A Study of Protein Regulation by Mass Spectrometry-Based Methods



Victoria Helen Beilsten-Edmands

Jesus College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2017

A Study of Protein Regulation by Mass Spectrometry-Based Methods

Victoria Helen Beilsten-Edmands

Jesus College, University of Oxford
A thesis submitted for the degree of Doctor of Philosophy, Michaelmas Term 2017

Abstract

Understanding the function and regulation of proteins is critical to medical and physiological research. To achieve this, the modification status, as well as the interactions of proteins need to be defined. Mass spectrometry is often used to catalogue modifications but their effect on interactions is less often reported. In this thesis these two strands are combined to address a series of challenging complexes that underpin cellular processes from stress responses to energy production.

Following an introduction to mass spectrometry and its applications to the study of protein interactions (Chapters 1 and 2), three protein assemblies are considered; Hsp70 chaperones, eukaryotic translation initiation factors, and F- and V-type ATP synthases. First, the oligomerisation of Hsp70 is studied (Chapter 3). Dimerisation is shown to be dependent on a highly-conserved phosphorylation site, suggesting a role for this modification in the regulation of protein complex assembly and targeting to antagonistic pathways.

A study of eukaryotic translation initiation factor 2B (eIF2B) is presented (Chapter 4), with proteomics revealing clusters of modifications at protein interfaces. Chemical cross-linking is applied, eliminating proposed structural models and supporting a decameric assembly. A cognition-enhancing drug (known as ISRIB) is an inhibitor of the integrated stress response, and was found to stabilise the active decameric eIF2B assembly, rationalising the observed therapeutic effects of ISRIB.

Finally, ATP synthases, membrane protein complexes which exhibit multiple forms of regulation, are investigated (Chapter 5). Positional information of small membrane subunits and the effect of post-translational modifications is deduced. The role of an endogenous inhibitor protein is investigated providing insights into its location, oligomeric state, and influence on nucleotide binding.

Overall this thesis applies mass spectrometry to three critical protein assemblies and contributes a more complete understanding of their activity by uncovering aspects of their carefully-tuned regulatory mechanisms.

This thesis is dedicated to my loving family

Acknowledgements

I owe my gratitude to many people for their contributions and support throughout my DPhil studies. Firstly, I am grateful to Professor Carol Robinson for giving me the opportunity to work on so many exciting and challenging research projects and for her continued optimism, support, and guidance. I would like to thank all of the members of the Robinson and Benesch groups, both past and present, with whom I have worked for these past years. In particular I would like to thank Professor Carla Schmidt (Martin Luther University of Halle-Wittenberg) for her endless support and advice, including discussing data, providing feedback on my research, technical advice and training, and not least her friendship. I am also grateful to Carla for her contributions to many of my projects, but particularly the cATP synthase project.

I would like to thank Dr Dror Chorev with whom I worked closely on the bovine ATP synthase project. I am grateful for his friendship, advice, technical assistance, and for always being willing to make time to discuss everything. I thank Dr Lucia Geis Asteggiante for her cheerful advice and for carefully reading this thesis. I would like to thank Dr Nisha Patel and Dr Frances Kondrat for their advice and for the time they spent training me in mass spectrometry and protein purification.

I acknowledge and thank Dr Yuliya Gordiyenko and Jocky Kung for providing purified eIF2 protein and performing preliminary experiments, and Dr Di Wu for his advice and expertise in performing denaturing LC-MS and lipidomics experiments. I express my gratitude to Professor Eugenia Clerico (University of Massachusetts), Dr Min Yang (University College London), Dr Ima-obong Ebong, and Professor Nina Morgner (Goethe University Frankfurt) who were helpful in providing advice on the Hsp70 project. I would like to thank Professor Shabaz Mohammed for his technical advice and allowing me to use instrumentation for proteomics analysis.

