDS-PAGE. ‘ml’ numbers refer to starting volume of the fraction collected. Fractions containing pure hPERK (II-IV) (68 – 76 ml) were pooled.
Protein samples were concentrated and used in subsequent biochemistry and crystallisation studies (Figure 5-6A). Although protein expression levels were relatively low (2 mg of protein per 10 L of cells) the final samples were pure, stable, homogenous and monodispersed dimers and therefore optimal for crystallisation studies. MALS analyses indicate that hPERK (II-IV), mPERK (II-IV) and dPERK (II-IV) have a molar mass of 66.3, 67.2 and 80.0 kDa respectively (Figure 5-6B). This corresponds to dimeric oligomeric states. No crystal hits were obtained for dPERK (II-IV); since this protein was not used in any subsequent experiments it will no longer be discussed in this thesis.

All PERK SeMet derivatives and mutant proteins were purified identically as for the wild type full-length hPERK (II-IV) protein.
Figure 5-6. Purification of PERK (II-IV) core luminal domain

(A) PERK (II-IV) luminal domain purified samples. SDS-PAGE gel of the purified and concentrated *H. sapiens*, *M. musculus* and *D. melanogaster* orthologues of PERK (II-IV) used for crystallisation.

(B) SEC MALS analyses (Superdex 200 PC 3.2/30) of PERK (II-IV) proteins. hPERK (II-IV) (blue), mPERK (II-IV) (purple) and dPERK (II-IV) (green) elute as single monodispersed dimeric species of 67.2, 66.3 and 80.0 kDa respectively.
5.3 Purification of hPERK and hIRE1 luminal domains for biophysical studies

5.3.1 Designation of hPERK and hIRE1 luminal domain regions I-V

To predict the boundaries of hPERK and hIRE1 luminal domains’ sub-regions, knowledge from previous studies and bioinformatics analyses were utilized. Kimata et al. first suggested that IRE1p luminal domain can be divided in five functional sub-regions designated I to V [85]. In their study, an 11 amino acid scanning deletion experiment was carried out and the truncated proteins were tested for their ability to splice Hac1p mRNA through a UPRE promoter sequence-LacZ reporter assay. IRE1p luminal domain shares highest sequence similarity with that of hIRE1. Initially, IRE1p and hIRE1 sequence alignment was used as an indication of the approximate location of sub-regions I to V in hIRE1 luminal domain. The known structural features of hIRE1 and IRE1p luminal domains were then used to modify the domain boundaries slightly [95,101]. In particular, care was taken to not truncate proteins in the middle of a secondary structure element and to not disrupt structural features such as stacked β-sheet bundles, hydrogen bonds or salt bridges, all which could affect the protein’s function if disturbed. Multiple sequence alignment and secondary structure prediction were carried out to extend hIRE1 luminal domain sub-region boundaries to hPERK luminal domain. The chosen boundaries for hPERK and hIRE1 regions I to V used for biochemistry experiments are summarised in Table 5-2.
Table 5-2. hPERK and hIRE1 sub-regions I to V boundaries

The boundaries of IRE1p, hIRE1 and hPERK luminal domain regions I, II, III, IV and V are shown. Residues for IRE1p were taken from Kimata et al. 2006 [85]. The hIRE1* column (grey) refers to residues designated solely based on IRE1p and hIRE1 sequence alignment; these were not used for constructs of expressed proteins. Rather, optimized hIRE1 (blue) and hPERK (purple) domain boundaries were used.

<table>
<thead>
<tr>
<th>Sub-region</th>
<th>IRE1p</th>
<th>hIRE1*</th>
<th>hIRE1</th>
<th>PERK</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>25-104</td>
<td>1-27</td>
<td>24-31</td>
<td>41-104</td>
</tr>
<tr>
<td>II</td>
<td>105-235</td>
<td>28-127</td>
<td>32-129</td>
<td>105-204</td>
</tr>
<tr>
<td>III</td>
<td>236-266</td>
<td>128-143</td>
<td>130-153</td>
<td>205-243</td>
</tr>
<tr>
<td>IV</td>
<td>266-477</td>
<td>144-379</td>
<td>154-390</td>
<td>244-403</td>
</tr>
<tr>
<td>V</td>
<td>448-517</td>
<td>380-443</td>
<td>391-440</td>
<td>404-510</td>
</tr>
</tbody>
</table>
5.3.2 Purification of hPERK and hIRE1 luminal domain regions I-V fragments

The following hPERK and hIRE1 luminal domain constructs were used to assess the functionality of regions I and V: I-V, I-IV, II-V and II-IV. All proteins were purified identically to hPERK (II-IV) (Chapter 5.2.2).

Regions I and III are very short segments compared to the other sub-domains. Additionally, as seen from the structure of hIRE1, they are unstructured loop regions [101]. Region V was not present in the hIRE1 crystallised fragment. From the evidence provided in Chapter 5.2.1, as well as disorder prediction results, region V is likely to be highly flexible. Regions I, III and V were therefore not purified as isolated sub-domains. Regions II and IV alone were cloned to further characterise the ER stress sensing luminal domain. However attempts at expressing these proteins were unsuccessful and will not be further discussed in this thesis.

5.4 Purification of BiP chaperone

The human orthologue of full-length (FL) BiP spans residues Gly28 to Leu654. BiP FL was expressed and purified by an initial TALON affinity step followed by SEC (Figure 5-7). The ion-exchange step was omitted since protein levels obtained were significantly more abundant than any *E. coli* contaminant, even just after only the TALON affinity step. Some BiP FL protein was lost in the 5 mM imidazole wash step. However since most of the *E. coli* protein contaminants were also removed in this step and sufficient protein was recovered in the elution fraction this step was included. The BiP FL that is lost on the wash may be unfolded and as a result does not bind the column. Alternatively the TALON resin was saturated with the high quantity of BiP protein. The TALON resin elution fraction were pooled and run on a Hi Load 16/60 Superdex 200 column. BiP FL initially eluted in the void volume during SEC (Figure 5-7C). On this column this is indicative of a protein species larger than 700 kDa and thus suggests the BiP FL protein was aggregating. The absorbance of the purified BiP FL sample was measured. Samples had an absorbance at 260 nm (A260) to absorbance at 280 nm (A280) ratio above 1.8. Proteins absorb almost exclusively at 280 nm mainly due to their Tryptophan, Tyrosine and Cysteine residues (with their molar absorption coefficients
decreasing in that order). Nucleic acids normally absorb at 260 nm due to the aromatic heterocyclic ring structures of pyrimidine and purine bases. Other factors that may cause a high reading at 260 nm include the presence of imidazole or dithiothreitol (DTT) and β-mercaptoethanol reducing agents. Since these reagents were not present in the BiP FL SEC buffer, the contaminating species was suspected to be DNA. Two identical BiP FL samples were run on a native gel and stained either with Coomassie Blue (protein-specific) or GelRed (DNA-specific) stains (Figure 5-7B). Large quantities of smeared DNA were detected confirming the cause of the high A260 to A280 ratio. DNA-contamination caused BiP FL aggregation and it was therefore essential to circumvent this in order to obtain functional protein for subsequent studies.

Firstly, the amount and incubation time with DNase during the lysis step were increased. Additionally, a 1 M NaCl wash step was added during TALON affinity purification so as to increase the ionic strength of the buffer and disrupt weak electrostatic interactions. However, neither of these changes prevented BiP FL from co-purifying with DNA. The cationic reagent PEI has been published as an effective means of precipitating protein-bound nucleic acids during early purification steps [176]. Addition of PEI during BiP FL lysis steps resulted in the elution of peaks corresponding to ∼140 and ∼70 kDa protein species during SEC, as approximated using the Hi Load 16/60 Superdex 200 calibration equation (Figure 5-7C). These MW are indicative of BiP FL dimers and monomers respectively. In summary, these results demonstrate that (i) the contaminating nucleic acid was responsible for BiP FL aggregation and (ii) that this can be overcome by the addition of PEI. Nonetheless, the effect of PEI on the functionality of proteins has never been investigated, at least in published studies, and therefore an alternative protocol to avoid BiP FL-DNA co-purification was developed.

BiP is a Hsp70 chaperone and adopts an open SBD conformation and low substrate affinity in the presence of ATP nucleotide. Following this reasoning, excess ATP was added, instead of PEI, during initial lysis and TALON affinity steps of BiP FL purification. ATP was not added in the final SEC step to ensure that BiP FL could be used in its different nucleotide-dependent functional conformations in subsequent experiments. The resulting BiP FL sample elutes as two distinct monomer and dimer peaks in SEC (Figure 5-7C-D). This SEC elution profile is similar to that obtained when BiP FL was purified using PEI. When BiP FL samples purified
either with PEI or ATP were stained with GelRed, no DNA was detected confirming their efficacy in preventing BiP FL-DNA contamination (Figure 5-7B). Addition of ATP avoids DNA contamination and BiP FL aggregation and was used to purify BiP FL protein.

The sequence of BiP, as for all Hsp70s, is highly conserved and proteins have been well characterised. BiP NBD and BiP SBD constructs were designed based on published studies and encompass residues Gly28 to Leu405 and Leu422 to Leu654 respectively. Their purification was carried out as for BiP FL protein. BiP NBD elutes as a single 40 kDa monomeric peak during SEC (Figure 5-8A). BiP SBD elutes as two peaks corresponding to 25 kDa monomeric and 50 kDa dimeric species (Figure 5-8A). All final protein samples are shown in Figure 5-8B.

5.5 Purification of C_H1

The immunoglobulin constant heavy chain domain C_H1 protein was purified as previously published from inclusion bodies [25]. C_H1 elutes in the void volume during SEC. Since C_H1 is intrinsically unfolded and only becomes ordered upon binding of its cognate binding partner C_L, this was expected. The purified sample used for subsequent studies is shown in Figure 5-8B.
Figure 5-7. BiP FL purification

(A) BiP FL TALON affinity purification. The flow-through (FT), first and last 5 mM imidazole washes (W1 and W2) and elution (E) fractions were analysed by SDS-PAGE. BiP FL is present in the elution fraction and migrates with an apparent MW of 70 kDa.

(B) DNA contamination of BiP FL. BiP FL samples purified without additives, + 0.1% PEI or + 5 mM ATP/10 mM MgCl₂ were run on a native PAGE gel and stained either with Coomassie Blue or GelRed. DNA was detected only in the sample purified without additives.

(C) BiP FL SEC (HiLoad 16/60 Superdex 200). BiP FL purified without additives elutes as a large aggregate in the void volume. BiP FL purified in the presence of 0.1% PEI or 5 mM ATP/10 mM MgCl₂ elutes as dimer and monomer oligomers. Fractions from 47 to 80 ml were analysed by SDS-PAGE. Fractions containing pure BiP FL (50 – 80 ml) were pooled.
Figure 5-8. BiP and C41 purification

(A) BiP NBD and SBD SEC (HiLoad 16/60 Superdex 200). BiP NBD elutes as a single monomeric species; BiP SBD elutes as dimer and monomer oligomers. Fractions containing pure protein (68 – 75 ml for NBD; 65 – 83 ml for SBD) were pooled.

(B) BiP and C41 purified proteins. SDS-PAGE of BiP FL, BiP NBD, BiP SBD and C41 purified proteins used in subsequent experiments.
5.6 Discussion

In this chapter the cloning, expression and purification of PERK, IRE1, BiP and C\textsubscript{H}1 proteins used in subsequent crystallisation and biochemical characterisation experiments are described.

Firstly, a stable PERK luminal domain core fragment for use in crystallisation studies was identified. Initially, two complete hPERK luminal domain constructs (41-510 and 54-510) were expressed and purified. Although these proteins were soluble they sustained specific and spontaneous cleavage of a large C-terminal fragment. This corresponded approximately to residues Arg396 to Ile510. Furthermore, the –50 N-terminal residues promoted the formation of two distinct populations of PERK dimer oligomers, which rendered the sample highly heterogeneous. These results were combined with bioinformatics analyses (PERK and IRE1 multiple sequence alignment, PERK secondary structure prediction and PERK disorder prediction) and knowledge from the published hIRE1 and IRE1\textsubscript{p} luminal domain structures to indentify a PERK luminal domain core region. This core region of \textit{H. sapiens} (105-403), \textit{M. musculus} (101-399) and \textit{D. melanogaster} (182-384) PERK orthologues was cloned expressed and purified. Although the proteins expressed in low abundance, the final samples contained pure, stable, homogenous and monodispersed PERK dimers. These samples were optimal for use in ensuing crystallisation experiments (Chapter 6).

Secondly, hPERK and hIRE1 luminal domains were subdivided into five regions according to previously published IRE1\textsubscript{p} boundaries. Constructs encompassing luminal domains regions I-V (complete luminal domain), I-IV, II-V or II-IV (core luminal domain) of hPERK and hIRE1 were cloned, expressed and successfully purified. These truncation proteins provide a means to asses the significance of these sub-regions in the biological function of hPERK and hIRE1 luminal domains \textit{in vitro} (Chapter 7).

Thirdly, a protocol for the purification of BiP chaperone purification was developed. Initially, purifications carried out using published protocols resulted in aggregation of BiP FL. Further analysis revealed this was due to co-purification of contaminating DNA species with BiP FL. Since BiP FL aggregates and BiP FL-DNA complexes are not biologically significant, protein
samples obtained following the initial protocol were unsuitable for biochemical characterisation experiments.

Initially, the cationic reagent PEI, which is widely used in cellular biology to promote attachment of weakly anchoring cells in cultures as well as a powerful DNA transfection reagent. However it has also been reported to be used as a DNA precipitant during protein purification [176]. PEI promotes the condensation of DNA into positively charged particles rendering it less prone to form electrostatic interactions with charged proteins in solution. Addition of PEI successfully avoided DNA contamination during BiP FL purification. However since the effect of PEI on recombinant proteins has not been investigated in published studies, an alternate protocol was developed to ensure the functionality of BiP FL.

BiP is a Hsp70 chaperone and in the ATP-bound state exists in an open SBD conformation that confers low affinity for its hydrophobic substrates. BiP FL purification buffers were therefore supplemented with ATP as a means of maintaining BiP in a low substrate affinity conformation to avoid the binding of DNA from the E. coli cellular extracts. Similarly to addition of PEI, ATP was also successful in preventing the unspecific binding of DNA to BiP FL. Since the latter protocol exploits a biologically relevant function of BiP it was chosen as a means of preventing co-purification of BiP FL and DNA. This was also a first indication that the purified recombinant BiP FL protein is functionally active since it can be regulated by ATP nucleotide. BiP FL was purified to near homogeneity as monomers and dimers in equimolar amounts. The two sub-domains of BiP, the NBD and SBD, were also purified in isolation. These respectively contained uniquely monomers, or a mixture of monomeric and dimeric species, with the equilibrium shifted towards dimers.

Finally, the intrinsically unfolded protein C_H1 was successfully purified from inclusion bodies following the previously published protocol.
6  X-ray crystal structure of PERK luminal domain
6.1 Summary

This chapter describes the crystallisation and structure determination of PERK luminal domain *H. sapiens* and *M. musculus* orthologues, solved at 3.1 Å and 3.3 Å resolution respectively.

Crystallisation of hPERK (II-IV) and mPERK (II-IV) proteins was optimized and diffraction was improved at best from 10 to 3.1 Å for the former. Phasing of hPERK (II-IV) diffraction data was achieved by heavy atom derivatization and MAD. The refined structure of hPERK (II-IV) was used to phase mPERK (II-IV) diffraction data by molecular replacement.

Overall the structures of hPERK (II-IV) and mPERK (II-IV) are very similar to one another and to those of *H. sapiens* and *S. cerevisiae* IRE1. Dimerisation occurs mainly by antiparallel stacking of a central β-strand between two monomers (Interface 1). This is identical to the dimerisation interface of IRE1 luminal domains. Strikingly, hPERK (II-IV) packed as tetramers in the crystal lattice. This is achieved by the binding of two extended C-terminal α-helices of a hPERK (II-IV) dimer with a hydrophobic cleft on an interacting hPERK (II-IV) dimer (Interface 2). This conformation of the ER stress sensors has never been reported in published studies. Mutant proteins were used to probe the biological relevance of this assembly in solution by disrupting Interface 2. Indeed, tetramerisation was highly compromised by selected mutations. Since the tetramer conformation is also observed in solution, it is likely to exist *in vivo* and as such has a functional role in the activity of PERK.

On the other hand, mPERK (II-IV) only formed dimers. No electron density for the entire region that mediated tetramerization in hPERK (II-IV) was observed suggesting this region was not structured. The differences between the oligomeric states of the luminal domains can be reconciled by representing differently activated states. A novel model for oligomerisation-dependent UPR activation is proposed.
6.2 Introduction

To date, only the structures of *H. sapiens* IRE1 (PDB: 2HZ6) and *S. cerevisiae* IRE1p (PDB: 2BE1) luminal domains have been solved [95,101]. These date back to almost a decade ago.

Up to 2005, all the biochemical and cellular studies implicated BiP chaperone as a major regulator of PERK, IRE1 and ATF6 signalling. However, based on the X-ray crystal structure of IRE1p luminal domain, a mechanism for ER stress sensing which is completely independent of BiP chaperone was proposed. This involves the direct binding of unfolded proteins to an MHC-like crevice of the luminal domain, which allows for oligomerisation/clustering and signalling activation. The X-ray crystal structure of hIRE1 luminal domain closely resembles to that of IRE1p. However *(i)* the MHC-like groove of hIRE1 is too narrow to accommodate peptide binding and *(ii)* conserved residues speculated to be essential for peptide binding in IRE1p are either not conserved or buried in hIRE1’s structure. The structure of hIRE1 points away from a direct unfolded protein binding model and backs up a BiP-dependent pathway.

Although the luminal domains of hPERK and hIRE1 share only 18% sequence identity, a multitude of biochemical studies have demonstrated that they are similarly activated by ER stress [76,103]. Moreover the two are functionally interchangeable to mediate UPR signalling by their cytoplasmic domains *in vivo*. Given their functional conservation, at least with respect to ER stress sensing, it is reasonable to assume they have similar mechanisms for detecting the presence of unfolded proteins and as a result allowing phosphorylation and activation of their cytoplasmic effector domains.
6.3 Crystallisation of PERK (II-IV) proteins

The final optimized crystallisation conditions of hPERK (II-IV) and mPERK (II-IV) proteins are discussed in this section and summarised in Table 6-1.

6.3.1 hPERK (II-IV) crystallisation

6.3.1.1 Optimization of crystallisation buffer

hPERK (II-IV) protein is highly soluble, even at concentrations up to 40 mg/ml. Initial nanolitre scale screening crystallisation trays were set up at 25 mg/ml protein. Showers of bipyramidal microcrystals appeared overnight in untagged hPERK (II-IV) drops equilibrated over 0.1 M Tris-HCl (pH 8.5), 0.2 M MgCl₂, 25% w/v PEG3350 (Figure 6-1A). An initial optimization screen around the hit condition was set up and screened conditions from 15 - 35% w/v PEG3350, 0 - 0.4 M MgCl₂ and 3 - 25 mg/ml protein concentration. Conditions containing at least 18% w/v PEG3350 and 5 mg/ml hPERK (II-IV) were necessary for the formation of microcrystals and therefore spontaneous nucleation. These were identified as the lower limits of the supersaturated labile region of crystallisation. Although nucleation was reduced, it was still too high to sustain growth of crystals larger than 5 µm. This is probably because most of the protein was depleted to form the high quantity (over 100) of initial microcrystals causing the protein concentration to quickly fall below the metastable zone into under-saturated conditions.