The work described in this thesis would not have been possible were it not for the contributions from my collaborators, in particular for the high-quality protein samples with which they provided me. I would like to acknowledge and thank the following people; Dr Elaine Kirschke and Zygy Roe-Zurz from Professor David Agard's group (UCSF) for

providing plasmids and the purified Hsp70$_{Sf9}$ protein, Professor Graham Pavitt (University of Manchester) for providing yeast strains and the purified yeast eIF2B protein, Jordan Tsai from Professor Peter Walter's group (UCSF) for providing the human eIF2B proteins, Thamiya Vasanthakumar and Stephanie Bueler from Professor John Rubinstein's group (University of Toronto) for the yeast V-type ATPase protein samples, and Professor Christoph Gerle (University of Osaka) for the bovine F-type ATP synthase protein samples.

I am immensely grateful to my family, to whom this thesis is dedicated, for their unfailing love, endless support, and always being wonderful. I would like to thank my sister Melanie for her love and friendship and my parents for their generosity, for always believing in me and encouraging me, and for their unconditional love. I would like to thank my husband James for being a constant source of encouragement, joy, and love, and for being a excellent father. Finally, I would like to thank my son Jonathan, born during the course of my DPhil studies, who has brought me so much happiness and a new perspective to everything I do.

Contents

1 Protein Structure and Function **1**
1.1 Structural Biology Techniques 1
1.2 Overview of the Thesis 7

2 Mass Spectrometry **9**
2.1 Electrospray Ionisation 10
2.1.1 Nanoelectrospray Ionisation 14
2.2 Ion Transmission 16
2.3 Mass Analysers 16
2.3.1 The Quadrupole Mass Analyser 17
2.3.2 The Time-of-Flight Mass Analyser 18
2.3.3 The Orbitrap Mass Analyser 20
2.3.4 Tandem Mass Spectrometry 22
2.4 Ion Detection 25
2.5 Instrumentation 27
2.5.1 Micromass Quadrupole-Time-of-Flight II Mass Spectrometer . 27
2.5.2 Thermo Fisher Scientific LTQ Orbitrap XL Hybrid Ion Trap-Orbitrap Mass Spectrometer 28
2.5.3 Thermo Fisher Scientific Q Exactive Plus Hybrid Quadrupole-Orbitrap Mass Spectrometer 28
2.6 Interpretation of Electrospray Mass Spectra 30
2.7 Liquid Chromatography 32
2.8 Cross-linking: Structural Biology Approaches at the Peptide Level . . 35
2.8.1 Comparative Cross-linking Mass Spectrometry 37

3 The Role of Post-Translational Modifications in the Dimerisation of the 70 kDa Heat Shock Protein **39**
3.1 Introduction 39
3.2 Protein Folding 40
3.3 Molecular Chaperones 42
3.3.1 70 kDa Heat Shock Protein 44
3.4 Research Aims 48
3.5 Results 49
3.5.1 The Dependence of Hsp70 Dimerisation on the Expression System 49
3.5.2 Identification of Post-Translational Modifications on Hsp70 . . 51

3.5.3 The Effect of Hsp70 Dephosphorylation on Dimerisation . . . 53
3.5.4 The Effect of Ionic Strength on Hsp70 Dimerisation 55
3.5.5 The Dimerisation of an Hsp70 Phosphomimic 56
3.5.6 The Effect of Mutation of the Substrate-Binding Domain on Dimerisation 58
3.6 Discussion 60
3.7 Conclusions 67
3.8 Methods 71
3.8.1 Protein Preparation 71
3.8.2 Mass Spectrometry 73
3.8.3 Homology Modelling 77

4 Regulation of the Interactions of Eukaryotic Translation Initiation Factor 2B 78
4.1 Introduction 78
4.2 Protein Translation 79
4.2.1 Translation Initiation 79
4.2.2 The Eukaryotic Translation Initiation Factors 80
4.2.3 The Integrated Stress Response 84
4.3 Research Aims 86
4.4 Results 87
4.4.1 Yeast eIF2B 87
4.4.2 Human eIF2B 95
4.5 Discussion 105
4.5.1 Post-Translational Modification of Yeast eIF2B 105
4.5.2 Chemical Cross-Linking of Yeast eIF2B 107
4.5.3 Native Mass Spectrometry of Human eIF2B 107
4.5.4 The Interaction of ISRIB and Human eIF2B 108
4.6 Conclusions 113
4.7 Methods 116
4.7.1 Protein Purification 116
4.7.2 Identification of Post-Translational Modifications 119
4.7.3 Homology Modelling 121
4.7.4 Chemical Cross-linking Mass Spectrometry 121
4.7.5 Native Mass Spectrometry 122

5 Regulation of the Rotary ATP Synthases 123
5.1 Introduction 123
5.2 ATP Synthases 124
5.2.1 The Structure of the F-type ATP Synthases 128
5.2.2 The Structure of the V-type ATPases 131
5.2.3 The Regulation of ATP Synthases 133
5.3 Research Aims 137
5.4 The Yeast V-type ATPase 139
5.4.1 Results and Discussion 139

5.5 The Spinach Chloroplast F-type ATP Synthase 144
5.5.1 Results and Discussion 144
5.6 Bovine Mitochondrial F-type ATP Synthase 149
5.6.1 Results 150
5.6.2 Discussion 171
5.7 Conclusions 175
5.8 Methods 179
5.8.1 Protein Purification 179
5.8.2 Proteomics Analysis 183
5.8.3 Cross-Linking Mass Spectrometry 184
5.8.4 Homology Modelling 186
5.8.5 Denaturing Mass Spectrometry 186
5.8.6 Lipidomics Analysis 187
5.8.7 Native Mass Spectrometry 188

6 Conclusions **189**

A Supplemental Information for Chapter 3 **196**

B Supplemental Information for Chapter 4 **198**

C Supplemental Information for Chapter 5 **201**

Bibliography **207**

Definitions of initialisms and acronyms

β-ME : β-Mercaptoethanol
5 UTR : 5 untranslated region
6.8PL : 6.8 kDa proteolipid
ACN : acetonitrile
ADP : adenosine diphosphate
aIF2 : archaeal initiation factor 2
AMP-PNP : adenosine 5-(β,γ-imido)triphosphate
Arg : arginine
Asp : aspartic acid
ATF4 : activating transcription factor 4
ATP : adenosine triphosphate
BS3 : bis(sulfosuccinimidyl)suberate
BS3-d0 : non-deuterated bis[sulfosuccinimidyl] suberate
BS3-d4 : deuterated bis[sulfosuccinimidyl] suberate
C. albicans : *Candida albicans*
CCS : collisional cross-section
CEM: chain ejection model
C. elegans : *Caenorhabditis elegans*
CHIP : C-terminal Hsp70-interacting protein
CID : collision-induced dissociation
CRM : charged residue model
DAPIT: diabetes-associated protein in insulin-sensitive tissues
DC : direct current
DDM : n-Dodecyl β-D-maltoside
DHB : 2,5-dihydroxybenzoic acid
D. melanogaster : *Drosophila melanogaster*
DM : decyl maltoside
DMSO : dimethyl sulfoxide
DNA : deoxyribonucleic acid
DSS : disuccinimidyl suberate
DSS-d0 : non-deuterated disuccinimidyl suberate
DSS-d12 : deuterated disuccinimidyl suberate
DTT : dithiothreitol
E. coli : *Escherichia coli*
ECD : electron capture dissociation
EDTA : ethylenediaminetetraacetic acid
eIF : eukaryotic initiation factor
EM : electron microscopy
EMR : extended mass range
ER : endoplasmic reticulum
ESI : electrospray ionisation
FA : formic acid
FT-ICR : Fourier-transform ion cyclotron resonance
GCN2 : general control depressible 2 kinase