The following variables were changed to promote the growth of large single crystals: pH, incubation temperature, reservoir volume and the composition of the crystallisation buffer. The pH changes the charge of proteins, which in turns affects electrostatic and polar interactions between macromolecules in solution and could help control nucleation and formation of protein-protein contacts and thus the crystal lattice. The pH was varied between 7 - 9, the buffering range of Tris-HCl.
Table 6-1. PERK luminal domain crystallisation conditions

The optimized conditions used to obtain diffracting crystals of *H. sapiens* and *M. musculus* PERK (II-IV) are summarised. The components of the crystallisation buffer, protein concentration (mg/ml) protein (P) to reservoir (R) ratio used to set up vapour diffusion drops is shown. Additional ‘treatments’ carried out such as seeding and stepwise dehydration are indicated where relevant.

<table>
<thead>
<tr>
<th>Condition</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Protein</strong></td>
<td>hPERK (II-IV)</td>
<td>hPERK (II-IV)</td>
<td>mPERK (II-IV)</td>
<td>mPERK (II-IV)</td>
</tr>
<tr>
<td><strong>Crystallization buffer</strong></td>
<td>0.1M Tris (pH 8.5), 0.2M MgCl₂, 25% w/v PEG3350, 7% glycerol</td>
<td>20% v/v glycerol, 26% w/v PEG1500</td>
<td>0.1M Tris (pH 8.5), 0.2M LiS, 47.5% w/v PEG400</td>
<td>0.1M MES/imidazole (pH 6.5), 0.09 M NPS mix, 12.5% w/v PEG1000, 12.5% w/v PEG3350, 20% v/v MPD</td>
</tr>
<tr>
<td><strong>[protein] mg/ml</strong></td>
<td>5</td>
<td>25</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td><strong>P : R ratio</strong></td>
<td>1:1</td>
<td>2:1</td>
<td>2:1</td>
<td>2:1</td>
</tr>
<tr>
<td><strong>Dehydration</strong></td>
<td>Stepwise up to 40% w/v PEG3350</td>
<td>none</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td><strong>Seeding</strong></td>
<td>Reiterative microseeding</td>
<td>+ microseeds from cond. A</td>
<td>none</td>
<td>none</td>
</tr>
</tbody>
</table>
Lowering the incubation temperature slows down chemical processes and diffusion of molecules whereas increasing the reservoir volume decreases the rate of equilibration with the crystallisation drop. Although in theory these variables should control rates of nucleation in this instance they had no effect on the size or quantity of crystals obtained. Different PEG chain lengths, buffer types, and commercially available additives were tested but these variables all inhibited crystallisation. Finally, glycerol increases the solubility of proteins. By causing a shift in a protein’s solubility curve, glycerol also affects the position of the labile region of nucleation in the phase diagram [177]. This reduces spontaneous nucleation occurring at a given protein concentration. 7% glycerol resulted in fewer larger bipyramidal crystals approximately 10 µm in length (Figure 6-1B).

6.3.1.2 Reiterative microseeding

Microseeding was used to introduce nucleation sites into metastable zone conditions, which should support crystal growth but not spontaneous nucleation. Addition of microseeds into 17% w/v PEG3350, 0.1 M Tris-HCl (pH 8.5), 0.2 M MgCl₂ did not lead to crystal growth. However, microseeding into conditions of higher PEG concentrations led to formation of few crystals approximately 50 µm in all dimensions overnight (Figure 6-1C). Although this result was unexpected it suggests that the metastable zone is small and 17% w/v PEG3350 conditions actually already represent the clear undersaturated zone. The presence of seeds, thus nucleation sites, in labile zone conditions reduced the spontaneous nucleation of hPERK (II-IV). hPERK (II-IV) favours the formation of contacts with the already established lattice (from the seeds) rather than with protein in solution to form new nucleation sites. Crystals were grown by mixing 1 µl of untagged hPERK (II-IV) at 5 mg/ml plus 1 µl of fresh seed (diluted 10⁶-fold) equilibrated over 1 ml of 0.1 M Tris-HCl (pH 8.5), 0.2 M MgCl₂, 25% w/v PEG3350 and 7% glycerol in 24-well hanging drop plates at 18 °C (Condition A). 10 rounds of seeding yielded crystals that grew up to 300 µm in length after 1 week (Figure 6-1D). Identical conditions were used for crystallisation of SeMet hPERK (II-IV).
Figure 6-1. Optimization of hPERK (II-IV) crystals (Condition A)
(A) Initial hit containing showers of hPERK (II-IV) microcrystals grown in 0.1 M Tris-HCl (pH 8.5), 0.2 M MgCl₂, 25% w/v PEG3350 at 18 °C.
(B) Crystals were grown as ‘A’ plus 7% glycerol. Nucleation was reduced and crystals grew to a maximum size of 10 μM in length.
(C) Crystals from ‘B’ were used for microseeding. A few large crystals (50 μM in length) appeared overnight.
(D) 10 rounds of re-iterative microseeding seeding, as in ‘C’, yielded large single crystals (300 μM in length) after a week.
6.3.1.3  hPERK (II-IV) microseed matrix screening

To obtain additional crystallisation conditions, microseed matrix screening was employed as a means of introducing nucleation sites in conditions where crystal growth is supported but spontaneous nucleation does not occur. hPERK (II-IV) Condition A crystals were used to produce microseeds, as previously, added to hPERK (II-IV) at 20 mg/ml and used to screen for new crystal hits. One new crystal form was found. Thin needles approximately 3 μm in length and less than 1 μm in width were observed in a drop equilibrated over 20% v/v glycerol and 24% w/v PEG1500 (Figure 6-2A). These crystals were observed a year after the drops were initially set up. It was speculated that they were a cleavage product of hPERK (II-IV) that appeared over time due to spontaneous specific degradation of the protein. The crystals were washed in the reservoir solution and analysed by SDS-PAGE to check the identity of the protein crystallised. The protein appeared intact and identical to the purified hPERK (II-IV) (Figure 6-2B). It is possible that these crystals were so small they were missed during initial screening. However, since these appeared only in the corner of the drop containing a 2:1 protein to reservoir ratio (the 1:1 ratio drop remained clear) it suggested that higher protein concentration may be key to support their growth. In such a scenario, the crystals may have only grown after extensive equilibration and evaporation of water from the drop. All subsequent experiments for this condition were set up with 25 mg/ml of protein and a 2:1 protein to reservoir volume ratio. Three-dimensional crystals measuring 50 μm in length were observed after 3 days in drops seeded with Condition A microseeds and equilibrated over 1 ml reservoir of 20% v/v glycerol and 26% w/v PEG1500 in 24-well hanging-drop plates at 18 °C (Condition B) (Figure 6-2C). Because crystal formation was dependent on the presence of microseed from the original hPERK (II-IV) crystals grown in condition A it suggests that Condition B does not support spontaneous nucleation of hPERK (II-IV).
Figure 6.2. hPERK (II-IV) microseeding matrix screen (Condition B)

(A) Thin hPERK (II-IV) protein needles grown in drops containing 20 mg/ml protein, 20% v/v glycerol and 24% w/v PEG1500 (1:1 protein to reservoir ratio) at 18 °C plus microseeds from hPERK (II-IV) condition A crystals.

(B) Crystals from ‘A’ were washed and analysed by SDS-PAGE. Crystals contained pure and undegraded hPERK (II-IV) protein.

(C) Optimized hPERK (II-IV) crystals grown as in ‘A’ with the following modifications: 25 mg/ml protein, 25% w/v PEG1500, 2:1 protein to reservoir ratio.
6.3.2 mPERK (II-IV) crystallisation

Two protein crystals hits were obtained from initial untagged mPERK (II-IV) screens: (i) 0.1 M Tris-HCl (pH 8.5), 0.2 M Li$_2$SO$_4$, 40% w/v PEG400 (Figure 6-3A) and (ii) 0.1 M MES/imidazole (pH 6.5), 0.09 M sodium phosphate salts (NPS) (mix containing 0.03 M of each NaNO$_3$, Na$_2$HPO$_4$, (NH$_4$)$_2$SO$_4$), 12.5% w/v PEG1000, 12.5% w/v PEG3350, 12.5% v/v MPD (Figure 6-3C). In both conditions, single crystals approximately 20 µm in their longest dimension were easily reproduced using conditions from the initial hit in 48 well sitting-drop MRC plates. Matrix optimization screens were set up as for hPERK (II-IV) but varying the crystallisation solution composition did not make a significant difference in the size of the crystals. It was noticed that the size of these crystals was directly proportional to the time they took to appear with small 5 µm crystals appearing after 2 days and reaching 20 µm after 5 days. Therefore it was reasoned that controlling the speed of crystal growth was key to obtaining large single crystals. Indeed, drops set up with a higher protein to reservoir solution ratio and increased reservoir volume grew slower and yielded larger crystals, as they required longer for equilibration. Final optimized conditions which gave few crystals 50 - 100 µm in size after 5 days when grown in 24-well hanging drop plates at 18 °C were as following: 2 µl protein plus 1 µl reservoir solution, 500 µl reservoir volume, 20 mg/ml of untagged mPERK (II-IV) MPD. The composition of the reservoir solution was identical to the initial hits except that PEG400 and MPD concentrations were increased to 47.5% w/v and 20% v/v respectively as a means of cryoprotecting the crystals (Figure 6-3B,D). Before modifying these, it was verified that it had no effect on the diffraction of the crystals. These will be referred to as Conditions C and D respectively. Once again, SeMet mPERK (II-IV) crystals were set up in identical conditions.
Figure 6-3. mPERK (II-IV) crystals (Conditions C and D)

(A) Initial hits of mPERK (II-IV) crystals grown in 0.1 M Tris-HCl (pH 8.5), 0.2 M Li$_2$SO$_4$, 40% w/v PEG400 at 18 °C.

(B) Optimized crystals grown as in ‘A’ except with 47.5% w/v PEG400 and equilibrated over higher reservoir volume (Condition C).

(C) Initial hits of mPERK (II-IV) crystals grown in 0.1M MES/imidazole (pH 6.5), 0.09 M NPS, 12.5% w/v PEG1000, 12.5% w/v PEG3350, 12.5% v/v MPD at 18 °C.

(D) Optimized crystals grown as in ‘B’ except with 20% v/v MPD and equilibrated over higher reservoir volume (Condition D).
6.3.3 Improving crystallisation by surface entropy reduction

To increase the likelihood of obtaining diffracting hPERK (II-IV) crystals, a surface entropy reduction mutant was created. The property of a protein’s surface, in part, dictates its propensity to crystallise [173]. Flexible regions, which usually contain Lysine, Glutamine and Glutamate residues, must be buried or ordered for crystal contacts to be formed. As such, these regions confer high surface entropy and increase the energy required for crystallisation. Mutating them to small non-polar residues, usually Alanine, has been used as a strategy to lower the energy barrier required for crystallisation to occur. This has been successful in both increasing the number of crystal hits obtained as well as their diffraction limits [178]. The latter is probably due to a tighter packing of the crystal lattice due to the decreased energy required for it to form.

The sequence of hPERK (II-IV) was analysed using the SERp server [173]. Three clustered Glutamate residues (Glu230, Glu231 and Glu233) gave highest scores for coil prediction and entropy values indicating they are probably highly flexible and disordered. In fact, this corresponds to the region just upstream of β9 of hIRE1 luminal domain. No electron density was observed for this region in the crystal structure indicating it is highly flexible. hPERK (II-IV) E230A_E232A_E233A mutant was created, expressed and purified as for wild type hPERK (II-IV). Crystal screens were set up however no hits were obtained.
6.4 PERK (II-IV) native X-ray diffraction data

6.4.1 Native hPERK (II-IV) diffraction data

Initially, only hPERK (II-IV) crystals grown in Condition A were available and therefore most of the work has focused on these. At the concentrations used, the reagents in Condition A do not act as cryoprotectants. Avoiding the formation of ice during flash freezing of the crystals is essential for accurate data collection therefore a number of reagents were tested as cryoprotectants. Solutions were made up containing the reservoir components and additional 20% v/v glycerol, 30% w/v sucrose, 30% v/v ethylene glycol, 30% v/v MPD or 30% w/v PEG400. Although these reagents were successful in preventing icing during freezing, the maximum resolution reached was only 8.5 Å in crystals cryoprotected with 20% v/v glycerol (Figure 6-4A). To test whether this was due to the cryoprotectant, crystals frozen without cryoprotection were mounted onto the goniometer. Diffraction extending to 4.9 Å resolution was achieved (Figure 6-4B). However, as expected, there were large ice rings visible and analysis of the diffraction data was indicative of poor crystal quality (Table 6-2). Firstly, the mosaicity was very high (1.54°). Crystal mosaicity is a measure of the degree of long-range order of the unit cells within the crystal and reflects the quality of crystal packing. Secondly, the \( R_{\text{sym}} \) was high (14.4% overall and 95.3% in the outer shell (despite \( I/\sigma(I) = 2.0 \) in the latter)). \( R_{\text{sym}} \) is a measure of the error of the measurement of each unique reflection throughout the dataset and describes the quality of the data collected. \( R_{\text{sym}} \) is calculated using Equation 10 where \( I \) is the intensity of a reflection and \( <I> \) is the mean intensity of the reflection throughout the dataset.

\[
R_{\text{sym}} = \frac{\sum |I-<I>|}{\sum I}
\]

Finally, the overall B-factor (153.5 Å\(^2\)) was very high and the Wilson Plot of the data varies highly from the linear ‘optimal’ line (Appendix Figure 10-5A). ‘B’ is the B- or temperature-factor and describes the relative motion of a value, the atoms in the crystal in this instance. Wilson Plot is also indicative of order of the scattering atoms in a crystal. The observed intensities (\( I_{\text{obs}} \)) are dependant on experimental factors such as the intensity of the X-ray
beam impinging the crystal, the volume of the crystal which is exposed to the beam as well as how long it is exposed. Thus the measured $I_{obs}$ are proportional, but not identical to, the absolute intensities ($I_{abs}$). In a Wilson Plot, ln ($<I_{obs}>$/]< I_{abs}>) values are plotted against resolution and, if the structure is perfectly ordered, this should yield a straight line with gradient $-2B$. Together these observations indicated that collection of data from non-cryoprotected crystals was not an option. As addition of 20% glycerol was found to decrease the diffraction quality of the crystals, a different means of cryoprotection was investigated. Concentrations of 30% w/v and above PEG3350 have been reported to act as cryoprotectant [179]. Increasing the PEG3350 concentration reduces the water content of the solution. Other than preventing the formation of ice crystals during freezing (cryoprotection), this can also result in more closely packed and better ordered crystal lattices (dehydration). Dehydration has been shown to sometimes extend the X-ray diffraction resolution. Gentle cryoprotection and dehydration of the crystals was achieved by a serial transfer of the coverslip holding the crystallised drop over reservoirs containing increasing concentrations of PEG. PEG3350 concentration was increased stepwise (by 2% w/v and 8 - 12 h incubation at each step) up to a final 40% w/v PEG3500 concentration. A complete 3.2 Å dataset was collected (Figure 6-4C). Compared to data collected without cryoprotection/dehydration, the processed data shows lower mosaicity (0.97 °), $R_{sym}$ (8.0%), overall B-factor (80.3 Å$^2$) and fit of the Wilson Plot (Table 6-2; Appendix Figure 10-5B). Together these indicate improved crystal and diffraction data quality. A Matthews coefficient (Vm) of 2.44 Å$^3$/Da and 49.5% solvent content (Vs) were obtained (0.98 probability). This corresponds to two hPERK (II-IV) molecules in the asymmetric unit.
Figure 6-4. Native hPERK (II-IV) Condition A crystal diffraction

(A-C) The looped crystal mounted onto the goinometer and a diffraction image of hPERK (II-IV) crystals (A) cryoprotected with 20% glycerol, (B) frozen without cryoprotection and (C) dehydrated over 40% w/v PEG 3350 are shown. The resolution limit is indicated. Glycerol cryoprotection increases the resolution of diffraction compared to non-cryoprotected crystals. However in the absence of cryoprotection, large ice rings are present. Gradual dehydration successfully cryoprotects the crystals and improves their diffraction.
<table>
<thead>
<tr>
<th></th>
<th>No cryoprotection</th>
<th>Dehydration</th>
<th>Dehydration + seeding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell dimensions (Å)</td>
<td>85.6 85.6 189.3</td>
<td>83.9 83.9 184.4</td>
<td>83.6 83.6 184.9</td>
</tr>
<tr>
<td>(Cell angles (°))</td>
<td>(90 90 90)</td>
<td>(90 90 90)</td>
<td>(90 90 90)</td>
</tr>
<tr>
<td>Space group</td>
<td>P41212</td>
<td>P41212</td>
<td>P41212</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>59.6 – 4.9</td>
<td>83.9 – 3.4</td>
<td>49.8 – 3.2</td>
</tr>
<tr>
<td></td>
<td>(5.0 – 4.9)</td>
<td>(3.5 – 3.4)</td>
<td>(3.3 – 3.2)</td>
</tr>
<tr>
<td>Mosaicity (°)</td>
<td>1.54</td>
<td>0.97</td>
<td>0.65</td>
</tr>
<tr>
<td>Rsym (%)</td>
<td>14.4 (95.3)</td>
<td>8.0 (81.9)</td>
<td>4.8 (65.7)</td>
</tr>
<tr>
<td>I/σ(I)</td>
<td>13.3 (2.0)</td>
<td>14.8 (2.0)</td>
<td>15.5 (2.7)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>3.9 (3.2)</td>
<td>4.1 (3.7)</td>
<td>6.0 (6.1)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>98.7 (93.7)</td>
<td>99.9 (97.5)</td>
<td>98.8 (98.8)</td>
</tr>
<tr>
<td>Overall R-factor (Å²)</td>
<td>153.5</td>
<td>80.3</td>
<td>89.7</td>
</tr>
</tbody>
</table>

Table 6-2. Native hPERK (II-IV) Condition A diffraction data

The diffraction images for hPERK (II-IV) crystals collected without cryoprotection, dehydrated over 40% w/v PEG3350 or dehydrated plus 10 rounds of reiterative seeding were integrated, merged and scaled; the results obtained are summarised. R_{sym}, I/σ(I), multiplicity and completeness values refer to values in the overall and outer shell, in brackets.
Crystals grown with and without reiterative seeding were compared and although the resolution limit achieved was not improved, crystals grown from with reiterative seeding diffracted to resolution higher than 4 Å more consistently. Defects accumulate in crystals as they grow and this may limit the size of the crystals achieved. Multiple rounds of reiterative seeding thus can improve crystal quality. In fact, the approximate ratio of crystals diffracting to 4 Å or higher was 1:30 for seeded and 1:3 for reiteratively seeded crystals. This difference was especially important for the collection of derivative datasets where extensive screening of heavy atom type, concentration and soaking time is already required and therefore good quality crystals are a prerequisite.