GDP : guanosine diphosphate
GEF : guanine nucleotide exchange factor
Glu : glutamic acid
GTP : guanosine triphosphate
HCD: higher energy collisional dissociation
HEPES : 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HETP : height equivalent to a theoretical plate
His : histidine
HOP : Hsp organising protein
HPLC : high performance liquid chromatography
HRI : heme-regulated eukaryotic initiation factor 2 α kinase
H. sapiens : *Homo sapiens*
Hsp70 : 70 kDa heat shock protein
IC : initiation complex
IEM : ion evaporation model
IMD : intermembrane space domain
IPA : isopropanol
IPTG : isopropylthiogalactopyranoside
ISR : integrated stress response
ISRIB : integrated stress reponse inhibitor
LB : lysogeny broth
LC : liquid chromatography
LC-MS : liquid chromatography mass spectrometry
LMNG : lauryl maltose neopentyl glycol
Lys : lysine
m/z : mass to charge ratio
Met-tRNA$_i^{Met}$: initiator methionyl transfer ribonucleic acid
MFC : multi factor complex
MS : mass spectrometry
MS/MS : tandem mass spectrometry
M. musculus : *Mus musculus*
NBD : nucleotide-binding domain
nESI : nanoelectrospray ionisation
NHS : N-hydroxysuccinimide
NMR : nuclear magnetic resonance spectroscopy
OSCP: oligomycin sensitivity-conferring protein
PBS : phosphate-buffered saline
PDB : protein data bank
PERK : double-stranded protein-like endoplasmic reticulum kinase
PIC : pre-initiation complex
PKR : double-stranded protein kinase
PMF : proton motive force
PMSF : phenylmethylsulfonyl fluoride
PTM : post-translational modification
Q-ToF : quadrupole/time-of-flight

RAVE : regulator of the ATPase of vacuolar and endosomal membranes
RF : radiofrequency
RNA : ribonucleic acid
rpm : revolutions per minute
SEC : size-exclusion chromatography
SILAC : stable isotope labelling by amino acids in culture
SBD : substrate-binding domain
S. cerevisiae : *Saccharomyces cerevisiae*
SDS-PAGE : sodium dodecyl sulfate polyacrylamide gel electrophoresis
S. pombe : *Schizosaccharomyces pombe*
S. solfataricus : *Solfalobus solfataricus*
SID : surface-induced dissociation
TBI : traumatic brain injury
TC : ternary complex
TCEP : tris(2-carboxyethyl)phosphine
TEV : tobacco etch virus
TFA : trifluoroacetic acid
ToF : time-of-flight
TPR : tetratricopeptide repeat
tRNA : transfer ribonucleic acid
Tyr : tyrosine
UHMR : ultra-high mass range
UPR : unfolded protein response
WT : wild type
YPD media: yeast extract peptone dextrose media

Chapter 1

Protein Structure and Function

Proteins are an indispensible part of living organisms and perform a wide range of functions. They are responsible for providing structural integrity, for the transportation and translocation of small molecules, proteins, cells and even other organisms, for regulation including signalling, and for catalysing countless biochemical reactions [1]. To fulfil many of these roles, proteins assemble into larger complexes, the assembly of which is often assisted by other proteins. Protein structure, along with changes in structure, dynamics, and intrinsic disorder [2] determine function and consequently the elucidation of the structure and conformational dynamics of proteins and multisubunit protein assemblies is essential to understanding protein function, regulation, and interactions.

1.1 Structural Biology Techniques

A large variety of techniques is available for the study of macromolecular structure with X-ray crystallography, electron microscopy (EM), mass spectrometry (MS), and nuclear magnetic resonance (NMR) among the most prominent. Each technique offers its own advantages in addition to limitations.

X-ray crystallography has for many years been at the forefront of structural biology, providing protein structures with unparalleled resolution. It is therefore no surprise that the majority of protein structures deposited in the protein data bank (PDB) structural repository [3] are derived from X-ray diffraction of protein crystals. X-ray crystallography is a well-established technique, with significant effort from the structural biology community resulting in an abundance of robust experimental procedures along with high-quality software for the automation of data acquisition and analysis [4]. In spite of the considerable contributions that X-ray crystallography has made to the field of structural biology, challenges and limitations of this technique persist. The first of these is implicit in the very nature of protein crystals and the difficulties involved in their production. Not all proteins are amenable to crystallisation and for those that are, finding the correct conditions to produce high-quality, well-diffracting crystals can be time consuming and require large quantities of sample, the production of which may be a significant hindrance. Furthermore, X-ray crystallography provides a static structure in a single conformation, which whilst very informative, often leaves the area of protein dynamics unexplored and allows only a single protein conformation to be determined from each crystal. Additionally, deriving a crystal structure from an experimental diffraction pattern requires complex data analysis to attempt to overcome the phase problem, a significant issue inherent in the interpretation of X-ray diffraction data [5]. While there are techniques that attempt to address the phase problem including protein engineering, these can be laborious and also result in the introduction of non-native protein features.

NMR is able to overcome some of the limitations of X-ray crystallography. NMR is compatible with samples in buffered native-like aqueous solutions foregoing the requirements to find the optimal conditions for protein crystallisation and the

resulting non-physiological environment protein crystals bring [6]. Furthermore, NMR excels at characterising protein dynamics while being particularly good at measuring shifts in protein conformation in response to stimuli such as ligand binding. On the other hand, NMR can require the use of isotopically-labelled protein samples, which can become costly while spectral assignment can prove difficult. One of the drawbacks of using NMR in analysing protein complexes is its size limitations, with analysis of larger complexes becoming very difficult, leaving NMR most suited to analysing smaller proteins and isolated protein domains [6].

EM and electron cryo-microscopy (cryo-EM) has for a long time provided somewhat lower resolution structural information on large protein complexes. Recent advances in detection methods and computational analysis however have resulted in an increase in higher resolution cryo-EM-generated structures bringing cryo-EM to the cutting edge of protein structure determination [7]. Currently, software and instrumentation for the automation of data acquisition and analysis are not as abundant and advanced as those for X-ray crystallography, although the significant resources being directed towards them means that it is expected that they soon will be. Freezing of samples for cryo-EM can result in breaking of multisubunit interactions, while samples may suffer artefacts arising from the freezing process and radiation damage during data acquisition. Often, the resolution of structures deduced from cryo-EM is not homogeneous throughout the structure with some parts remaining at resolutions too low to accurately determine protein structure. In addition, the study of protein dynamics is somewhat challenging by cryo-EM, while analysis of smaller complexes is more difficult, albeit with this limitation being reduced [8].

Although X-ray crystallography, NMR, and cryo-EM are able to offer

high-resolution structures of many proteins and protein complexes, there are still areas of structural biology with questions that cannot be answered by such analyses. There are many techniques that, while not providing the high-resolution structural data produced by the aforementioned techniques, are still able to provide important information. These techniques are too numerous to discuss in detail here [9–13], however for the purposes of this thesis I will focus on MS and its ability to provide unique and valuable insights into characterising protein complexes that are inaccessible or unsuited to other techniques.

MS has become an indispensable tool in the study of proteins, particularly in the field of proteomics, with MS becoming a routine method for protein identification. MS allows accurate mass measurements of proteins and peptides allowing substituents of protein complexes to be identified alongside providing information on interaction partners, protein expression levels, and protein isoforms. MS-based proteomics can be used in conjunction with some of the high-resolution structural techniques discussed above, as protein identification may be required to provide the starting point for model building or identification of unknown features in electron density maps [14]. Furthermore, although the genome encodes a vast array of different proteins, the experimentally-determined proteome is greater still. Part of this proteome expansion arises through pre-mRNA splicing prior to protein translation, while a significant level of diversity is generated through chemical modifications occurring co- and post-translationally, with these modifications intimately involved with protein structure, dynamics, and behaviour [15]. Post-translational modifications (PTMs) are a key mechanism for regulation in cells and identifying their location and function therefore is a key area of research. The mass shift imparted by PTMs means that MS has emerged as a leading technique in their identification, allowing accurate PTM site determination, assessment of global

PTM status and quantification of PTM site occupancy [16].

MS-based techniques are not limited to protein and PTM identification. Techniques including chemical cross-linking coupled to MS are able to provide structural information down to individual amino acid residues, yet do not require native protein structure to be maintained for the duration of cross-linking analysis while being compatible with near-native buffers during the chemical cross-linking reaction [17].