Although native Condition A crystals diffracted to an acceptable resolution, initially (prior to attempting reiterative seeding) the very small proportion of crystals reaching 4 Å or more was a bottleneck for heavy atom derivatization. Attempts to obtain alternative crystal forms included matrix microseeding, which gave rise to Condition B crystals, and crystallisation of murine and drosophila orthologues as discussed below. Condition B crystals were cryoprotected using 30% v/v glycerol and 30% w/v PEG1500. Since these reagents are present in the original condition, albeit at lower concentrations, it was a reasonable and gentle cryoprotection technique to try in a first instance. Indeed, by comparing crystals harvested with and without cryoprotection, the resolution limit achieved was not affected but cryoprotection did avoid the formation of ice crystals as seen by the absence of ice rings. However, the maximum resolution achieved was only 8.7 Å (Table 6-2). Condition B crystals were unsuitable for further experiments; no complete data sets were collected.
6.4.2  Native mPERK (II-IV) diffraction data

mPERK (II-IV) native crystals from Condition C and D were tested for diffraction (Table 6-3). These crystals consistently diffracted better than hPERK (II-IV) crystals. The cell dimension and space group of two mPERK (II-IV) datasets were almost identical suggesting both crystals represent the same crystal form. Crystals grown in Condition D diffracted better than those grown in Condition C. At best, a complete 3.3 Å dataset from Condition D was collected. The statistics of the processed diffraction data were within typical ranges of low-resolution data. However the $R_{sym}$ values (8.0% overall and 81.9% in the outer shell) were high. This was particularly of hindrance for experimental phasing as discussed later. A $V_m$ of 2.47 Å$^3$/Da and $V_s$ of 50.2% were obtained (1.00 probability) which corresponds to one mPERK (II-IV) molecule in the asymmetric unit.
<table>
<thead>
<tr>
<th></th>
<th>Condition B</th>
<th>Condition C</th>
<th>Condition D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell dimensions (Å)</td>
<td>161.1 161.1 61.2 (90 90 120)</td>
<td>89.9 89.9 76.7 (90 90 120)</td>
<td>87.6 87.6 73.6 (90 90 120)</td>
</tr>
<tr>
<td>(Cell angles (*)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>P3</td>
<td>P3,21</td>
<td>P3,21</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>58.3 – 8.7 (8.8 – 8.7)</td>
<td>54.6 – 4.2 (4.3 – 4.2)</td>
<td>52.8 – 3.2 (3.3 – 3.2)</td>
</tr>
<tr>
<td>Mosaicity (*)</td>
<td>-</td>
<td>1.34</td>
<td>0.97</td>
</tr>
<tr>
<td>Rsym (%)</td>
<td>-</td>
<td>15.2 (84.3)</td>
<td>8.0 (81.9)</td>
</tr>
<tr>
<td>I/σ(I)</td>
<td>-</td>
<td>14.8 (3.8)</td>
<td>14.5 (2.6)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>-</td>
<td>7.6 (7.8)</td>
<td>6.1 (6.4)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>-</td>
<td>99.9 (100.0)</td>
<td>99.7 (99.7)</td>
</tr>
<tr>
<td>Overall B-factor (Å²)</td>
<td>-</td>
<td>93.6</td>
<td>85.3</td>
</tr>
</tbody>
</table>

Table 6-3. Native hPERK (II-IV) Condition B and mPERK (II-IV) Condition C and D diffraction data
The diffraction images were integrated, merged and scaled; the results obtained are summarised. $R_{sym}$, $I/\sigma(I)$, multiplicity and completeness values refer to values in the overall and outer shell, in brackets. No data set was collected for Condition B.
6.5 Phasing

6.5.1 Molecular replacement

The luminal domains of hPERK and hIRE1 are functionally interchangeable and appear to behave similarly with respect to UPR activation function [76,103]. In addition, secondary structure prediction suggests they share overall similar structural folds. However, hPERK (and mPERK) and hIRE1 luminal domains only share 18% sequence identity.

Molecular replacement was attempted to solve the structure of hPERK (II-IV) [164,165]. The crystal structures of hIRE1 (PDB: 2HZ6) and IRE1p (PDB: 2BE1), and in silico models generated based on PERK luminal domain sequences using the Phyre server were all used as search ensembles [174]. hIRE1 and IRE1p coordinates were inputted both as monomers and dimeric assemblies. Regions which were not conserved or highly flexible were truncated to generate additional search models. However, the translation function Z-score (TFZ) (<2) and log-likelihood gain (LLG) scores (<20) of any solution found were very low. TFZ and LLG scores reflect agreement between the Patterson maps of the experimental data and those of the search ensemble in a specific orientation. Nonetheless, these solutions were not discarded outright.

Symmetry related mates were generated using the known crystal symmetry operations. In these orientations, there was extensive clashing and/or improbably large solvent channels between the proteins. Together these results are dissuasive of correct molecular replacement solutions. Experimental phasing was therefore used to solve the structure of PERK luminal domain.

6.5.2 Phasing by SeMet anomalous scattering

6.5.2.1 hPERK (II-IV) and mPERK (II-IV) SeMet incorporation

SeMet-labelled proteins were expressed as described and purified identically to the native proteins. MALDI-TOF spectrometry was used to measure the precise mass of the proteins.
and thus determine the quantity of SeMet incorporation (Appendix Figure 10-6). The masses obtained were 34.431 kDa for SeMet hPERK (II-IV) and 34.221 kDa for SeMet mPERK (II-IV). These values indicate the presence of 7.5 and 6.9 Se atoms respectively per protein molecule. Given that hPERK (II-IV) and mPERK (II-IV) have 8 and 7 Methionine residues respectively per monomer, the obtained masses indicate near complete SeMet incorporation (94% for SeMet hPERK (II-IV) and 99% for SeMet mPERK (II-IV)).

6.5.2.2 SAD and MAD experiments

SeMet hPERK (II-IV) Condition A and SeMet mPERK (II-IV) Condition D crystals were used to collect derivative datasets. SAD data was collected at the peak wavelength; MAD data was collected on a single crystal at peak, inflection and higher energy remote wavelengths. These values were experimentally obtained by carrying out fluorescence scans prior to data collection. SeMet hPERK (II-IV) Condition A derivative crystals diffracted too poorly (5 Å and above) for use in phasing.

SeMet mPERK (II-IV) Condition D derivative crystals reached a maximum of 3.4 Å resolution. The scattering factors were \( f'' = 4.85 \) and \( f' = -7.65 \) at the peak wavelength (0.9789 Å). Crank, Shelix C, Phenix AutoSol programs were used to attempt to locate the Se scatters and obtain phases [164,165]. However no solutions were obtained using this dataset.

The approximate anomalous signal can be calculated by estimating the ratio of anomalous scattering to the expected total scattering of a protein (Equation 11 and Equation 12 where \( N \) is the number of anomalous scatterers, \( n \) is the total number of residues, \( f''_A \) is the scattering factor of the anomalous data set and \( F_T \) is the total scattering).

**Equation 11**

\[
\sqrt{2N \frac{f''_A}{rms|F_T|}}
\]

**Equation 12**

\[
rms|F_T| \approx \sqrt{346n}
\]
Assuming 100% SeMet incorporation, and that these residues are ordered, a 5.7% anomalous signal is expected for this 3.4 Å SeMet mPERK (II-IV) dataset. The anomalous signal is below the noise (overall Rsym (8.9%)). Additionally the limit of the anomalous signal, determined by using a 30% anomalous correlation coefficient (CCanom) cut-off, only reached a maximum of 5.6 Å. Se has 34 electrons. Although this is significantly different from nitrogen, oxygen and carbon atoms of proteins, it is low for a heavy atom. The size of an atom is partly dependent on the number of electrons it has. As such, the more electrons an atom has, the higher the minimum resolution at which differences between it and the atoms in a protein can be discerned. As a rule of thumb, anomalous signal to ~4 Å resolution is required to locate Se atoms for phasing. The exact number depends on other factors such as Se incorporation, how ordered they are, their scattering properties and the quality of the dataset. Taken together, these observations can help explain why phases could not be obtained from this dataset. Finally, since these crystals were not isomorphous to the native dataset (20.1% overall Rmerge), SIR and Single Isomorphous Replacement with Anomalous Scattering (SIRAS) experiments were not attempted.

The anomalous signal of scatterers is highly dependent on its redox state [180]. In the case of SeMet, this affects both the magnitude and energy of the anomalous signal. The presence of a mixture of SeMet oxidation states in the crystal thus results in a heterogeneous and dispersed absorption peak. This reduces the overall anomalous signal and also makes it difficult to determine precise peak and inflection absorbance energies. SeMet mPERK (II-IV) Condition D crystals were oxidized or reduced immediately prior to freezing. The better diffracting dataset was obtained with reduced SeMet mPERK (II-IV) Condition D crystals. The absorbance peak (0.9794 Å) was shifted, significantly more sharp and homogenous compared to that of untreated crystals (Figure 6-5). In addition, the scattering factors were increased (f"=7.59 and f’ = -8.91). This amounts approximately to 9.3% anomalous signal. This time, the anomalous signal was measured up to 4.3 Å. Although reduction of the SeMet significantly improved the anomalous signal, no phases could be obtained in this instance either. Heavy atom derivatization was attempted next for phasing.
**Figure 6-5. SeMet mPERK (II-IV) Condition D anomalous signal**

(A) (i) Fluorescence scan of untreated SeMet mPERK (II-IV) Condition D crystals. The absorbance peak at 12666.2 eV (0.9789 Å) is wide. Scattering factors were $f'' = 4.85$ and $f' = -7.65$. (ii) The CC$\text{anom}$ falls below 30% at 5.61 Å.

(B) (i) Fluorescence scan of reduced SeMet mPERK (II-IV) Condition D crystals. The absorbance peak at 12659.5 eV (0.9794 Å) is sharper than in ‘A’. The scattering factors ($f'' = 7.59$ and $f' = -8.91$) are higher than in ‘A’. (ii) The CC$\text{anom}$ falls below 30% at 4.27 Å.
6.5.3 Heavy atom derivatization

Crystals were soaked with numerous heavy atom salts for derivatization. Owing to the large number of Cysteine, Methionine and Histidine residues, Mercury (Hg), Platinum (Pt) and Gold (Au) were mainly used. The sulphur atom of Cysteine and imidazole ring of Histidine are most reactive at pH above 7 and 6, respectively. The crystallisation buffer of hPERK (II-IV) and mPERK (II-IV) was compatible with that. The results obtained are summarised in Appendix Table 10-2 and Appendix Table 10-3. The data can be grouped into three categories: (i) no heavy atom binding, (ii) binding but not adequate diffraction or (iii) binding and diffraction. For derivatives in the latter group (iii), SAD datasets were collected at the experimentally derived peak wavelength. Where crystal quality and radiation damage permitted, MAD experiments were also collected at peak, inflection and higher energy remote wavelengths. Where the derivative and native datasets were isomorphous, SIR, MIR, SIRAS and/or Multiple Isomorphous Replacement with Anomalous Scattering (MIRAS) were also attempted. Crank, Shelx C, Phenix AutoSol programs were used for phasing [164,165]. However, in a first instance, all derivative datasets were unsuccessful in yielding any phase information. A number of factors including poor diffraction and outer limit of the anomalous signal, non-specific or insufficient binding of the heavy atoms, disorder in the regions of heavy atom binding, poor merging data statistics (higher noise than anomalous signal) might be responsible for this.

A wide range of heavy atom salts, concentrations, times of soaking and method (or absence of) backsoaking were tested. This extensive number of variables limited the number of crystals that could be tested for each condition. Native gel shift assays were thus carried out to screen more efficiently for heavy atom binding (Appendix Figure 10-7). This was done on hPERK (II-IV) at first since the quality of these crystals and the native dataset were better compared to those of mPERK (II-IV). Soaking of hPERK (II-IV) with a Tungsten (W) salt (Na₂WO₄) caused a large shift of the protein compared to unbound hPERK (II-IV). hPERK (II-IV) Condition A crystals were then soaked with 0.5 - 10 mM Na₂WO₄ for 0.5 - 24 h and were frozen with and without backsoaking in the mother liquor. A three-wavelength MAD set of datasets was collected on a hPERK (II-IV) Condition A crystal soaked with 2 mM Na₂WO₄ for 5 h (no backsoaking). The data were integrated, merged and scaled (Table 6-4).
Phasing was carried out using Shelx C/D/E via the AutoSharp pipeline [167]. Two W atoms (1 and 0.93 occupancy) were located within the asymmetric unit with an overall correlation coefficient (CCall) of 47% (Appendix Figure 10-8). These appear to be mainly coordinated by the aromatic ring of a solvent exposed Trp175 residue. The figure of merit (FOM) of the original hand (P4₁2₁2) (0.46) was significantly higher compared to that of the inverted hand (P4₁2₁2) (0.21) indicative of a correct solution for the former.
<table>
<thead>
<tr>
<th></th>
<th>Peak</th>
<th>Inflection</th>
<th>Remote</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelength (Å)</td>
<td>1.2148</td>
<td>1.2152</td>
<td>0.9795</td>
</tr>
<tr>
<td>Cell dimensions (Å) (Cell angles (°))</td>
<td>83.9 83.9 186.2 (90 90 90)</td>
<td>84.1 84.1 186.5 (90 90 90)</td>
<td>84.2 84.2 186.9 (90 90 90)</td>
</tr>
<tr>
<td>Space group</td>
<td>P4₁2₁2</td>
<td>P4₁2₁2</td>
<td>P4₁2₁2</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>75.6 – 3.1 (3.2 – 3.1)</td>
<td>76.8 – 3.3 (3.4 – 3.3)</td>
<td>76.8 – 3.6 (3.7 – 3.6)</td>
</tr>
<tr>
<td>Mosaicity (°)</td>
<td>0.63</td>
<td>0.67</td>
<td>0.84</td>
</tr>
<tr>
<td>Rsym (%)</td>
<td>4.8 (65.7)</td>
<td>5.7 (82.8)</td>
<td>6.7 (93.2)</td>
</tr>
<tr>
<td>I/σ(I)</td>
<td>32.3 (4.4)</td>
<td>30.9 (3.4)</td>
<td>25.9 (2.9)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>10.4 (11.0)</td>
<td>10.3 (10.2)</td>
<td>10.3 (11.0)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.9 (99.5)</td>
<td>99.9 (100.0)</td>
<td>99.9 (99.8)</td>
</tr>
<tr>
<td>Overall B-factor (Å²)</td>
<td>96.5</td>
<td>90.3</td>
<td>89.9</td>
</tr>
<tr>
<td>Anom compl. (%)</td>
<td>99.5 (100.0)</td>
<td>99.6 (100.0)</td>
<td>99.4 (99.4)</td>
</tr>
<tr>
<td>Anom multiplicity</td>
<td>5.6 (5.8)</td>
<td>5.6 (5.3)</td>
<td>5.7 (5.9)</td>
</tr>
<tr>
<td>CCAnom (%)</td>
<td>73.4 (9.3)</td>
<td>47.3 (3.7)</td>
<td>34.3 (-0.3)</td>
</tr>
</tbody>
</table>

Table 6-4. Na₂WO₄ derivative hPERK (II-IV) Condition A MAD diffraction data
The diffraction images were integrated, merged and scaled; the obtained results are summarised. \( R_{sym} \), \( I/\sigma(I) \), multiplicity, completeness, anomalous completeness (Anom compl.), anomalous multiplicity (Anom multiplicity) and CCAnom refer to values in the overall and outer shell, in brackets.
6.6 Model building and structure refinement

6.6.1 hPERK (II-IV)

The two W atoms were used to phase the reflections of the MAD peak dataset. Since this was the best data obtained it was used as the native dataset for structure solution. Initial electron density maps were very poorly defined (Figure 6-6A). Density modification was carried out to improve solvent boundaries [167]. From this map, polyAla chains were threaded where electron density continuity could be seen. Eight extended β-strands were apparent; this feature is also present in IRE1 luminal domain (Figure 6-6B). The hIRE1 luminal domain structure was superimposed, using secondary-structure matching (SSM) alignment, uniquely based on this β-sheet region to help with initial assignment of residues.

The use of the FEMs as implemented by Phenix was of particular help during model building in poorly defined areas [164]. In FEMs a Hoppe-Gassmann (polynomial) distribution is used during density modification to reduce/remove noise and enhance weak signals. Additionally, local map scaling (as opposed to a global refinement) avoids obscuring weak signal (such as that by partially occupied sites) by strong signal (such as heavy atom signals). This allowed for model building and map improvement in areas for which initial density was very poor or in flexible regions at the end of chain fragments (Figure 6-6C).
Figure 6-6. hPERK (II-IV) structure refinement  
(A) Initial electron density map of phased hPERK (II-IV) diffraction data. The map shows poor connectivity and few solvent boundaries.  
(B) Improved electron density map after density modification. Solvent boundaries and secondary structural features (eight β-strands) are discernable.  
(C) Electron density (2Fo-Fc map) around (i) an α-helix and (ii) a β-strand of hPERK (II-IV) (a) before and (b) after refinement. The density was significantly improved. This is reflected in the R-factors: 39.3% R-work and 43.7% R-free before refinement and 24.2% R-work and 29.3% R-free after completion of refinement.
Two hPERK (II-IV) monomers were present in the asymmetric unit. These have identical structural features and when they are superimposed by residue-based least square (LSQ) alignment, the central β-sheets overlay perfectly. However, there is a shift of ~2 - 4 Å in the position of the remainder of the structure. Since there was a lack of perfect symmetry between two monomers within the asymmetric unit, non-crystallographic symmetry (NCS) could not be applied. Several rounds of model improvement and refinement were carried out until the model could not be improved further. Final refined statistics are summarised in Table 6-5.

6.6.2 mPERK (II-IV)

The refined structure of hPERK (II-IV) was used as a model for molecular replacement to phase mPERK (II-IV) Condition D native diffraction data (LLG = 416; TFZ =20.2) [165]. The poor quality of the diffraction data and low resolution made it challenging to achieve a well-refined structure. The final statistics are summarised in Table 6-5.
### Table 6-5. hPERK (II-IV) and mPERK (II-IV) refinement

The final refinement statistics are summarised. Resolution values refer to outer and inner diffraction limits. R-work and R-free values refer to the overall and outer shell, in brackets, data. Protein residues present in chain A / chain B of the final dimer models (298 residues expected per monomer).