MS is not limited to the analysis of denatured protein subunits and peptides. The development of native MS has opened up the field to the analysis of protein samples under non-denaturing conditions with instrumental conditions allowing the maintenance of non-covalent interactions, which has resulted in a wealth of information on various proteins and protein complexes including both soluble and membrane proteins [18–21]. Native MS particularly excels in the determination of protein complex stoichiometry, with highly-precise mass determination allowing elucidation of protein complex association and oligomerisation with high accuracy [22]. Another strength of MS is its ability to detect multiple different protein species within a single spectrum allowing analysis of heterogeneous samples, in contrast to other techniques such as X-ray crystallography which may only detect a single protein complex and may have difficulty with heterogeneous samples.

In spite of the advantages of MS-based proteomics and native MS, analysis takes place in the gas phase. Proteins exist *in vivo* in a primarily aqueous environment, constantly interacting with various other species such as salts, small molecules, lipids, and other proteins. With hydrogen-bonding networks and the hydrophobic effect playing such important roles in protein structure and interactions, this

aqueous medium is of paramount importance to the maintenance of native functional protein structure and activity [23]. To understand the behaviour of proteins *in vivo*, solution-phase structure must be preserved. Studying proteins in the gas phase, as is the case in MS, in the absence of water-mediated interactions may not seem a natural choice for studying protein structure and interactions and yet MS has made numerous contributions to the field of structural biology [20, 24–26], with an abundance of studies demonstrating that native protein interactions can be preserved in the gas phase (reviewed in [27]).

Much of the evidence for gas-phase protein structure originates from ion mobility mass spectrometry (IM-MS) studies [28]. Ion mobility separates gas-phase ions by their mobility in a carrier gas and allows calculation of the gas-phase collisional cross-sections (CCS) of protein ions. Hence, IM-MS can be used to compare the topology of gas-phase ions with that expected from the native structure. Multiple studies have shown that the measured gas-phase CCS corresponds well to the calculated CCS of the native structure [29–32] showing that native structure of both soluble proteins and membrane proteins, in the presence of detergent micelles, can be maintained in the gas phase for extended lengths of time. Furthermore, these studies have shown that high charge states in MS correspond to more unfolded protein conformations, while lower charge states are indicative of more compact native-like structures [28]. In addition, it has been reported that proteins can maintain their enzymatic activity after transmission through the mass spectrometer, followed by soft landing [33]. As protein function derives from structure, a functional complex is a strong indication of maintenance of native structure, thus indicating that native structure can be maintained for the duration of gas-phase MS analysis.

In practice, proteins are often amenable to a number of structural biology techniques with each analysis excelling at elucidating various aspects of protein structure and function. Many recent studies have combined multiple techniques to generate complementary data sets, giving additional insights that would not be possible without such integration [34–40].

1.2 Overview of the Thesis

It is using the technique of native MS, along with MS-based proteomics with which I will characterise protein complexes throughout this thesis. I aim to further our understanding of protein interactions, function, and regulation by combining multiple MS-based methods to study three particular protein systems, motivated by the insights gained from complementary techniques.

After introducing the theory and instrumentation of MS, I will apply MS to the study of the oligomerisation of the 70 kDa heat shock protein. I will identify post-translational modifications present and subsequently explore the role of phosphorylation, and specifically phosphorylation at an identified phosphosite, on oligomerisation. I will finally consider the conformation of identified oligomers and discuss the MS results within their biological context.

Next, a key enzyme in eukaryotic translation initiation will be studied; eukaryotic translation initiation factor 2B. Firstly, the identification of PTMs and their location within the structure will be explored. The structure of the initiation factor complex, still an area of confusion, will be investigated by cross-linking MS. Finally the interaction of the initiation factor with the recently-identified memory-enhancing drug ISRIB will be studied and insights into the mechanism of

action will be discussed.

Proteomics and native MS techniques will then be combined to interrogate ATP synthases from three different species. The many techniques used in the thesis will be brought together with proteomics and denaturing MS analyses used to identify protein isoforms and determine PTM status. Cross-linking MS will be employed to explore the subunit association, organisation, and the conformational effects of PTMs in ATP synthases. Finally, native MS will be used to probe the subunit architecture and stoichiometry, including binding of lipids and nucleotides, and effects of the binding of the natural inhibitor IF_1. These analyses will be brought together, integrated, and discussed in the context of previous studies.

Chapter 2

Mass Spectrometry

Proteomics refers to the large-scale study of the proteome, that is the study of the set of proteins expressed by a particular organism under defined conditions [41, 42], and addresses a wide variety of questions including those surrounding protein expression, localisation, modifications, and interactions. MS has emerged as an invaluable tool in proteomics investigations [43], with advances in instrumentation allowing increasing levels of detail to be uncovered. One of the most commonly-used and effective areas of proteomics is termed bottom-up proteomics and employs enzymes to hydrolyse protein sequences into shorter peptides, which are more amenable to transmission and fragmentation in the mass spectrometer. These peptides are usually separated by high-performance liquid chromatography (HPLC), the eluent of which is continuously transferred to a mass spectrometer via an electrospray source. Fragmentation of peptides during MS analysis allows identification of the amino acid sequence and modifications such as phosphorylation. These techniques can also be applied to intact proteins under denaturing conditions, with MS allowing determination of the global PTM status of protein subunits. Advances in instrumentation and methodology have allowed non-denaturing conditions to be used under which non-covalent protein interactions can be

maintained, including those of membrane proteins [19]. The advent of non-denaturing non-covalent MS (native MS) has allowed MS to gain new insights into protein structure and interactions, with studies of protein-protein interactions, complex stoichiometry, and ligand binding now amenable to MS analysis. Recent developments in instrumentation and methodology have allowed analysis to be extended to increasingly large protein complexes, while mass accuracy, resolution, and sensitivity continue to increase [44].

2.1 Electrospray Ionisation

The first step in MS is the ionisation and desolvation of the analyte. Today, electrospray ionisation (ESI) is the most commonly-used technique in protein MS to generate gas-phase protein ions. Although the theory of electrospray was first described by Taylor in 1964 [45] and proposed as a means to transfer large molecules to the gas phase by Dole *et al.* in 1968 [46], it was not until 1988 that Fenn and co-workers were able to transfer large molecules, such as proteins, to the gas phase for mass analysis, without fragmentation [47, 48]. As ESI ionises analytes from liquids, ESI-MS can be interfaced to a wide variety of liquid chromatographic techniques, prior to ionisation.

In ESI, a protein sample in solution is introduced to a conductive capillary held at a high voltage relative to a downstream plate electrode with a central orifice. In positive ion mode[1], the electric field causes polarisation of the solution, predominantly at the tip of the capillary, with negative ions being attracted to the capillary, enriching positive ions at the capillary tip [49].

[1] As positive ion mode is the most commonly-used mode, and is used for most of the spectra reported in this thesis, positive ion mode will be assumed from this point forth unless explicitly stated otherwise.

The electric field then causes the positively-charged solution to experience a downstream force, distorting the meniscus into a Taylor cone (Figure 2.1), with the force due to the electric field being counteracted by the surface tension of the liquid. It has been calculated that the electric field can be given by the approximate form

$$E_c = \frac{2V_c}{r_c \ln(4d/r_c)}, \tag{2.1}$$

i.e. proportional to the voltage V_c and increased by reducing the capillary radius r_c and the distance d between the capillary and downstream electrode. As the electric field must overcome the surface tension to form a jet of charged droplets, there is a