<table>
<thead>
<tr>
<th></th>
<th>hPERK (II-IV)</th>
<th>mPERK (II-IV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution</td>
<td>76.5 – 3.1</td>
<td>54.4 – 3.3</td>
</tr>
<tr>
<td></td>
<td>(3.2 – 3.1)</td>
<td>(3.4 – 3.3)</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>22118</td>
<td>11682</td>
</tr>
<tr>
<td>R-work (%)</td>
<td>24.2 (28.7)</td>
<td>28.6 (36.5)</td>
</tr>
<tr>
<td>R-free (%)</td>
<td>29.3 (32.3)</td>
<td>30.6 (40.5)</td>
</tr>
<tr>
<td>Protein residues</td>
<td>231 / 197</td>
<td>211 / 211</td>
</tr>
<tr>
<td>RMSD (bonds) (Å)</td>
<td>0.004</td>
<td>0.003</td>
</tr>
<tr>
<td>RMSD (angles) (°)</td>
<td>0.991</td>
<td>0.935</td>
</tr>
<tr>
<td>Ramachandran favoured (%)</td>
<td>91.2</td>
<td>83.9</td>
</tr>
<tr>
<td>Ramachandran outliers (%)</td>
<td>1.2</td>
<td>3.1</td>
</tr>
<tr>
<td>Clash score</td>
<td>13.5</td>
<td>11.7</td>
</tr>
</tbody>
</table>
6.7 PERK luminal domain structure

6.7.1 hPERK (II-IV)

6.7.1.1 Monomers

The overall fold of hPERK (II-IV) luminal domain is very similar to that of IRE1 (3.8 Å and 4.2 Å rmsd values between hPERK (II-IV) and IRE1p or hIRE1 luminal domain monomers respectively). hPERK (II-IV) monomer comprises of a triangular assembly of β-sheet sandwiches (19 β-strands in total) interspersed by two α-helices (Figure 6-7). The amino acid chain zigzags back and forth between these structural elements. These features therefore do not fold independently and linearly based on the amino acid sequence. hPERK (II-IV) contains a central bed of extended β-strands (β8, β11, β12) with a transverse short α-helix (α1) and two flanking short β-strands (β9, β10). This central β-sheet/α-helix core is flanked by two Lobes of β-sheets (Lobe 1: β1-3, β13-17; Lobe 2: β4-7, β18, β19, α2).
Figure 6-7. hPERK (II-IV) monomer
(A) Cartoon diagram of hPERK (II-IV) monomer comprising regions II to IV. hPERK (II-IV) is composed of a triangular assembly of β-sheets interspaced by two α-helices (core = magenta; Lobe 1 = blue; Lobe 2 = cyan).
(B) Topology diagram of hPERK (II-IV). β-strands are shown as arrows; α-helices are shown as cylinders. Continuous black lines indicate resolved residues without secondary structure; dotted grey lines indicate flexible regions for which no electron density was observed.
6.7.1.2  Dimers

Two hPERK (II-IV) monomers come together via zippering-up of two adjacent monomers (Figure 6-8). hPERK (II-IV) dimerisation interface is referred to as Interface 1. Interface 1 involves 38 amino acids and buries a total solvent-exposed surface area of ~2300 Å² of an otherwise solvent-exposed surface of each monomer. It is stabilized mainly by five hydrogen bonds between two antiparallel β8 strands of adjacent molecules (Figure 6-9A; Appendix Table 10-4). Compared to IRE1 luminal domains, hPERK (II-IV) has two additional β-strands (β9 and β10). These are closely associated with the central β-barrel core and also contribute to Interface 1. β9 and β10 form four hydrogen bonds with β11 and the loop between β11 and β12 of two interacting monomers (Figure 6-9B; Appendix Table 10-4). Finally, Glu184 in helix α1 forms a salt bridge with Lys198 in strand β8 of an adjacent monomer (Figure 6-9C).

A point mutation was designed to disrupt dimer formation based on previous work on hIRE1 luminal domain [101]. The hIRE1 L194P mutant shifted the equilibrium towards a monomeric population in sedimentation equilibrium experiments. Leu196 of hIRE1 corresponds to Leu200 of hPERK. hPERK (II-IV) L200P was expressed and purified similarly to the wild type protein. Its oligomeric state was analysed by MALS (Figure 6-9D). Whereas wild type hPERK (II-IV) eluted as a homogenous 67.8 kDa dimer, hPERK (II-IV) L200P eluted as a single homogenous species of 33.5 kDa, corresponding to monomers. This confirms the dimer observed in the crystal structure is also present in solution and thus likely to represent the true in vivo dimeric assembly.
Figure 6-8. hPERK (II-IV) dimers

(A) Schematic representation of two hPERK (II-IV) monomers coming together to form dimers. Each luminal domain sub-region is coloured differently according to

(B) Cartoon (left) and surface (right) diagram of hPERK (II-IV) dimer. Each monomer is coloured differently (yellow and purple).
Figure 6-9. hPERK (II-IV) Interface 1

(A-C) Interface 1 between hPERK (II-IV) monomer A (yellow) and monomer B (purple) is stabilized by extensive hydrogen bonding between (A) two interacting β8, (B) β9-β10 and β11 and the loop between β11 and β12, and (C) a salt bridge between α1 and β8. Residues involved and atom distances (Å) are annotated.

(D) SEC MALS analysis (Superdex 200 PC 3.2/30) of hPERK (II-IV) (100 μM) and hPERK (II-IV) L200P (100 μM). The calculated MW (kDa) of the peaks obtained is indicated. The two species elute as homogeneous dimeric and monomeric species respectively.
6.7.1.3  
Tetramers

The crystal lattice of hPERK (II-IV) is stabilized by the packing of helix α2 of a dimer A against the β-sheets and loops of Lobe 2 (β4-7, β18, β19) of a dimer B (Figure 6-10; Figure 6-11A). This is referred to as Interface 2. As such, ‘swapping’ of helices α2 mediates intertwining of two hPERK (II-IV) dimers to form a circular tetramer. All four α2 helices participate in tetramer formation.

Interface 2 involves 38 residues and buries a total solvent exposed surface area of ~2500 Å². This area is solvent-accessible in the dimer assembly. Interface 2 is strongly hydrophobic (ΔG = -17.9 kcal/mol) (Figure 6-11B). The main contributors to the hydrophobic core are found on helix α2 (Ile374, Val375, Ala377, Ala378, Gly380, Ala381, Ser385, Val386), β18 (Tyr387, Leu388, Gly389, Met390), β19 (Gly393, Leu395, Try396, Leu397, Ser399, Ser400) and the loop between β6 and β7 (Trp165, Met172) from two interacting molecules. Seven hydrogen bonds, which mainly involve interaction of the side chains of Lobe 2 (β5-7, β19-20) residues with the backbone oxygen atoms of helix α2 of a facing dimer also stabilize Interface 2 (Appendix Table 10-4).
Figure 6-10. hPERK (II-IV) tetramerisation

Two hPERK (II-IV) dimers (purple/yellow and magenta/orange) come together to form closely packed tetramers via interaction of helix $\alpha2$ with Lobe 2 of an interacting dimer. The structure is represented as cartoon (left) and surface (right).
Figure 6-11. hPERK (II-IV) Interface 2

(A) Packing of hPERK (II-IV) in the unit cell. Only a few molecules are shown for clearer visualisation. The tetrameric assembly in the lattice is apparent.

(B) The hydrophobic nature (rich in Valine, Isoleucine, Leucine, Methionine, Alanine, Tyrosine, Serine and Glycine residues) is shown. These are found on helix α2, β18 and β19 that are the major contributors to Interface 2.
SEC MALS analysis showed that hPERK (II-IV) appears to form homogenous dimers in solution (Figure 6-9D). The presence of hPERK (II-IV) tetramers was detected in cross-linking and AUC experiments. To investigate whether tetramerisation via Interface 2 could be biologically relevant or if it is simply an artifact of crystallisation, targeted mutations were introduced. Firstly, the conserved hydrophobic Trp165 of β6 that stabilizes the accommodation of helix α2 of an opposite dimer was mutated to a non-polar Alanine (W165A). Secondly, the hydrophobic Leu388 of β18 that forms a hydrogen bond with Arg379 of the interacting α2 was mutated to a polar Asparagine (L388N). Finally, a truncation protein lacking the α2 helix (hPERK 105-384) was cloned; this will be referred to as hPERK (II-IV) Δα2. AUC was carried out at identical protein concentrations on wild-type hPERK (II-IV) and the three described mutants to determine the dimer to tetramer ratio (Figure 6-12). hPERK (II-IV) W165A, L388N, W165A_L388N and Δα2 proteins reduced the amounts of tetramer species compared to the wild-type protein. This confirms that in solution tetramer formation similarly occurs as in the crystal lattice via binding of the α2-helices with Lobes 2 of interacting dimers at Interface 2.
Figure 6-12. Disruption of hPERK (II-IV) Interface 2

(A) α2 helix binding of dimer A (orange) is stabilized by hydrophobic interaction with residues of Lobe 2 of dimer B (purple). Mutated Trp165 and Leu388 residues are shown. Distances (Å) of Trp165 with α2 backbone and of the hydrogen bond between Leu388 with Arg379 are annotated.

(B) AUC analysis of hPERK (II-IV) wild-type, W165A, L388N, W165A_L388N and Δα2 mutants (20 μM). The c(s) distribution is plotted versus the sedimentation coefficient (S) in Svedberg unit to show the distribution of the different species according to their molar mass. The expected position of dimers (70 kDa) and tetramers (140 kDa) are shown (grey). The dimer to tetramer ratio of the curves in ‘B’ is shown. The change in amount of tetramers compared to wild-type hPERK (II-IV) (%) was calculated. Tetramerisation is greatly compromised in all mutant proteins. AUC experiments and data analysis was carried out by Dr K. Stott.
Looking at the electrostatic potential of PERK’s surface, the groove formed by β5-7 and β19-20 is strongly negatively charged (Figure 6-13). On the other hand, α2 is mostly positively charged. If these structures are not stabilized by tetramer formation, they are solvent exposed. In biology, large charged surfaces are rarely solvent exposed unless they serve a function. In the case of PERK Lobe 2, the opposite charges of Lobe 2 α2 and its β-sheet cluster allude to an energetically favourable ‘lock and key’ analogy for tetramerisation. This further highlights a probable biological importance of Interface 2.

6.7.2 mPERK (II-IV)

Structurally, the overall core and Lobe 1 of mPERK (II-IV) are identical to those of hPERK (II-IV) (1.1 Å rmsd). No electron density was observed for Lobe 2, which is involved in mediation tetramer formation in hPERK (II-IV), and no tetrameric assemblies were seen (Figure 6-14). Crystals were washed and analyzed by SDS-PAGE to confirm they contained the intact mPERK (II-IV) protein. The crystal lattice contacts are between antiparallel stacking of two β15 and β9-β10 interaction with β11-β12 of symmetry related molecules. These observations highlight the role of Lobe 2 in mediating tetramer formation at Interface 2. Helix α2 and the β-strands of Lobe 2 (β5-7, β19-20) are implicated in stabilising tetramers and consequently are themselves stabilized by tetramer formation in hPERK (II-IV). In mPERK (II-IV) these regions are disordered because tetramers do not form and thus do not stabilise helix α2 and the Lobe 2 β-strands. It follows that they were not observed in electron density maps.
Figure 6-13. PERK Lobe 2 electrostatic charge

(A) The electrostatic potential of the surface around Lobe 2 is shown. The α2 helix, especially at its tip, is positively charged whereas the cavity to which it binds to form PERK tetramers is strongly negatively charged.

(B) The cartoon representation of PERK Lobe 2 is superimposed on the surface charge as shown in ‘A’.
Figure 6.14. mPERK (II-IV) dimeric structure

(A) The structure of dimeric mPERK (II-V) (red) was superimposed onto that of tetrameric hPERK (II-IV) (only one dimer is shown) (grey). Structures are shown as ribbons.

(B) Crystal packing of mPERK (II-IV) dimers in the unit cell.
6.8 Discussion

This chapter presents the novel structure of the *H. sapiens* and *M. musculus* orthologues PERK luminal domain, a major constituent of ER stress sensing and UPR signalling in higher eukaryotes. The structures of tetrameric hPERK (II-IV) and dimeric mPERK (II-IV) were solved by X-ray crystallography to 3.1 and 3.3 Å, respectively.

Early studies observed that induction of ER stress resulted in the formation of higher PERK and IRE1 oligomers *in vivo* [76,87]. Nonetheless to date, the oligomeric states of inactive and active PERK and IRE1 have not been elucidated. From the results presented in this chapter, a biological role for PERK's tetrameric assembly can be envisaged.

The monomer of PERK luminal domain is composed of a triangular assembly of β-sheets interspaced by two α-helices. Two monomers come together to form dimers. Superimposition of all known PERK and IRE1 luminal domain structures reveals that their core remains mostly immobile (Appendix Figure 10-9 and Appendix Figure 10-10). In particular, Interface 1, which mediates dimerisation, is entirely conserved. Given their functional resemblance, this is not surprising.

Similarly to IRE1 structures, PERK luminal domain has a short α-helix (α1) that lies transversely along the central bed of β-sheets (β8, β11-12). In the IRE1p structure this was postulated to mimic an MHC-like peptide-binding groove, and allow for direct binding of unfolded proteins and consequently lead to IRE1p activation during ER stress. The structures of PERK luminal domain do not support this mechanism of UPR activation by direct binding of unfolded proteins. Firstly, although the secondary structures in this region are highly conserved, DALI fold recognition program did not link PERK's luminal domain core to any known structural motif other than IRE1 luminal domains [175]. Secondly, the width of the groove between two PERK α1 helices is too narrow to accommodate unfolded proteins (15 Å in hPERK and mPERK; 14 Å in hIRE1; 17 Å in IRE1p). Thirdly, residues proposed to be essential for peptide binding in IRE1p are Met229, Phe285, and Tyr301. Based on secondary structure alignment, these correspond to Lys198, Lys243 and Val261 respectively in PERK. These residues are all non-conservative mutations. In addition, Lys198 (on β8) forms a salt bridge with Glu184 (on α2) of an adjacent monomer and sterically occludes access to the binding
groove. Similar observations were made for hIRE1 luminal domain. Together these results point away from a direct peptide binding mechanism for ER stress sensing by UPR sensors of higher eukaryotes.

The most striking difference between PERK and IRE1 luminal domain structures is within Lobe 2 and the tetramerisation Interface 2 (Appendix Figure 10-9). In the tetrameric structure of hPERK, the C-terminal α-helix (α2) of a dimer is stabilized against the β-strand cluster of Lobe 2 (β5-7, β19-20) of an interacting dimer. Targeted mutations of Interface 2 designed to disrupt its hydrophobic nature, as well as by truncation of α2, greatly compromised PERK tetramerisation. These results confirm that Interface 2 is present in solution and is not an artefact of the unphysiologically high protein concentrations used during crystallisation.

Based on structural alignment, the region implicated in Interface 2 is conserved between PERK and IRE1 orthologues. Most remarkable is the presence of the hydrophobic C-terminal α-helix as well as the VYLG-X-Y consensus sequence, where X is a positively charged amino acid. These non-polar residues are major contributors to the hydrophobic nature of Interface 2. ‘X’ is accountable for the electrostatic forces that also stabilize α-helix binding to Lobe 2.

Interestingly, no electron density for the entire Lobe 2 was observed for the mPERK luminal domain implying that it has low occupancy in any given orientation. In the hPERK (II-IV) structure, Lobe 2 interacts only with an adjacent dimer. In an isolated dimer, Lobe 2 would be solvent exposed. Given that mPERK (II-IV) adopts exclusively a dimeric arrangement in the crystal lattice, Lobe 2 is probably mobile. It is therefore not surprising that this region was not observed in electron density maps.

_H. sapiens_ and _M. musculus_ PERK luminal domains sequences share 90% identity and 98% similarity. Differences occur almost exclusively in their most N-terminal region, which was not present in the PERK (II-IV) constructs used for crystallisation. The C-terminal Lobe 2 region that is key for tetramer formation is 100% conserved. Their different oligomeric states thus probably reflect the assemblies in which they were trapped during crystallisation rather than a genuine difference between the two species. The structures of hPERK and mPERK luminal domains can therefore be used to describe PERK luminal domain tetrameric and dimeric assemblies respectively.
Three-dimensional (3D) domain swapping was first proposed by Moore and colleagues to explain the aggregation of RNase A after lyophilisation [181]. Eisenberg and colleagues later defined its mechanistic framework [182]. Domain swapping involves the replacement of a domain of a protein with the same domain from an identical protein chain. This can be as large as a complete tertiary domain or as small as an α-helix or β-strand. Domain swapping results as intertwined higher oligomers and serves as an efficient and thermodynamically favourable means of inter-converting between lower and higher oligomeric assemblies. This implies that there are two (or more) energetically stable and favourable assemblies a protein can adopt somewhat amending Anfinsen’s early thermodynamic theory.

3D domain swapping has been studied in many proteins and is often accompanied by a change in the protein’s functionality [183]. To date there are fewer than 60 structures of domain-swapped proteins. Although the interfaces formed by the lower and higher oligomers are largely the same, this switch allows for co-operativity between proteins or formation of active sites at the interface. Variations such as changes in pH, ionic concentration, redox state, temperature or the presence of co-factors can favour one oligomeric state over another and lead to domain swapping [184,185]. Understanding how domain swapping is regulated is important as it can shed light on the biological role it plays.

Domain swap can result in the formation of a new active site by (i) creation of a new interface (between two RNase A dimers) or (ii) inducing conformational changes in distant catalytic domains (α-isopropylmalate synthase (α-IPMS) dimerisation). Moreover it can allow docking of partner proteins by (iii) changing the dynamic properties of binding regions (phosphorylated substrates preferentially binding to dimers of the sucrose transport protein SUC1) [183]. Finally linear and ‘open ended’ domain swapping (iv) mediates formation of amyloid fibrils and plays a major role in pathologies associated with protein aggregation [186].

Based on the structural (and biochemical) data discussed in this chapter, a biological role for the intertwining of PERK (and probably IRE1) luminal domains via Interface 2 is proposed. Although strictly speaking this is not ‘domain swapping’, it is analogous to this phenomenon. Different models in which dimer and tetramer assemblies represent differently activated states can be envisaged (Figure 6-15). In a possible scenario, dimers and tetramers represent
inactive and active oligomers respectively. Alternatively, both conformations might mediate UPR signalling but activate downstream pathways differently.

Although more experiments will be key to elucidate the precise mechanism, speculations on how α-helix ‘swapping’ in the luminal domain regions may convey a divergence in PERK and IRE1 cytoplasmic signalling can be made. Firstly, it has been well established that the cytosolic regions of PERK and IRE1 must autophosphorylate in order to initiate eIF2α phosphorylation and xbp1 mRNA splicing respectively. Tetramerisation could allow for cooperativity of the cytoplasmic domains and increase/alter downstream signalling. Secondly, tetramer formation may provide new surfaces for different co-factors to bind and as such lead to different physiological effects. This could include activation of Nrf2 and TRAF2 pathways downstream of PERK and IRE1 respectively.

Future experiments will be key to (i) validate PERK and IRE1 oligomeric state-dependent signalling in vivo and (ii) appreciate its role in UPR signalling. This will include testing the effect of mutations which compromise dimerisation and tetramerisation in PERK and IRE1 activation. The use of full-length proteins and in vivo cellular experiments will be critical. Both levels of phosphorylation of their cytoplasmic domains and quantification of signalling (e.g. eIF2α phosphorylation and GADD34 expression for PERK; xbp1 mRNA splicing and TRAF2 binding for IRE1) will be important to fully understand the importance of dimers/tetramers. This will shed light on which oligomeric state represents their inactive and active conformation. Additionally it will be possible to appreciate if these are responsible for detecting differing conditions/duration/strength of ER stress and tuning downstream pathways accordingly. Together these studies will have great implications in the understanding, and manipulation, of ER stress sensing and UPR signalling.
In physiological conditions, PERK exists as an inactive monomer. In this arrangement, helix α2 is unstructured or flexible. In the presence of ER stress, two PERK monomers come together and form dimers capable of UPR signalling. Elevated levels or prolonged ER stress could two PERK dimers could come together via α2 helix-Lobe2 interaction to form tetramers. α2 is stabilized in the interaction with an adjacent dimer and adopts its helical structure. This could allow for increased co-operativity of the cytoplasmic domains to increase eIF2α phosphorylation. Alternatively this could activate differential signalling such as Nrf2 activation.
7 Biophysical characterisation of BiP interaction with *H. sapiens* PERK and IRE1 luminal domains
7.1 Summary

This chapter presents the biophysical characterisation of the interaction between BiP chaperone and the luminal domains of *H. sapiens* PERK and hIRE1 *in vitro*. A range of biophysical techniques including MST, ITC and pull-down experiments were used.

The binding of BiP to the luminal domains is shown to occur *via* BiP’s NBD and luminal domain regions II-IV; with low micromolar affinity. This represents an unconventional signalling interaction that has never been previously conclusively demonstrated nor characterised. Furthermore, unlike chaperone-substrate interaction, this interaction is independent of nucleotide binding, reinforcing the notion that this is a UPR signalling interaction distinct from BiP’s chaperone function.

In order to recapitulate conditions of ER stress, and as such probe the mechanisms of BiP-dependent UPR activation, the BiP-luminal domain protein complexes were also analysed in the presence of the unfolded protein mimic ∆EspP and the authentic unfolded protein CH₁. Although binding of hPERK and hIRE1 luminal domains to ∆EspP was observed, this peptide had no discernable effect on their oligomeric state nor on BiP-luminal domain complexes. In contrast, CH₁ bound specifically to BiP’s SBD; this is a conventional BiP-unfolded substrate interaction. The addition of CH₁ not only prevented the formation of BiP-luminal domain complexes but was also capable of disrupting pre-assembled complexes. This indicated that the ‘conventional’ binding of unfolded proteins to BiP SBD is transduced to the ‘unconventional’ BiP NBD-luminal domain UPR signalling interaction.

This study elegantly illustrates how the accumulation of unfolded proteins during ER stress leads to UPR activation *via* PERK and IRE1, with BiP chaperone playing a major role in this regulation.
7.2 Introduction

The correlation between overexpression of BiP and attenuation of UPR signalling has long been established [79,80]. Furthermore, many studies dating to over a decade ago, provide strong evidence for the interaction between BiP and IRE1, PERK and ATF6 [76,81,82]. In these studies, immunoprecipitation experiments using whole cell extracts were carried out to study the protein complexes. These cellular studies have provided the principal basis for our understanding of protein complexes involving UPR components. However, the direct interaction between BiP and IRE1, PERK and ATF6 has never actually been demonstrated. Importantly, in part due to the experimental set up, these studies have been unsuccessful in discerning the mechanistic events that take place in the ER during conditions of stress to lead to downstream UPR signalling.

Several hypotheses of how the presence of unfolded proteins in the ER lumen lead to activation of the UPR mediators have been proposed [77]. Initially, a competition model was proposed where BiP binds to the luminal domains of IRE1, PERK and ATF6 to repress UPR activation. The accumulation of unfolded protein within the ER would cause for BiP to be titrated off the luminal domains and thus releasing the inhibition. More recently, a BiP-independent model whereby the direct binding of unfolded proteins to the luminal domains of the ER stress sensors causes their oligomerization resulting in UPR activation. This mechanism is backed up by the X-ray structure of S. cerevisiae IRE1p luminal domain; this model has not yet been linked to the mammalian proteins [95]. In fact, based on the X-ray crystal structure of H. sapiens IRE1, although similar to S. cerevisiae IRE1p’s, it was suggested that this mechanism is incompatible with the mammalian protein [101].

Clearly, a detailed and conclusive mechanistic understanding of ER stress sensing and UPR activation remains to be established.
7.3 Direct binding of BiP to the luminal domains of hPERK and hIRE1

7.3.1 MST experiment setup

The interaction of the complete hPERK (I-V) and hIRE1 (I-V) luminal domain proteins with BiP was initially characterised using MST. Proteins were labelled with an amine-reactive red fluorescent dye. Preliminary experiments to ensure the suitability of MST experimentation with the purified proteins were carried out.

Initially, labelled BiP FL was titrated with 16 two-fold dilutions of unlabelled hPERK (I-V), samples were loaded into standard grade capillaries and a ‘resting state’ capillary scan was carried out (Figure 7-1A). Firstly, the peaks obtained are symmetric, indicating that BiP FL is uniformly dispersed in the capillaries. Secondly, readings above 200 absorbance units, the minimal sensitivity of the instrument, were obtained. This confirms that the dilution used BiP FL produces sufficient signal above noise levels. Thirdly, the height of the peaks is constant and independent of the concentration of hPERK (I-V) in the samples. This implies that BiP FL fluorescence is not quenched by hPERK (I-V), and that hPERK (I-V) does not cause BiP FL to be irregularly distributed in the capillary. These factors were similarly checked for each MST experiment.

Typical raw curves obtained for an MST binding experiment are shown in Figure 7-1 Bi. The curves are smooth throughout the course of the experiment. This indicates that there was no sample aggregation and that the protein sample, buffer and experiment parameters used are suitable. The curves were then normalized to take into account differences in their ‘resting state’ initial fluorescence ($F_{\text{cold}}$). $F_{\text{hot}}/F_{\text{cold}}$ values were calculated for each titration point and plotted against the concentration of the titrant, in a logarithmic scale (Figure 7-1Bii).
Figure 7-1. Typical MST binding experiment

(A) Capillary scan of labelled BiP FL titrated with unlabelled hPERK (I-V). 16 two-fold dilutions from highest concentration of 500 μM were used. Samples were loaded into standard grade capillaries and scanned using a Monolith NT-115 instrument (20% LED; 80% IR-laser; 25 °C). X-axis position values refer to the physical distance along the tray in which the capillaries are aligned; capillary numbers (Cap.) are shown. The measured raw fluorescence is given in absolute absorbance units. The peaks are symmetrical and show little variance in shape and height throughout the titration.

(B) MST curves for labelled BiP FL titrated with unlabelled hPERK (I-V). After 6" equilibration, the LED was switched on (100% power) for 30". After the LED laser was switched off, the capillary fluorescence was measured for a further 6". The range of data used for \( F_{\text{cold}} \) (blue) and \( F_{\text{hot}} \) (red) values are shown. The (i) raw and (ii) normalized fluorescence curves are shown.
Initially, two laser powers (80 and 100% LED) were tested for the titration of unlabelled hPERK (I-V) and hIRE1 (I-V) into labelled BiP FL (Figure 7-2A). The $K_D$ values obtained when using 80% LED ($K_D = 3.76 \, \mu M$) and 100% LED ($K_D = 2.22 \, \mu M$) are similar. The fit of the data to the Boltzmann equation used to determine affinity values was determined (R square = 0.74 for 80% LED curve and 0.99 for 100% LED curve). Using 100% LED power gave a better fit of the data. The broader temperature gradient induced by the higher LED power probably provokes a more pronounced difference between the thermophoresis of BiP FL alone versus BiP FL bound to hPERK (I-V). 100% LED was used for all subsequent experiments.

Furthermore, experiments were carried out using both permutations of labelled and unlabelled partners (labelled BiP FL proteins titrated with unlabelled luminal domains and vice versa). This ensured the results obtained were specific and reproducible. Figure 7-2B shows MST curves for labelled BiP FL titrated with unlabelled hPERK (I-V) ($K_D = 2.37 \, \mu M$) and the reverse, labelled hPERK (I-V) titrated with unlabelled BiP FL ($K_D = 1.99 \, \mu M$). The affinity values obtained are almost identical. The first setup (labelled BiP proteins titrated with unlabelled hPERK or hIRE1 proteins) was chosen for all subsequent experiments. This combination provided less variability between experiments allowing for more reliable comparisons. This was especially of importance in experiments screening for different luminal domain constructs binding to the same BiP protein but also for comparison of identical experiments carried out with either hPERK or hIRE1.
Figure 7-2. Optimization of MST experiments

(A)Labelled BiP FL titrated with unlabelled hPERK (I-V) using 80% (green) or 100% (blue) LED to induce thermophoresis. $K_D$ values obtained are similar. $R$ square values of the fit of the data to the Boltzmann equation are shown. Using 100% LED yields a curve that better fits the data.

(B)Interaction between hPERK (I-V) and BiP FL measured by alternating the labelled and unlabelled species. $K_D$ values obtained show no difference between the two experimental setups.
7.3.2 hPERK (I-V) and hIRE1 (I-V) binding to BiP sub-domains

Although many studies have highlighted a function of BiP in the activation/deactivation of the UPR sensor proteins, their direct association has never been demonstrated. Initially, BiP FL binding to hPERK (I-V) and hIRE1 (I-V) was measured using MST. BiP FL binds with similar low micromolar $K_D$ to hPERK (I-V) ($K_D = 1.92 \mu M$) and hIRE1 (I-V) ($K_D = 1.33 \mu M$) (Figure 7-3A). To identify the domain of BiP to which the luminal domain proteins bind to, their association with BiP NBD and BiP SBD was explored. MST measurements show that BiP NBD binds specifically to hPERK (I-V) ($K_D = 2.05 \mu M$) and hIRE1 (I-V) ($K_D = 1.97 \mu M$) (Figure 7-3B). No binding to BiP SBD was observed. This is the first time the direct interaction between BiP and the luminal domain proteins has been measured and categorically mapped to BiP’s NBD.

7.3.3 Influence of sub-regions I to V in hPERK and hIRE1 interaction with BiP NBD

The luminal domains of PERK and IRE1 have been previously divided into five sub-regions (I, II, III, IV and V) [85-87]. The functional relevance of these in binding to BiP and UPR signalling has been investigated. However so far, published studies have not given consistent and definitive results. To assign the importance of the luminal domain sub-regions I-V in binding to BiP NBD, affinities were compared using the different luminal domain constructs (I-V, I-IV, II-V and II-IV) (Figure 7-4). Similar $K_D$ values of approximately 2 $\mu M$ were obtained throughout indicating that regions II-IV, the core region of the luminal domains, are necessary and sufficient for hPERK and hIRE1 binding to BiP NBD.
Figure 7-3. Interaction of hPERK and hIRE1 luminal domains with BiP sub-domains

(A) MST binding curves of labelled BiP FL titrated with unlabelled (i) hPERK (I-V) or (ii) hIRE1 (I-V). Similar $K_D$ values in the low micromolar range are obtained.

(B) MST binding curves for labelled BiP NBD (magenta) and SBD (teal) binding to (i) hPERK (I-V) or (ii) hIRE1 (I-V). BiP NBD binds to the luminal domain proteins with similar affinity as BiP FL. No binding is observed to BiP’s SBD. SE values were calculated from $K_D$ of three independent binding experiments.
Figure 7-4. Defining the core interaction of hPERK and hIRE1 with BiP NBD

(A) Schematic representation of hPERK and hIRE1 luminal domain region I to V truncation constructs used.

(B) MST binding curves of labelled BiP NBD titrated with various length constructs of unlabelled (i) hPERK and (ii) hIRE1 luminal domains.

(C) List of $K_D$ values (μM ±SE) for MST binding curves in 'B' of BiP NDB interacting with hPERK and hIRE1 luminal domain constructs. The regions comprised by each construct are indicated. All affinity values obtained are similar. Regions II-IV of the luminal domains are necessary and sufficient for binding to BiP NBD.

SE values were calculated from $K_D$ of three independent binding experiments.
7.3.4 ITC of BiP FL binding to hPERK (II-IV) and hIRE1 (II-IV)

ITC was used to verify MST results as well as to gain additional information on the stoichiometry of the complexes. hPERK (II-IV) and hIRE1 (II-IV) luminal domain constructs were chosen since they bind to BiP identically as region I-V complete luminal domain constructs but are most stable and express more abundantly. For these same reasons, region II-IV constructs were used hereon for all subsequent experiments.

Firstly, the heat of dilution of BiP FL was calculated by titrating BiP FL, at identical concentration as was used for the titration experiments, into the cell containing buffer only (Appendix Figure 10-11). The peaks were integrated and a straight line was fitted through the data. This represents BiP FL heat of dilution (5.8 kcal/mol). For binding experiments, BiP FL was titrated into hPERK (II-IV) and hIRE1 (II-IV) separately. The experimental data was integrated and, after subtraction of the heat of BiP FL dilution values, they were fitted with a one site binding model (Figure 7-5). BiP FL binds to hPERK (II-IV) ($K_D = 6.00 \, \mu M$) and hIRE1 (II-IV) ($K_D = 1.36 \, \mu M$). These values are within the same range as those obtained by MST ($K_D = 1.92$ and $1.33 \, \mu M$ respectively). In both cases, the data yielded a 1:1 binding stoichiometry. The significance of this will be discussed further in Chapter 7.3.7.2.
**Figure 7-5. ITC of BiP FL binding to hPERK (II-IV) and hIRE1 (II-IV)**

(A-B) BiP FL (500 μM) was titrated into (A) hPERK (II-IV) and (B) hIRE1 (II-IV) (55 μM) (VP-ITC MicroCalorimeter). BiP FL heat of dilution (5.8 kcal/mol) was subtracted from the integrated experimental data and a one-binding site model was fitted. The association constant ($K_A$), changes in enthalpy ($\Delta H$) and entropy ($\Delta S$) and stoichiometry ($N$) values obtained are shown. Dissociation constants ($K_D$) were calculated. Both hPERK (II-IV) and hIRE1 (II-IV) bind to BiP FL with affinities in the same range as those obtained by MST experiments.
7.3.5 Effect of nucleotides on BiP FL-luminal domains binding affinity

Since BiP is a member of the Hsp70 family and its ATPase activity is directly linked to its chaperoning function and conformational cycling, the effect of nucleotide binding and hydrolysis was investigated.

Firstly, to ensure recombinant BiP possesses its functional ability to bind nucleotides, DSF experiments were carried out to measure changes in its thermal stability upon binding different nucleotides. The normalized thermal denaturing curves indicate two separate unfolding events are taking place (Figure 7-6Ai). This is suggestive of the presence of two domains, which have very different thermal stabilities. In the case of BiP, the two melting transitions curves are likely to represent the sequential denaturation of BiP NBD and SBD. Only the first denaturing event shows differences depending on the nucleotide present. Since nucleotides bind only to the NBD, these first melting curves are likely to represent unfolding of BiP NBD. In any case, since all that was of interest in this experiment was to measure BiP binding to nucleotides, it is not important to know which domain is unfolding.

Maximum and minimum values for the first transition curves were calculated and fitted with a Boltzmann equation (Figure 7-6Ai/Aii). BiP FL Tm values were obtained from the inflection points of the curves. Significant differences between the apo (Tm = 46.8 °C), AMPPNP (Tm = 50.0 °C), and ADP (Tm = 55.9 °C) bound states were observed (Figure 7-6Aii). Binding of AMPPNP and ADP incrementally stabilizes BiP FL. This is consistent with the conformational changes known to be induced by AMPPNP and ADP binding to Hsp70s. More specifically, they promote a closed NBD/open SBD and open NBD/closed SBD states respectively. In the ATP bound state (Tm = 55.2 °C) BiP behaves as in the ADP bound state (Tm = 55.9 °C). This could be due to BiP’s ATPase activity or spontaneous degradation of ATP during the experiment, both of which would result in ATP hydrolysis into ADP. In either case, these results demonstrate BiP’s ability to bind nucleotides and undergo conformational rearrangements upon doing so.
Figure 7-6. Thermal stability of BiP FL upon nucleotide binding

(A) DSF was used to measure BiP FL Tm in the presence of 5 mM ATP, ADP, AMPPNP or no nucleotide (+10 mM MgCl$_2$). (i) Normalized denaturing curves obtained show two melting transition curves (1 and 2). Only transition curve 1 is affected by the presence of different nucleotides. (ii) The maximum and minimum values of transition curves 1 were calculated and fitted with Boltzmann equations. Inflection points were used to obtain Tm values.

(B) Tm values obtained from the experiments in ‘B’ are listed. ATP and ADP bound Tm are very similar suggesting ATP is being hydrolysed into ADP. Nucleotide binding increases BiP FL stability, with highest stability achieved in the ADP-bound states.
Next, the affinity of interaction of BiP FL binding to hPERK (II-IV) and IRE1 (II-IV) in the presence of the different nucleotides was measured by MST (Figure 7-7). Affinity values obtained are comparable to those obtained in the apo state. These measurements demonstrate that the direct association of BiP FL with the luminal domain proteins is not dependent on, or at least not greatly affected by, nucleotide binding.

7.3.6 BiP FL-luminal domain proteins pull down experiments

Pull-down experiments were used (i) to confirm MST and ITC results and (ii) to have a tool to visualize protein complexes. Negative controls where the His₆-tagged species were replaced by buffer were included to verify that the untagged proteins did not bind to the TALON resin. BiP chaperone has a proneness to bind charged molecules and surfaces. Experiments in which His₆-luminal domain proteins captured untagged BiP were unsuitable. When this set up was tested, untagged BiP bound to TALON resin in negative control experiments (results not shown). On the other hand, the untagged hPERK and hIRE1 proteins tested did not bind to TALON resin (Figure 7-8 lanes 1-4; Figure 7-9 lanes 1-2). His₆-BiP proteins were therefore chosen as the bait to capture the untagged luminal domain proteins.

When designing the pull down experiment, care was taken to allow for capturing of transient and low affinity complexes. Firstly, 10-fold excess of hPERK and hIRE1 proteins were added onto immobilized His₆-BiP proteins to ensure concentrations exceeded not only $K_D$ values but also reached levels where most of the complex would be formed. Secondly, multiple small volume wash steps rather than fewer larger volume ones were carried out. As such, it was ensured that the protein concentrations did not fall below 10 μM. At 10 μM protein concentrations, proteins that interact with 1 μM $K_D$ can be estimated to be roughly 75% bound.
Figure 7-7. Effect of nucleotide in BiP FL binding to hPERK and hIRE1

(A) labelled BiP FL was titrated with unlabelled (i) hPERK (II-IV) and (ii)hIRE1(II-IV) in the presence of 5 mM ATP, ADP, AMPPNP or no nucleotide and 10 mM MgCl$_2$.

(B) List of $K_d$ values (μM ±SE) for MST binding curves in A and B of BiP FL interacting with hPERK (II-IV) and hIRE1 (II-IV), respectively, in the presence of ATP, ADP or AMPPNP nucleotides. All affinity values obtained are similar.

BiP FL interaction with the luminal domain proteins is not dependent on nor affected by its binding to nucleotides.
In a first instance, the binding of the luminal domain complete (region I-V) and core (region II-IV) constructs to BiP FL was tested (Figure 7-8). BiP FL binds to both length constructs of hPERK (lanes 6 and 7) and hIRE1 (lanes 8 and 9) luminal domains. Since region II-IV constructs are much more stable, and the data obtained throughout this study indicates that it behaves identically to region I-V constructs, at least with respect to interaction with BiP, they were used for subsequent pull down experiments.

Using an identical setup, binding of hPERK (II-IV) and hIRE1 (II-IV) to the different BiP subdomains was tested (Figure 7-9). hPERK (II-IV) and hIRE1 (II-IV) bind specifically to BiP FL (lanes 4 and 5) and its NBD (lanes 7 and 8) but not to BiP SBD (lanes 10 and 11). These results are consistent with those previously obtained by MST.
<table>
<thead>
<tr>
<th>MW</th>
<th>Control</th>
<th>Elution</th>
</tr>
</thead>
<tbody>
<tr>
<td>His$_6$-BIP FL</td>
<td>- - - -</td>
<td>+ + + + +</td>
</tr>
<tr>
<td>hPERK (I-V)</td>
<td>+ - - -</td>
<td>- + - - -</td>
</tr>
<tr>
<td>hPERK (II-IV)</td>
<td>- + - -</td>
<td>- - + - -</td>
</tr>
<tr>
<td>hIRE1 (I-V)</td>
<td>- - + -</td>
<td>- - - + -</td>
</tr>
<tr>
<td>hIRE1 (II-IV)</td>
<td>- - - +</td>
<td>- - - - +</td>
</tr>
</tbody>
</table>

Figure 7-8. His$_6$-BIP FL pull down of hPERK and hIRE1 luminal domains

Pull down assay using His$_6$-BIP FL and untagged hPERK (lanes 6-7) and hIRE1 (lanes 8-9) constructs spanning regions I-V (complete) and II-IV (core) of their luminal domains. Both the complete and core luminal domain constructs bind to His$_6$-BIP FL. In control experiments the luminal domains only were incubated with TALON resin (lanes 1-4). The MW (kDa) is indicated.
Figure 7-9. His<sub>6</sub>-BiP sub-domains pull down of hPERK (II-IV) and hIRE1 (II-IV)

Pull down assay using His<sub>6</sub>-BiP FL (lanes 3-5), His<sub>6</sub>-BiP NBD (lanes 6-8) and His<sub>6</sub>-BiP SBD (lanes 9-11) with untagged hPERK (II-IV) and untagged hIRE1 (II-IV). hPERK (II-IV) and hIRE1 (II-IV) bind to His<sub>6</sub>-BiP FL and His<sub>6</sub>-BiP NBD. No binding is observed to His<sub>6</sub>-BiP SBD. In control experiments the luminal domains only were incubated with TALON resin (lanes 1-2). The MW (kDa) is indicated.
7.3.7 Stoichiometry of hPERK (II-IV)-BiP complexes

To investigate the stoichiometry of hPERK (II-IV) interaction with BiP, native gel shift assays, chemical crosslinking, SEC, AUC and ITC were carried out. This thesis is a study mainly on hPERK protein and thus only the stoichiometry of hPERK-BiP complexes was analysed. For all experiments, proteins were mixed in equimolar amounts at concentrations at least 10-fold above $K_D$ values to ensure sufficient complex was formed.

7.3.7.1 Gel shift assays

To identify the size of the complex, gel shift assays were initially carried out. These included native gels and chemical-crosslinking. hPERK (II-IV) was incubated with BiP FL and run on a native gel (Appendix Figure 10-12A). However, all bands in the sample containing both hPERK (II-IV) and BiP FL could be attributed to species also visible when the proteins were run in isolation. This could be due to the complex being transient. Chemical-cross linking was then used as a means of capturing and stabilizing protein-protein interactions. Experiments were prepared similarly as for native gel shift assays but a 25-fold molar excess of EGS was added (Appendix Figure 10-12B). The MW of BiP FL ($\sim$65 kDa) is a multiple of the MW of hPERK (II-IV) ($\sim$32 kDa). Because of this, it was difficult to differentiate between bends belonging to protein complexes or those belonging to higher oligomers of either species. At least as visible by Coomassie staining, no band corresponding to hPERK (II-IV)-BiP complex could be discerned; all the visible bands could be assigned to specific oligomers of the two proteins.

7.3.7.2 SEC MALS

BiP FL-hPERK (II-IV) and BiP NBD-hPERK (II-IV) complexes were analysed by SEC MALS (Figure 7-10). When run alone hPERK (II-IV), BiP FL and BiP NBD alone elute as homogenous 70 kDa dimers, 139 kDa dimers and 37 kDa monomers respectively. As shown by the co-elution traces, hPERK (II-IV) and BiP proteins come off separately on the SEC column and no additional complex peak is observed. Again, this could be due to the complex not being strong enough to tolerate diffusion through the Superdex resin. Alternatively, the proteins could have been diluted to below $K_D$ concentrations in the column and as such no, or
insufficient, complex is formed. Given their low affinity values this is not an unexpected result.

7.3.7.1 AUC

Consequently, AUC was used as a more gentle approach to capture BiP FL-hPERK (II-IV) complexes (Figure 7-11). Results show that in solution hPERK (II-IV) exists as a dimer (70 kDa) and tetramer (140 kDa); BiP FL exists as a monomer (~68 kDa) and dimer (140 kDa). In the profile of BiP FL-hPERK (II-IV) complex samples, large peaks corresponding to the two individual species are clear. An additional smaller peak of 220 kDa was also detected. This could correspond to the BiP FL-hPERK (II-IV) complex with a 2:2 ratio. However, due to the poor resolution of the peaks, a 1:1 complex stoichiometry cannot be excluded. Such a species would be of approximately 110 kDa and would be masked by the poorly defined BiP FL peak profiles. A 290 kDa species was also observed. This could represent a number of hPERK (II-IV) to BiP FL complex ratios (2:3, 4:2, 6:1); it is not possible to discern between the different stoichiometries from this data. Additionally, in proportion to the other peaks this is very small and is likely to be due to aggregated protein rather than a biologically relevant oligomer.
Figure 7-10. SEC MALS of hPERK (II-IV)-BiP complexes
(A-B) SEC MALS analysis (Superdex 200 PC 3.2/30) of hPERK (II-IV) (100 μM) and (A) BiP FL (100 μM) or (B) BiP NBD (100 μM). The calculated MW (kDa) of the peaks obtained are indicated. In either experiment, no peak corresponding to BiP-hPERK (II-IV) complexes could be discerned.
Figure 7-11. AUC of hPERK (II-IV)-BiP FL complex

AUC analysis of hPERK (II-IV) (11 μM) and BiP FL (11 μM). The c(s) distribution is plotted versus the sedimentation coefficient (S) in Svedberg unit to show the distribution of the different species according to their molar mass. The expected MW (kDa) corresponding to sedimentation coefficient (S) values are shown (grey). Alone, hPERK (II-IV) exists as a dimer and tetramer whereas BiP FL forms monomers and dimers. When hPERK (II-IV) and BiP FL are run together, the profile is poorly resolved. Additionally to the peaks which are also present when the two proteins are run in isolation, a shoulder peak corresponding to approximately 290 kDa is present. AUC experiments and data analysis was carried out by Dr K. Stott.
Finally, ITC was used to analyse the stoichiometry of BiP FL-hPERK (II-IV) complexes. As discussed in Chapter 7.3.4, titration of BiP FL into hPERK (II-IV) yields a sigmoidal curve that fits a one-site of binding. The two proteins interact with 6 µM $K_D$ in an equimolar ratio. A stoichiometry of $N = 1$ in ITC cannot be used directly to distinguish between a ratio of 1:1, 2:2, 3:3 and so on. As measured by SEC MALS, hPERK (II-IV) is a dimer in solution, at least at the 55 µM working concentration of the ITC experiment. If hPERK (II-IV) dimers were to break into monomeric species upon titration of BiP FL, an event necessary to yield a 1:1 stoichiometry, it would contribute to the overall enthalpy of the reaction. In such a scenario, it would probably difficult to fit a one site of binding curve, as this is usually reflective of a reaction consisting of a single event. Although this is a first indication of a 2:2 (as opposed to a 1:1) hPERK (II-IV) to BiP FL stoichiometry, it is not unambiguous evidence.

To further address the stoichiometry, and to gain further insight on this interaction, BiP FL was titrated into hPERK (II-IV) L200P monomeric mutant (Figure 7-12). After subtraction of BiP FL heat of dilution to the integrated data, a one-site of binding curve was fitted. The data gave a 5.92 µM $K_D$ and 1:1 stoichiometry. The affinity value obtained is identical to that of binding to wild-type dimeric hPERK (II-IV). In addition, the overall enthalpy of the two reactions (BiP binding to hPERK (II-IV) monomers or dimers) are identical. This strongly suggests that the same reactions are taking place. It can be extrapolated that hPERK (II-IV) wild-type and L200P mutant remain dimers and monomers respectively upon binding of BiP. Together, the ITC data suggest that (i) BiP does not rely the dimerisation groove of hPERK luminal domain for binding and (ii) BiP-hPERK (II-IV) complexes comprise of a 2:2 ratio.
Figure 7.12. Titration of hPERK (II-IV) L200P with BiP FL

BiP FL (500 μM) was titrated into hPERK (II-IV) L200P (55 μM) (VP-ITC MicroCalorimeter). BiP FL heat of dilution (5.8 kcal/mol) was subtracted from the integrated experimental data and a one-binding site model was fitted. The association constant ($K_A$), changes in enthalpy ($\Delta H$) and entropy ($\Delta S$) and stoichiometry (N) values obtained are shown. Dissociation constants ($K_D$) were calculated. At least with respect to affinity and enthalpy, hPERK (II-IV) L200P binds to BiP FL almost identically as wild-type hPERK (II-IV).

$K_A = 1.69 \pm 5 \mu M$
$\Delta H = -3114 \text{ kcal mole}^{-1}$
$\Delta S = 13.5 \text{ cal mole}^{-1} \text{ K}^{-1}$

$N = 1.19$
$K_D = 5.92 \mu M$
7.4 Investigating UPR activation using model unfolded proteins

7.4.1 ∆EspP unfolded protein mimic

7.4.1.1 ∆EspP binds specifically to hPERK (II-IV) and hIRE1 (II-IV) luminal domains

As a preliminary screen for ∆EspP binding to hPERK and hIRE1 proteins, DSF was carried out at first (Figure 7-13). ∆EspP decreases the thermal stability of hPERK (II-IV) and hIRE1 (II-IV) by 8.5 and 4.4 °C respectively. Although DSF demonstrates binding, it gives no indication of affinity values nor stoichiometry. MST relies on differences in size, shape and charge of labelled proteins upon binding to an unlabelled partner, which changes its thermophoretic properties. Association of ∆EspP peptide, with a mass less than 3 kDa, to macromolecules larger than 70 kDa is likely to not result in sufficient changes to the hydration shell of the latter to be measurable by MST. Therefore, MST was not deemed suitable for studying ∆EspP binding to hPERK and hIRE1 luminal domains. ITC was employed instead.

Firstly, ∆EspP heat of dilution was measured (<0.1 kcal/mol) (Appendix Figure 10-11B). This is negligible and thus was not taken into account in the titration runs. ∆EspP was titrated into hPERK (II-IV) and hIRE1 (II-IV) proteins separately. The data were integrated and fitted to a one-site binding model (Figure 7-14). ∆EspP binds to the luminal domain proteins with 9.4 µM and 6.4 µM $K_D$ respectively. A 1:2 peptide to protein ratio was obtained in both cases indicating that one peptide molecule binds to each protein dimer.
Figure 7-13. ΔEspP decreases the thermal stability of hPERK (II-IV) and hIRE1 (II-IV)

(A-B) DSF was used to test for ΔEspP peptide (MKKHKLALCFLGGLQSSYSAAKKKK) binding to (A) hPERK (II-IV) and (B) hIRE1 (II-IV). Tm values indicate that binding of ΔEspP significantly decreases the Tm of hPERK (II-IV) (by 8.5 °C) and hIRE1 (II-IV) (by 4.4 °C).
Figure 7-14. ΔEspP binding to hPERK (II-IV) and hiRE1 (II-IV)

(A-B) ΔEspP (500 μM) was titrated into (A) hPERK (II-IV) or (B) hiRE1 (II-IV) (55 μM) (VP-ITC MicroCalorimeter). The data was integrated and fitted to a one-binding site model. The association constant (K_A), changes in enthalpy (Δ H) and entropy (Δ S) and stoichiometry (N) values obtained are shown. Dissociation constants (K_D) were calculated. ΔEspP binds similarly to hPERK (II-IV) and hiRE1 (II-IV) with micromolar affinity and a 1:2 ΔEspP to luminal domain proteins ratio.

\[ K_A = 1.07 \times 10^{-5} \text{ μM} \]
\[ Δ H = -522 \text{ kcal mole}^{-1} \]
\[ Δ S = 21.3 \text{ cal mole}^{-1} K^{-1} \]

\[ N = 1.98 \]
\[ K_D = 9.35 \text{ μM} \]

\[ K_A = 1.56 \times 10^{-5} \text{ μM} \]
\[ Δ H = -827 \text{ kcal mole}^{-1} \]
\[ Δ S = 21.0 \text{ cal mole}^{-1} K^{-1} \]

\[ N = 1.97 \]
\[ K_D = 6.41 \text{ μM} \]
To verify the specificity of ∆EspP for the luminal domain proteins, similar experiments were carried out using BiP FL. In DSF, ∆EspP only reduced the Tm of BiP FL by 0.4 °C (Appendix Figure 10-13A). This value is within experimental error values and therefore not significant. In ITC, no heat changes were measured upon titration of ∆EspP into BiP FL (Appendix Figure 10-13B). Together these results demonstrate the specificity of ∆EspP for the luminal domain proteins with no binding to BiP FL observed.

7.4.1.2 ∆EspP does not induce clustering of hPERK (II-IV) and hIRE1 (II-IV)

∆EspP peptide has been shown to induce formation of higher oligomeric species upon binding to *S. cerevisiae* IRE1p orthologue. This ∆EspP-induced clustering effect was tested on hPERK and hIRE1 luminal domains. hPERK (II-IV) and hIRE1 (II-IV) were incubated with ∆EspP and samples were analysed by SEC MALS (Appendix Figure 10-14). hPERK (II-IV) and hIRE1 (II-IV) remain ∼68 and ∼104 kDa dimers respectively. ∆EspP does not have an effect on the oligomeric state of the luminal domain proteins, at least as measured by SEC MALS.

Given that ∆EspP does bind to the luminal domain proteins, its effect on complexes with BiP FL was tested. Competition pull down experiments were carried out similarly to those with Cn1. Briefly, hPERK (II-IV)-BiP FL and hIRE1 (II-IV)-BiP FL complexes were formed and ∆EspP was added (Appendix Figure 10-15). This had no visible effect on protein complexes suggesting ∆EspP has no function in the mechanism of UPR activation by hPERK and hIRE1.
7.4.2  C\textsubscript{H}1 authentic unfolded protein

7.4.2.1  C\textsubscript{H}1 is a BiP-specific unfolded protein mimic

C\textsubscript{H}1 is a known BiP substrate and has been extensively used to study its chaperoning function. MST was used to investigate C\textsubscript{H}1 binding to BiP, hPERK and hIRE1 proteins (Figure 7-15A). C\textsubscript{H}1 binds specifically, and similarly, to BiP FL and to its SBD in the presence of ADP ($K_\text{D} = 8.67$ and $5.11 \, \mu\text{M}$ respectively). No binding to BiP NBD, hPERK (II-IV) nor hIRE1 (II-IV) was observed. These findings were confirmed in pull-down experiments (Figure 7-15B). His\textsubscript{6}-BiP FL and His\textsubscript{6}-BiP SBD were able to pull down untagged C\textsubscript{H}1. No binding of untagged C\textsubscript{H}1 to His\textsubscript{6}-BiP NBD, His\textsubscript{6}-hPERK (II-IV) or His\textsubscript{6}-hIRE1 (II-IV) was observed. These results confirm C\textsubscript{H}1 as a BiP SBD-specific unfolded protein mimic.

7.4.2.2  C\textsubscript{H}1 prevents formation, and can promote dissociation, of luminal domain-BiP complexes

C\textsubscript{H}1 was exploited as an authentic unfolded protein to probe the interactions between the luminal domains and BiP in pull-down experiments. Firstly, His\textsubscript{6}-BiP FL-C\textsubscript{H}1 complexes were formed and a 10-fold molar excess of untagged hPERK (II-IV) or untagged hIRE1 (II-IV) was added. As previously described, the luminal domain proteins normally bind to BiP FL in identical pull-down experiments. However, shows that if BiP FL is already bound to C\textsubscript{H}1, hPERK (II-IV) (lane 5) and hIRE1 (II-IV) (lane 7) do not bind to BiP FL. This was a first indication that C\textsubscript{H}1 unfolded protein has an inhibitory effect on BiP’s association with the luminal domain proteins.

Next, the reverse experiment was carried out to test whether, in addition to inhibiting luminal domain association with BiP, C\textsubscript{H}1 was also able to disrupt already formed complexes. His\textsubscript{6}-BiP FL was complexed with untagged hPERK (II-IV) and untagged hIRE1 (II-IV) and a 10-fold molar excess of untagged C\textsubscript{H}1 was added (Figure 7-17). hPERK (II-IV) (lane 5) and hIRE1 (II-IV) (lane 7) dissociate from BiP FL upon addition of C\textsubscript{H}1. This result demonstrates that C\textsubscript{H}1 can actively promote dissociation of BiP FL-luminal domain complexes.
Figure 7-15. Binding of Cn1 to BiP and luminal domain proteins

(A) Labeled Cn1 (+ 5 mM ADP, 10 mM MgCl2) was titrated with unlabelled BiP FL (blue), SBD (red), NBD (*), hPERK (II-IV) (x) or hIRE1 (II-IV) (|). Cn1 binds to BiP FL and BiP SBD with similar micromolar affinities. No binding of Cn1 to BiP NBD, hPERK (II-IV) or hIRE1 (II-IV) is observed.

(B) Pull down assay using His6-tagged BiP (FL, SBD and NBD), hPERK (II-IV) and hIRE1 (II-IV) proteins with untagged Cn1 (lanes 2-6). Cn1 only binds to His6-BiP FL and His6-BiP SBD. In control experiments Cn1 only was incubated with TALON resin (lane 1). The MW (kDa) is indicated.
Figure 7-16. hPERK (II-IV) and hIRE1 (II-IV) do not bind BiP FL-Cn1 complexes

His6-BiP FL was pre-incubated with untagged Cn1 (+ 5 mM ADP, 10 mM MgCl2) to from BiP FL-Cn1 complexes (lanes 4-7). Untagged hPERK (II-IV) (lane 5) and untagged hIRE1 (II-IV) (lane 7) were then added. The luminal domain proteins do not bind to His6-BiP FL when it is already bound to Cn1. In control experiments the untagged proteins only were incubated with TALON resin (lanes 1-3). The MW (kDa) is indicated.
Figure 7-17. C1 promotes dissociation of BiP FL-hPERK (II-IV) and BiP FL-hIRE1 (II-IV) complexes

His6-BiP FL was pre-incubated with untagged hPERK (II-IV) (lane 4-5) or untagged hIRE1 (II-IV) (lane 6-7) to from His6-BiP FL-luminal domain complexes. Untagged C1 (+ 5mM ADP, 10 mM MgCl2) was then added (lanes 5 and 7). C1 actively promotes the release of the luminal domain proteins from His6-BiP FL. In control experiments the untagged proteins only were incubated with TALON resin (lanes 1-3). The MW (kDa) is indicated.
7.5 Discussion

The work in this chapter set out to investigate the interaction between the ER Hsp70 chaperone BiP and the luminal domains of *H. sapiens* PERK and IRE1 proteins. The main aims were to *(i)* demonstrate and biochemically characterise their interaction and *(ii)* understand its importance in the activation of UPR signalling during ER stress.

To this end hPERK and hIRE1 proteins were purified and their binding to BiP FL was measured using MST and ITC *in vitro*. hPERK and hIRE1 bind to BiP FL with \( K_D \approx 1-2 \mu M \). This is the first time their direct interaction has been demonstrated and captured *in vitro*. Low micromolar \( K_D \) values are typical of weak protein-protein interactions that occur *in vivo*. Since in cells, and the ER, proteins are present at low micromolar concentration ranges, affinity values obtained for BiP-luminal domain complexes fit in with the biology of this interaction. The dissociation of BiP from PERK and IRE1 as a result of ER stress has long been recognised to activate UPR signalling pathways. It is therefore important that this interaction is sufficiently strong to be stable during physiological conditions, so as to inhibit UPR signalling, but also that it can be easily disrupted by stress signals to quickly activate the UPR and restore ER homeostasis. Given a 2 \( \mu M \) \( K_D \), it is reasonable to speculate that this is a transient interaction, which allows BiP to readily associate and dissociate from PERK and IRE1.

Previous studies have attempted to understand the functional significance of the sub-regions of PERK and IRE1 luminal domains. The luminal domains were classified into five sub-regions and truncation proteins were used in functional assays. Published studies have found that deletion of region V of PERK and IRE1 led to the loss of association with BiP [85-87]. However there has been ambiguity on the activation state of \( \Delta V \) IRE1 proteins. Whilst some of these studies reported constitutively active UPR signalling, another found that this had no effect on downstream signalling. Furthermore, binding of BiP to hIRE1 proteins lacking region V has been demonstrated [88,89]. Clearly, clarification of the role of the sub-domains has eluded cellular studies carried out to date. The results presented in this chapter indicate that regions II-IV of hPERK and hIRE1 are solely responsible for associating with BiP. Affinity values obtained for constructs encompassing the complete (regions I-V) luminal domains were identical to those which comprised only of the core (regions II-IV) portion. This demonstrates that regions II-IV are necessary and sufficient for binding to BiP. Regions I and V are
dispensable for the direct binding of hPERK and hIRE1 to BiP. It is possible that the truncated proteins used in previous studies were not correctly folded. As such they may have been unable to accommodate binding to BiP or, being unfolded proteins, may have themselves acted as inducers of UPR activation. The use of folded recombinant proteins, as well as the measurement of a direct interaction which precludes influence by the multitude of factors which influence cellular experiments, successfully provided a tool to unambiguously determine the role of regions I to V in binding to BiP. However these results do not elucidate the correlation between the association/dissociation of BiP with the luminal domains and UPR signalling.

The interaction of the luminal domain proteins was further mapped specifically to BiP’s NBD; no binding to BiP SBD was measured. The affinity of hPERK and hIRE1 binding to BiP NBD is in the same range as that measured for binding to BiP FL. These results show that BiP NBD is indispensable for binding to the luminal domains of PERK and IRE1. BiP NBD interaction with a substrate domain is unprecedented for Hsp70s. This result demonstrates that this is a specific and functional protein-protein interaction rather than a chaperone-substrate interaction, for which binding to BiP SBD would be observed. This had eluded many of the cellular based experiments of previous studies. Furthermore, the affinity of BiP FL for the luminal domain proteins was not affected by the presence of ATP, ADP or AMPPNP nucleotides. The ATP-ADP cycle of BiP is known to be key for its chaperoning function. Here BiP’s ATPase activity was uncoupled from its function in binding PERK and IRE1. This further emphasizes that it is a specific UPR signalling relevant interaction rather than a general Hsp70-substrate chaperoning one.

PERK and IRE1 have a distinct BiP binding site to that of unfolded proteins, which are known to bind to BiP SBD. As such, a different functionality of these interactions can be proposed. Whereas unfolded proteins target BiP’s chaperoning activity, BiP-luminal domain complexes probably play a role in ER stress sensing. This hypothesis was tested by using unfolded proteins to probe BiP-PERK and BiP-IRE1 complexes.

In a first instance, the intrinsically unfolded protein Cn1 was used. Cn1 bound exclusively to BiP SBD with similar affinity to that obtained in published studies [22-26]. No binding to BiP NBD or to the luminal domain proteins was observed. Cn1 was used as a BiP SBD-specific
authentic unfolded substrate to study its effect on the interaction between BiP and the luminal domain proteins. Strikingly, the luminal domains were unable of binding to BiP-C$_{II}$1 complexes. Moreover, C$_{II}$1 was capable of dissociating already formed BiP-luminal domain complexes. Clearly the binding of the luminal domain proteins and C$_{II}$1 to BiP are mutually exclusive. This confirms a negative regulatory role for unfolded proteins in the association of BiP with PERK and IRE1 luminal domains. The data presented supports a BiP-dependent mechanism of mammalian ER stress sensing by PERK and IRE1.

Similarly, ΔEspP, an SPR-derived peptide that has been used as an IRE1p-specific unfolded protein mimic, was used to probe the BiP-independent/direct unfolded protein binding mechanism of ER stress sensing. ΔEspP bound exclusively to hPERK and hIRE1 proteins; no binding to BiP proteins was observed. However, in contrast to previously published data, ΔEspP did not induce higher oligomer formation of the luminal domains. Additionally, ΔEspP did not affect BiP-luminal domain complexes. Since BiP is known to repress signalling when it is bound to PERK and IRE1, even if ΔEspP does bind to the luminal domains, this is probably not sufficient to activate UPR signalling during conditions of ER stress. Together, these results point away from a BiP-independent ER stress sensing mechanism, at least for the mammalian system.

This study makes a significant contribution to the mechanistic understanding of BiP-dependent ER stress sensing by the luminal domains of hPERK and hIRE1. In summary, the direct and unprecedented interaction between BiP NBD and the core region of hPERK and hIRE1 luminal domains was demonstrated. This interaction is not affected by nucleotides and there is no contribution from BiP SBD. Together these results emphasize that this is a genuine protein-protein interaction rather than a chaperone-substrate interaction. In the presence of an authentic unfolded protein, which binds solely to BiP SBD, BiP-luminal domain complexes actively dissociate.

These results support an elegant model whereby BiP normally represses UPR signalling by interacting with the luminal domains of the sensor proteins via its NBD. Binding of unfolded proteins to BiP SBD leads to dissociation of these complexes. BiP is known to undergo conformational cycling upon binding to and dissociating from unfolded proteins, and specifically C$_{II}$1. The luminal domain proteins bind to BiP only when it is not bound to C$_{II}$1.
substrate. C1 binding to BiP causes dissociation of the luminal domains. Although no definite conclusion can be made from the results carried out in this work, an allosteric BiP-dependent mechanism of ER stress sensing and UPR activation could be speculated (Figure 7-18).

In the absence of substrate, BiP NBD adopts a conformation where its NBD Lobes IA and IIA divide to create a hydrophobic cleft where the interdomain linker can bind. This also leads to docking of the SBD-α lid onto the NBD. Separation of SBD-α and SBD-β results in an open lid conformation and confers low substrate affinity. During physiological conditions, the luminals bind to substrate-free BiP and maintain it in the open lid conformation/low substrate affinity conformation. During ER stress, binding of unfolded substrates to BiP SBD leads to a conformational change, which consists of ‘closing’ of the SBD-α lid to trap the substrate between SBD-α and SBD-β. Substrate binding to the SBD of BiP is also known to lead to conformational changes in the NBD. The NBD could thus become unfavourable for binding to the luminal domain proteins and cause the BiP-luminal domain complexes dissociate.

Addition of different nucleotides had no effect on binding of BiP to hPERK and hIRE1. This fits in biologically in that depletion of ATP in the ER lumen is not sufficient to breakdown BiP-luminal domains that would result in UPR signalling. Rather, the presence of unfolded proteins is critical for this key activation step.

Further studies will be essential to fully appreciate (i) the interaction of BiP-NBD and the luminal domains of hPERK and hIRE1 and (ii) the precise mechanism by which C1 promotes dissociation of these complexes. More extensive screening using hPERK, hIRE1 and BiP truncations or mutant proteins will be essential to map the precise location and nature of their interaction surface. Obtaining the crystal structure of the complexes will reveal invaluable information to not only appreciate how the complexes physically form but also how this relates to their biological role. It would allow for answering of the fundamental question of how does binding of BiP repress activation of the luminal domains. For instance is it a steric hindrance of regions on the luminal domain proteins that is important for activation, or do the luminal domains undergo a conformational/oligomeric state change upon binding to BiP? Additionally, spFRET and NMR studies could be used to study the
movement of BiP and the luminal domains during complex dissociation. This would allow for a clearer and definitive understanding of the mechanism of how C\textsubscript{1,1}, or more generally the presence of unfolded proteins, leads to dissociation of BiP from the luminal domains.

To highlight the biological significance of the findings by this study, functional experiments using full-length hPERK and hIRE1, or transfected cells, to measure autophosphorylation of their cytoplasmic, eIF2\textalpha phosphorylation and splicing of \textit{xbp1} mRNA as appropriate should be carried out. Furthermore, this study could be extended to ATF6 as well as IRE1p to examine if this is a global mechanism of ER stress sensing and UPR activation or just specific to mammalian PERK and IRE1 pathways.
Figure 7-18. Model for BiP-dependent ER stress sensing by PERK and IRE1 luminal domains

(A) In physiological conditions, BiP, via its NBD, is normally bound to the luminal domains of PERK and IRE1 thereby repressing them.

(B) ER stress leads to the accumulation of misfolded proteins in the ER lumen.

(C) Misfolded proteins bind to BiP SBD which initiates a global conformational change in BiP. This leads to ‘closing’ of the SBD-α lid and rearrangement of the NBD.

(D) In the substrate-bound state, BiP NBD exists in a conformation which is unfavourable for binding to the luminal domains of PERK and IRE1 and as such dissociates. PERK and IRE1 are no longer repressed by BiP and thus can mediate downstream UPR signalling.
8 Conclusion
8.1 Overview of findings

Over the last two decades the mechanisms underlying UPR signalling have been studied in great detail in vivo. Research has focused on elucidating two major events: ER stress sensing (ER luminal) and UPR signal transduction (cytosolic). Despite the considerable effort that has been put into these fields, their definitive molecular mechanisms have not yet been established. The work presented in this thesis provides evidence of (i) a functional role for the tetrameric assembly of PERK (Chapter 6) and (ii) an allosteric mechanism of BiP-dependent ER stress sensing (Chapter 7).

Firstly, the structure of PERK luminal domain is described (Chapter 6). The dimeric and tetrameric assemblies of PERK luminal domain were solved and these can help enlighten our understanding of UPR signalling. Of major importance is the identification of Interface 2. Its hydrophobic nature mediates interchanging of the C-terminal α-helices between two interacting dimers to form stable tetramers. These are also present in solution and are likely to play a functional role. To date, the oligomeric state of active and inactive PERK and IRE1 has not been conclusively established. Based on published data and the results obtained from the structure of PERK luminal domain, two possible scenarios can be envisaged. PERK may exist as inactive monomers and active dimers and tetramers, each of which represents a differently activated state. Alternatively, they may function as inactive dimers and active tetramers. Further experiments will be key to elucidate this.

Secondly, by dissecting the complex pathway of ER stress sensing, an elegant and novel allosteric mechanism is presented (Chapter 7). Normally, BiP NBD is bound to the luminal domains of PERK and IRE1. It has been well established that when BiP is bound to the UPR sensor proteins, their activation is repressed [76,81,82]. During ER stress, unfolded protein binding to BiP SBD leads to dissociation of BiP NBD from the luminal domains PERK and IRE1 and alleviates their inhibition. These unconventional events have never been previously demonstrated and will be of major impact in the UPR field. However, how the release of BiP dissociation leads to activation of downstream UPR signalling remains unexplained.

Although these results have been thoroughly discussed in their respective chapters, the relationship between the two findings has not yet been examined. By taking them into
consideration together a more comprehensive understanding of their important contribution to the UPR field can be appreciated.

A scenario in which BiP NBD binds to inactive dimers of PERK and IRE1 luminal domains to inhibit the formation of active tetramers can be proposed (Figure 8-1A). Analysis of the electrostatic potential of PERK luminal domain reveals that the surface, which is inwards facing in the tetramer assembly, has a strong negative charge (Figure 8-1B). This surface is solvent exposed in dimers. The outer surface of the protein is mostly uncharged. On the other hand, BiP NBD has a positively charged solvent exposed patch. The complementarily between these two surfaces makes them ideal binding partners in vivo. As such, BiP NBD binding to the inner PERK luminal domain face would sterically hinder the formation of tetramers. As such two PERK dimers would be able to come together to form active tetramers only once BiP is released and PERK’s inner face is not obstructed.

Further work will be key to reliably draw conclusions on the mechanism by which ER stress signals result in active UPR signalling. For instance to verify whether BiP inhibits tetramerisation of PERK and IRE1, mutational analysis or spFRET experiments could be carried out both in vitro and in vivo. Furthermore, obtaining a structure of BiP-PERK/IRE1 luminal domains would significantly broaden the understating of UPR activation at atomic detail.
Figure 8.1. A proposed model for BiP-dependent initiation of UPR signalling by steric hindrance of active PERK and IRE1 tetramer formation

(A) (i) Normally BiP NBD is associated with the luminal domains of inactive PERK and IRE1 dimers. (ii) During ER stress, the accumulating misfolded proteins bind to BiP SBD. This results in a global conformational change in BiP that is unfavourable for binding to the luminal domains of PERK and IRE1. The dissociation of BiP relieves its physical hindrance and permits PERK and IRE1 active tetramers to form.

(B) Electrostatic potential of the surface of (i) dimeric PERK (Chapter 6) and (ii) BiP NBD (PDB: 3LDL) [45]. These structures are oppositely charged and thus could represent the interacting surfaces between PERK luminal domain and BiP NBD.
8.2 Implications for the UPR field and protein misfolding disorders

The UPR is the major mechanism by which cells respond to ER stress. UPR signalling aims to adapt the cells to handle the unfolded protein load and restore ER homeostasis. If the damage is too severe, apoptotic pathways are activated instead. Impaired ER and UPR function has been linked to many disease states including diabetes, cancer and neurodegenerative disorders. Gaining a detailed understanding of how UPR signalling is activated has important therapeutic potential.

The pharmacological modulation of UPR signalling has led to successful drugs against important pathologies such as cancers, mainly multiple myeloma, neurodegenerative diseases, diabetes, obesity, glaucoma and ischemia. Similarly ‘chaperone addiction’, including that of BiP, in tumour cells has been important drug target [150].

Discovery of the unconventional model of BiP-dependent ER stress sensing, and appreciation of the precise role of BiP sub-domains in the UPR, can be exploited for extremely targeted drug design in cancers. Inhibition of BiP SBD would reduce the folding efficiency of the cell, thus more quickly result in the lethal accumulation of unfolded proteins. In addition, this would ensure that BiP NBD remains functional and thus successfully represses UPR signalling by PERK and IRE1. This prevents the activation of resolving UPR pathways, which may otherwise serve as a negative feedback loop to alleviate ER stress. Together, this would quickly drive the cell towards apoptotic pathways. In a similar fashion, inhibiting BiP NBD binding to the luminal domains PERK and IRE1 would allow for constitutive UPR activation and more efficient removal of unfolded protein aggregates that cause many of the pathologies mentioned above.

The novel X-ray structure of PERK luminal domain has shed light on biologically important tetramers. Based on it, novel compounds that can modulate activation of PERK, and IRE1, independently of ER stress could be designed. These would allow for the chemical manipulation of UPR signalling independently of exogenous environmental signals.
The discovery of the effect of protein misfolding, ER stress and the UPR is ongoing. The findings of this thesis will clearly be of interest and beneficial for scientists in the field. They will also importantly provide powerful tools for the therapeutic exploitation of the UPR.
9 References


166. Smith JL. Multiple Anomalous Diffraction in Molecular Crystallography; 1997; Daresbury Laboratory.


10 Appendix
<table>
<thead>
<tr>
<th>Protein</th>
<th>Species</th>
<th>F/R</th>
<th>#</th>
<th>Enzyme</th>
<th>Nucleotide sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ggaattcgctagcaggccgcttgccctc</td>
</tr>
<tr>
<td>PERK</td>
<td></td>
<td>41</td>
<td></td>
<td></td>
<td>ggaattcgctagcatatgcagcgagctga</td>
</tr>
<tr>
<td></td>
<td>H. sapiens</td>
<td>F</td>
<td>105</td>
<td>Nhel</td>
<td>ggaattcgctagcttttagtaaatatagccagtta</td>
</tr>
<tr>
<td></td>
<td></td>
<td>146</td>
<td></td>
<td></td>
<td>ggaattcgcagcttagaggtatttgagaataagatg</td>
</tr>
<tr>
<td></td>
<td></td>
<td>229</td>
<td></td>
<td></td>
<td>ggaattcgctagcataaggaacaggaagagtcc</td>
</tr>
<tr>
<td></td>
<td></td>
<td>207</td>
<td></td>
<td></td>
<td>gcgcgccgtattccatatgtgatcagaga</td>
</tr>
<tr>
<td></td>
<td></td>
<td>355</td>
<td></td>
<td></td>
<td>gcgcgccgttaaagctgatgggaatgaac</td>
</tr>
<tr>
<td></td>
<td>M. msucul</td>
<td>R</td>
<td>384</td>
<td>NotI</td>
<td>gcgcgccgttagttttctgtgtgctctct</td>
</tr>
<tr>
<td></td>
<td></td>
<td>403</td>
<td></td>
<td></td>
<td>gcgcgccgttaaattctgcagctgactg</td>
</tr>
<tr>
<td></td>
<td></td>
<td>510</td>
<td></td>
<td></td>
<td>gcgcgccggccgcttagatattctttgttaatg</td>
</tr>
<tr>
<td>PERK</td>
<td>D. melano-</td>
<td>F</td>
<td>W165A</td>
<td></td>
<td>gcagtcagccctctcagtctctgagagcagacgagagt</td>
</tr>
<tr>
<td>gaster</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>gcagtggacctttcagcgccgagagagtagc</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R</td>
<td>101</td>
<td>Nhel</td>
<td>gcagtcagccctctcagtctctgagagcagacgagagt</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R</td>
<td>399</td>
<td>NotI</td>
<td>gcgcgccgttagctaccgtgacgagagt</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F</td>
<td>182</td>
<td>Nhel</td>
<td>gcagtcagccctctcagtctctgagagcagacgagagt</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R</td>
<td>384</td>
<td>NotI</td>
<td>gcgcgccgttagctaccgtgagagcagacgagagt</td>
</tr>
</tbody>
</table>

Table 10-1. Cloning of PERK, IRE1, BIP and C_H1 proteins

The complete sequence of primers used for the cloning of PERK, IRE1, BIP and C_H1 proteins are shown (continued on next page). The position (♯) of the 5’ N-terminal starting residue, for forward (F) primers, and 3’ C-terminal ending residue, for reverse (R) primers, are indicated. Restriction enzymes used for cloning are listed. For primers used for site-directed mutagenesis, the target mutation residue and amino acid changes are indicated in the ♯ column.
<table>
<thead>
<tr>
<th>Protein</th>
<th>Species</th>
<th>F/R</th>
<th>#</th>
<th>Enzyme</th>
<th>Nucleotide sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>F</td>
<td>24</td>
<td></td>
<td>ggaattgctagccatagcagctttcctgaaacctgttgg</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>32</td>
<td>Ndel</td>
<td>ggaattccatatgtggtgttgggtcgaacgctgg</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>155</td>
<td></td>
<td>ggaattccatatgcgtatccttgggcgagaacagaat</td>
</tr>
<tr>
<td>IRE1</td>
<td>H. sapiens</td>
<td>R</td>
<td>129</td>
<td></td>
<td>aagcttttagctaactaaccacagagtc</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>300</td>
<td>HindIII</td>
<td>ggaccgaagccttttaaccgagcaagggagaggc</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>390</td>
<td></td>
<td>ggcggaagcctttacacattcccgagctttgg</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>440</td>
<td></td>
<td>ggcggaagcctttgaactttacatggagttcacaaggggggctc</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F</td>
<td>28</td>
<td></td>
<td>ggaattgctagccatagcagccagctttggt</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>422</td>
<td>Ndel</td>
<td>ggaattgctagccatagcagctttggt</td>
</tr>
<tr>
<td>BiP</td>
<td>H. sapiens</td>
<td>R</td>
<td>405</td>
<td></td>
<td>ggcggaagcctttggtaccccttagagacacccacccctgacagcggcc</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>654</td>
<td>HindIII</td>
<td>tagaggtacctcatactttccagatctgtgaagctta</td>
</tr>
<tr>
<td>C_{41}</td>
<td>H. sapiens</td>
<td></td>
<td>1</td>
<td>Ndel</td>
<td>ggaattccatatgccacccctcagcagctttatcc</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>88</td>
<td>HindIII</td>
<td>ggcggaagccttttaacgcgaagagttagtgttctgaacc</td>
</tr>
</tbody>
</table>

(continued from previous page)
<table>
<thead>
<tr>
<th>Heavy atom</th>
<th>Total res (Å)</th>
<th>Anom res (Å)</th>
<th>Isom</th>
<th>'group'</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mercury (II) acetate</td>
<td>3.8</td>
<td>4.6</td>
<td>Y</td>
<td>iii</td>
</tr>
<tr>
<td>Mercury (II) iodide</td>
<td>10</td>
<td>NA</td>
<td>Y</td>
<td>i</td>
</tr>
<tr>
<td>Mercury (II) chloride</td>
<td>10</td>
<td>NA</td>
<td>Y</td>
<td>i</td>
</tr>
<tr>
<td>Mercury (II) cyanide</td>
<td>6.8</td>
<td>NA</td>
<td>Y</td>
<td>i</td>
</tr>
<tr>
<td>Phenyl mercury acetate</td>
<td>7.1</td>
<td>NA</td>
<td>Y</td>
<td>i</td>
</tr>
<tr>
<td>Ethylmercurithiosialicylic acid</td>
<td>4.1</td>
<td>5.3</td>
<td>Y</td>
<td>ii</td>
</tr>
<tr>
<td>Ethylmercury phosphate</td>
<td>7.2</td>
<td>NA</td>
<td>Y</td>
<td>i</td>
</tr>
<tr>
<td>Gold (III) chloride</td>
<td>4.3</td>
<td>NA</td>
<td>Y</td>
<td>ii</td>
</tr>
<tr>
<td>Gold (I) potassium cyanide</td>
<td>4.6</td>
<td>NA</td>
<td>Y</td>
<td>ii</td>
</tr>
<tr>
<td>Sodium tetrachloroaurate (III) hydrate</td>
<td>3.8</td>
<td>NA</td>
<td>Y</td>
<td>ii</td>
</tr>
<tr>
<td>Ammonium tetrachloroplatinate (II)</td>
<td>10</td>
<td>NA</td>
<td>Y</td>
<td>i</td>
</tr>
<tr>
<td>Potassium tetrachloroplatinate (II)</td>
<td>10</td>
<td>NA</td>
<td>Y</td>
<td>i</td>
</tr>
<tr>
<td>Potassium tetracyanoplatinate (II) hydrate</td>
<td>10</td>
<td>NA</td>
<td>Y</td>
<td>i</td>
</tr>
<tr>
<td>Sodium tungstate dihydrate</td>
<td>3.1</td>
<td>4.0</td>
<td>Y</td>
<td>iii</td>
</tr>
<tr>
<td>Silver nitrate</td>
<td>10</td>
<td>NA</td>
<td>Y</td>
<td>i</td>
</tr>
<tr>
<td>Samarium (III) acetate hydrate</td>
<td>4.2</td>
<td>NA</td>
<td>Y</td>
<td>ii</td>
</tr>
<tr>
<td>Europium (III) chloride hexahydrate</td>
<td>4.8</td>
<td>NA</td>
<td>Y</td>
<td>ii</td>
</tr>
<tr>
<td>I3C magic triangle</td>
<td>10</td>
<td>NA</td>
<td>Y</td>
<td>i</td>
</tr>
<tr>
<td>Tantulum (V) bromide</td>
<td>10</td>
<td>NA</td>
<td>Y</td>
<td>i</td>
</tr>
</tbody>
</table>

Table 10-2. hPERK (II-IV) condition A crystals heavy atom datasets

The total and anomalous outer resolution limit (NA indicates none measured), isomorphism between the derivative and native datasets (yes (Y) or no (N)) are summarised. The results are categorized into three groups: (i) no heavy atom binding, (ii) binding but not adequate diffraction or (iii) binding and diffraction.
<table>
<thead>
<tr>
<th>Heavy atom</th>
<th>Total res (Å)</th>
<th>Anom res (Å)</th>
<th>Isom</th>
<th>‘group’</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mercury (II) acetate</td>
<td>4.3</td>
<td>7.8</td>
<td>Y</td>
<td>ii</td>
</tr>
<tr>
<td>Mercury (II) iodide</td>
<td>10</td>
<td>NA</td>
<td>Y</td>
<td>i</td>
</tr>
<tr>
<td>Mercury (II) chloride</td>
<td>4.1</td>
<td>6.8</td>
<td>Y</td>
<td>ii</td>
</tr>
<tr>
<td>Mercury (II) cyanide</td>
<td>10</td>
<td>NA</td>
<td>Y</td>
<td>i</td>
</tr>
<tr>
<td>Phenyl mercury acetate</td>
<td>10</td>
<td>NA</td>
<td>Y</td>
<td>i</td>
</tr>
<tr>
<td>Ethylmercurithiosalicylic acid</td>
<td>3.9</td>
<td>5.6</td>
<td>N</td>
<td>iii</td>
</tr>
<tr>
<td>Ethylmercury phosphate</td>
<td>4.7</td>
<td>NA</td>
<td>N</td>
<td>ii</td>
</tr>
<tr>
<td>Gold (III) chloride</td>
<td>3.7</td>
<td>NA</td>
<td>Y</td>
<td>ii</td>
</tr>
<tr>
<td>Gold (I) potassium cyanide</td>
<td>4.3</td>
<td>7.5</td>
<td>Y</td>
<td>ii</td>
</tr>
<tr>
<td>Sodium tetrachloroaurate (III) hydrate</td>
<td>3.9</td>
<td>5.2</td>
<td>N</td>
<td>iii</td>
</tr>
<tr>
<td>Ammonium tetrachloroplatinate (II)</td>
<td>10</td>
<td>NA</td>
<td>Y</td>
<td>i</td>
</tr>
<tr>
<td>Potassium tetrachloroplatinate (II)</td>
<td>6.9</td>
<td>NA</td>
<td>N</td>
<td>i</td>
</tr>
<tr>
<td>Potassium tetracyanoplatinate (II) hydrate</td>
<td>10</td>
<td>NA</td>
<td>N</td>
<td>i</td>
</tr>
<tr>
<td>Sodium tungstate dihydrate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Silver nitrate</td>
<td>10</td>
<td>10</td>
<td>Y</td>
<td>ii</td>
</tr>
<tr>
<td>Samarium (III) acetate hydrate</td>
<td>4.0</td>
<td>NA</td>
<td>Y</td>
<td>ii</td>
</tr>
<tr>
<td>Europium (III) chloride hexahydrate</td>
<td>4.3</td>
<td>NA</td>
<td>Y</td>
<td>ii</td>
</tr>
<tr>
<td>I3C magic triangle</td>
<td>10</td>
<td>NA</td>
<td>Y</td>
<td>ii</td>
</tr>
<tr>
<td>Tantulum (V) bromide</td>
<td>10</td>
<td>NA</td>
<td>Y</td>
<td>ii</td>
</tr>
</tbody>
</table>

Table 10-3. mPERK (II-IV) condition D crystals heavy atom datasets

The total and anomalous outer resolution limit (NA indicates none measured), isomorphism between the derivative and native datasets (yes (Y) or no (N)) are summarised. The results are categorized into three groups: (i) no heavy atom binding, (ii) binding but not adequate diffraction or (iii) binding and diffraction.
Table 10-4. Hydrogen bonding of PERK luminal domain Interface 1 and 2

The atoms involved in the dimerisation Interface 1 and tetramerisation Interface 2 of PERK luminal domain are listed. Distances are given in Å.

<table>
<thead>
<tr>
<th></th>
<th>Interface 1</th>
<th></th>
<th>Interface 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Monomer ‘A’</td>
<td>Dist. (Å)</td>
<td>Monomer ‘B’</td>
<td>Dist. (Å)</td>
</tr>
<tr>
<td>Leu200 (N)</td>
<td>2.9</td>
<td>Gly196 (O)</td>
<td>Met172 (N)</td>
<td>3.8</td>
</tr>
<tr>
<td>Lys198 (N)</td>
<td>3.1</td>
<td>Lys198 (O)</td>
<td>Met172 (N)</td>
<td>2.6</td>
</tr>
<tr>
<td>Gly196 (N)</td>
<td>2.7</td>
<td>Leu200 (N)</td>
<td>Asp258 (N)</td>
<td>3.7</td>
</tr>
<tr>
<td>Arg246 (NE)</td>
<td>2.6</td>
<td>Ala218 (O)</td>
<td>Arg379 (NE)</td>
<td>2.7</td>
</tr>
<tr>
<td>Tyr214 (OH)</td>
<td>3.8</td>
<td>Ser252 (O)</td>
<td>Arg379 (NH2)</td>
<td>3.8</td>
</tr>
<tr>
<td>Gln223 (NE2)</td>
<td>3.4</td>
<td>Asn254 (OD1)</td>
<td>Asn384 (ND2)</td>
<td>3.1</td>
</tr>
<tr>
<td>Cys216 (CG)</td>
<td>3.6</td>
<td>Gly253 (O)</td>
<td>Leu388 (O)</td>
<td>2.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Arg379 (NE)</td>
<td></td>
</tr>
</tbody>
</table>
Figure 10-1. pTWO-E and pET-26b plasmid maps

The pTWO-E and pET-26b plasmids used for cloning are illustrated. Each contains an origin of replication (Ori) (red), lacI gene (green) and T7 promoter (orange) region that allow for IPTG-controlled expression of the gene inserted at the multiple cloning site (blue). The His$_6$-tag, PreScission protease cleavage site (PreSc) and restriction enzyme sites present in the multiple cloning sites are indicated. Their antibiotic resistance genes (purple) are shown.
Figure 10-2. Calibration of HiLoad 16/60 Superdex 200 column

(A) Thyroglobulin (1), apoferretin (2), β-amylase (3), alcohol dehydrogenase (4), albumin (5) and carbonic anhydrase (6) proteins were run on a HiLoad 16/60 Superdex 200 column. Their MW (kDa) and elution volumes (Ve) are indicated. V0 was obtained separately from the elution of Blue Dextran in the HiLoad 16/60 Superdex 200 column. V0/Ve values used to plot the calibration curve are shown.

(B) V0/Ve values were plotted against log MW and a nonlinear regression line was fitted. The resulting function (y = 0.127ln(x) – 0.7681) was used to estimate MW of samples from their elution volumes on the HiLoad 16/60 Superdex 200 column.
Figure 10.3. PERK and IRE1 luminal domain sequence and secondary structure alignment. The sequences of hPERK (magenta), mPERK, dPERK, hIRE1 (blue), mIRE1 and IRE1p luminal domains were aligned using the T-coffee server. Residues are coloured according to conservation (red for 100% identical; yellow for >70% similarity). The known structural features of hIRE1 luminal domain (PDB: 2HZ6) are shown, and named according to the published structure [95]. The secondary structure of hPERK luminal domain was predicted using Jpred3 and is also shown (red).
Figure 10-4. hPERK luminal domain disorder prediction

(A) The complete hPERK luminal domain sequence (1-510) was analysed using RONN software to predict protein disorder. A probability of disorder >0.5 suggests disorder.

(B) hPERK sequence and disorder prediction for each residue is shown. ‘*’ indicates disordered residues; ‘_’ indicates ordered residues. The major disordered portion is suggested to be the N-terminal region (residues 1-103).
Figure 10-5. Wilson Plot of hPERK (II-IV) Condition A crystals diffraction data

(A-B) The Wilson Plot of hPERK (II-IV) Condition A crystals collected on crystals (A) without cryoprotection and (B) after gradual dehydration over a final 40% w/v PEG3500.
Figure 10-6. hPERK (II-IV) and mPERK (II-IV) SeMet incorporation

(A-B) MALDI-TOF analysis of (A) SeMet hPERK (II-IV) and (B) SeMet mPERK (II-IV) purified samples. Peaks represent multiple charge states of component peptides; numbers indicate the charge state. The mass (Da) was calculated using the equation: mass (Da) = (m/z value x charge state) – charge state. SeMet hPERK (II-IV) and SeMet mPERK (II-IV) protein masses were 34431 and 34221 Da respectively. These indicate the presence of 7.5 and 6.9 Se atoms per protein that corresponds to 94% and 99% SeMet incorporation respectively.
### Figure 10-7. hPERK (II-IV) heavy atom binding

hPERK (II-IV) was mixed with heavy atoms and samples were analysed on a native gel (3-8% Tris-Acetate). Numbers at the top of the lanes refer to the heavy atom type (legend on the left). A significant shift was observed for hPERK (II-IV) in the presence of sodium tungstate dihydrate (lane 15) compared to the native sample (lane 1).

<table>
<thead>
<tr>
<th>Lane</th>
<th>Heavy atom</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Native</td>
</tr>
<tr>
<td>2</td>
<td>Mercury (II) acetate</td>
</tr>
<tr>
<td>3</td>
<td>Mercury (II) iodide</td>
</tr>
<tr>
<td>4</td>
<td>Mercury (II) chloride</td>
</tr>
<tr>
<td>5</td>
<td>Mercury (II) cyanide</td>
</tr>
<tr>
<td>6</td>
<td>Phenyl mercury acetate</td>
</tr>
<tr>
<td>7</td>
<td>Ethylmercurithiosalicylic acid</td>
</tr>
<tr>
<td>8</td>
<td>Ethylmercury phosphate</td>
</tr>
<tr>
<td>9</td>
<td>Gold (III) chloride</td>
</tr>
<tr>
<td>10</td>
<td>Gold (I) potassium cyanide</td>
</tr>
<tr>
<td>11</td>
<td>Sodium tetrachloroaurate (III) hydrate</td>
</tr>
<tr>
<td>12</td>
<td>Ammonium tetrachloroplatinate (II)</td>
</tr>
<tr>
<td>13</td>
<td>Potassium tetrachloroplatinate (II)</td>
</tr>
<tr>
<td>14</td>
<td>Potassium tetracyanoplatinate (III) hydrate</td>
</tr>
<tr>
<td>15</td>
<td>Sodium tungstate dihydrate</td>
</tr>
<tr>
<td>16</td>
<td>Silver nitrate</td>
</tr>
<tr>
<td>17</td>
<td>Samarium (III) acetate hydrate</td>
</tr>
<tr>
<td>18</td>
<td>Europlum (III) chloride hexahydrante</td>
</tr>
<tr>
<td>19</td>
<td>I3C magic triangle</td>
</tr>
<tr>
<td>20</td>
<td>Tantulum (V) bromide</td>
</tr>
<tr>
<td>21</td>
<td>Gadolinium (III) chloride hydrate</td>
</tr>
<tr>
<td>22</td>
<td>Thallium (I) chloride</td>
</tr>
<tr>
<td>23</td>
<td>Lead (II) nitrate</td>
</tr>
<tr>
<td>24</td>
<td>Cadmium iodide</td>
</tr>
</tbody>
</table>
Figure 10-8. Location of Na₂WO₄ heavy atom sites

(A) CCall versus CCweak plot of MAD dataset processing using ShelxD. Correct solutions are clustered at high CCall and CCweak values.

(B) FoFc difference map of hPERK (II-IV) data. A large positive density (green) is observed at the site of Na₂WO₄ (red). Binding is mainly coordinated by Trp165 residue. Trp165 distance (Å) to W atom is shown.
Figure 10-9. Comparison of PERK and IRE1 luminal domain structures

(A-B) The crystal structure of *H. sapiens* PERK luminal domain (blue) is aligned to that of (A) *H. sapiens* IRE1 (PDB: 2HZ6) (cyan) and (B) *S. cerevisiae* IRE1 (PDB: 2BE1) (purple) [95,101]. Structures are shown as ribbons. The central core portions, which include dimerisation Interface 1, are almost identical. The most marked differences occur in the outer regions (Lobe 1 and Lobe 2), especially near the C-terminal α-helix.
The sequences of *H. sapiens* PERK (purple), *H. sapiens* IRE1 (blue) and *S. cerevisiae* IRE1 (red) were aligned based on their secondary structures. Their overall features are similar, especially between the *H. sapiens* proteins.
**Figure 10-11. BiP and ΔEspP heat of dilution in ITC experiments**

(A) BiP (500 μM) was titrated into buffer K. The resulting isotherm was integrated and fitted with a straight line to calculate the heat of dilution. A constant value of 5.8 kcal/mol was obtained.

(B) ΔEspP (500μM) was titrated into buffer K. The resulting isotherm was integrated and fitted with a straight line to calculate the heat of dilution. A constant value of 0.1 kcal/mol was obtained; this heat of dilution is negligible.
Figure 10-12. Gel shift assays of hPERK (II-IV)-BiP FL complexes

(A) Purified hPERK (II-IV) (11 μM) and BiP FL (11 μM) were pre-incubated and analysed directly on a native gel (3-8% Tris-Acetate). hPERK (II-IV) and BiP FL were run alone (+ buffer) as controls. No additional bands could be discerned in the complex sample compared to hPERK (II-IV) (blue) and BiP FL (red) control samples.

(B) Purified hPERK (II-IV) (11 μM) and BiP FL (11 μM) were cross-linked with a 25-fold molar excess of EGS prior to analyzing by SDS-PAGE (4-12% Bis-Tris gel). DMSO was used instead of EGS as a control. hPERK (II-IV) and BiP FL were run alone (+ buffer) as controls. Once again, bands in the complex lanes can be attributed to hPERK (II-IV) (blue) or BiP FL (red) proteins.
Figure 10-13. ΔEspP binding to BiP FL

(A) DSF was used to test for ΔEspP peptide (MKHKRILALCFLGGLLQSSYSAKKKK) (+ 5mM ADP, 10 mM MgCl₂) binding to BiP FL. The Tm of BiP FL is not significantly affected by the presence of ΔEspP suggestive of no interaction.

(B) ΔEspP (500 μM) (+ 5mM ADP, 10 mM MgCl₂) was titrated into BiP FL (55 μM) (VP-ITC MicroCalorimeter). The data was integrated and fitted to a one-binding site model. The association constant (Kₐ), changes in enthalpy (ΔH) and entropy (ΔS) and stoichiometry (N) values obtained are shown. Dissociation constants (Kₐ) were calculated. ΔEspP does not bind to BiP FL.
Figure 10.14. Oligomeric state of hPERK (II-IV) and hIRE1 (II-IV) in complex with ΔEspP

(A-B) SEC MALS analysis (Superdex 200 PC 3.2/30 column) of (A) hPERK (II-IV) (100 μM) and (B) IRE1 (II-IV) (100 μM) with or without ΔEspP (100 μM). The calculated MW of the peaks obtained is indicated. In both experiments, no change in the oligomeric state of the proteins is observed.
Figure 10-15. ΔEspP does not induce dissociation of BiP FL-hPERK (II-IV) and BiP FL-hIRE1 (II-IV) complexes

His₆-BiP FL was pre-incubated with untagged hPERK (II-IV) (lane 3-4) or untagged hRE1 (II-IV) (lane 5-6) to from His₆-BiP FL-luminal domain complexes. ΔEspP was then added (lanes 4 and 6). ΔEspP does not have a discernable effect on the association of the luminal domain proteins with His₆-BiP FL in pull down assays. In control experiments the untagged proteins only were incubated with TALON resin (lanes 1-2). The MW (kDa) is indicated.