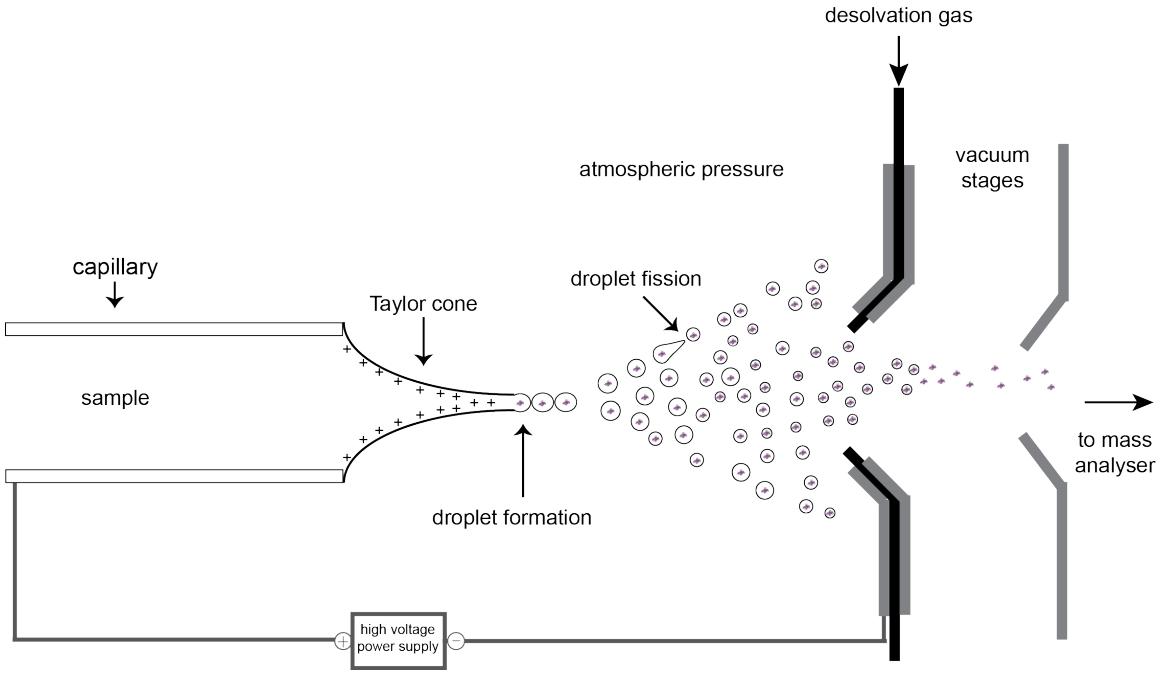


Figure 2.1: Illustration of the electrospray ionisation process. The sample solution is loaded into a conductive capillary which is held at high voltage relative to the downstream electrode. Positive ions are enriched at the capillary tip, forming a Taylor cone from which budding droplets form. Droplets shrink by evaporation until the increase in charge density and therefore Coulombic repulsion induces droplet fission. Solvent evaporation continues, forming desolvated, charged analyte ions which are transmitted into the mass spectrometer for analysis. Figure based on Kerbarle and Verkerk [49].

minimum voltage needed to induce electrospray. This voltage has been calculated by Smith *et al.* [50] to have the approximate form

$$V_{on} \approx \left(\frac{r_c \gamma \cos\theta}{2\epsilon_o}\right)^{1/2} ln\left(\frac{4d}{r_c}\right), \tag{2.2}$$

where γ is the surface tension, θ is the Taylor cone half angle, and ϵ_o is the electric permittivity of the surrounding medium. Importantly, V_{on} is proportional to the square root of the surface tension, giving considerations for experimental solvent preparation.

In each droplet, as the solvent evaporates and the droplet size reduces, the Coulombic force due to the surface charges increase. At a certain droplet radius, this repulsion overcomes the surface tension of the droplet and Coulombic fission occurs. The condition for instability is given by the Rayleigh equation

$$Q_{Ry} = 8\pi(\epsilon_o \gamma R^3)^{1/2}, \tag{2.3}$$

relating the droplet charge Q_{Ry} to the surface tension and droplet radius R. This process occurs repeatedly as the solvent evaporates, producing small charged progeny droplets. In order to aid in this process, a nebulising gas such as N_2 can be applied in the opposite direction to droplet motion, accelerating solvent evaporation.

The exact mechanism of ion formation has been the subject of debate with two main mechanisms proposed. The first of these models is the ion evaporation model (IEM, Figure 2.2A), proposed by Iribarne and Thomson in 1976 [51], and thought to be the mechanism via which low molecular weight, charged analytes are transferred to the gas phase. In the IEM, a droplet is generated containing the pre-charged analyte, often assisted in protonation by acidic solution conditions. As

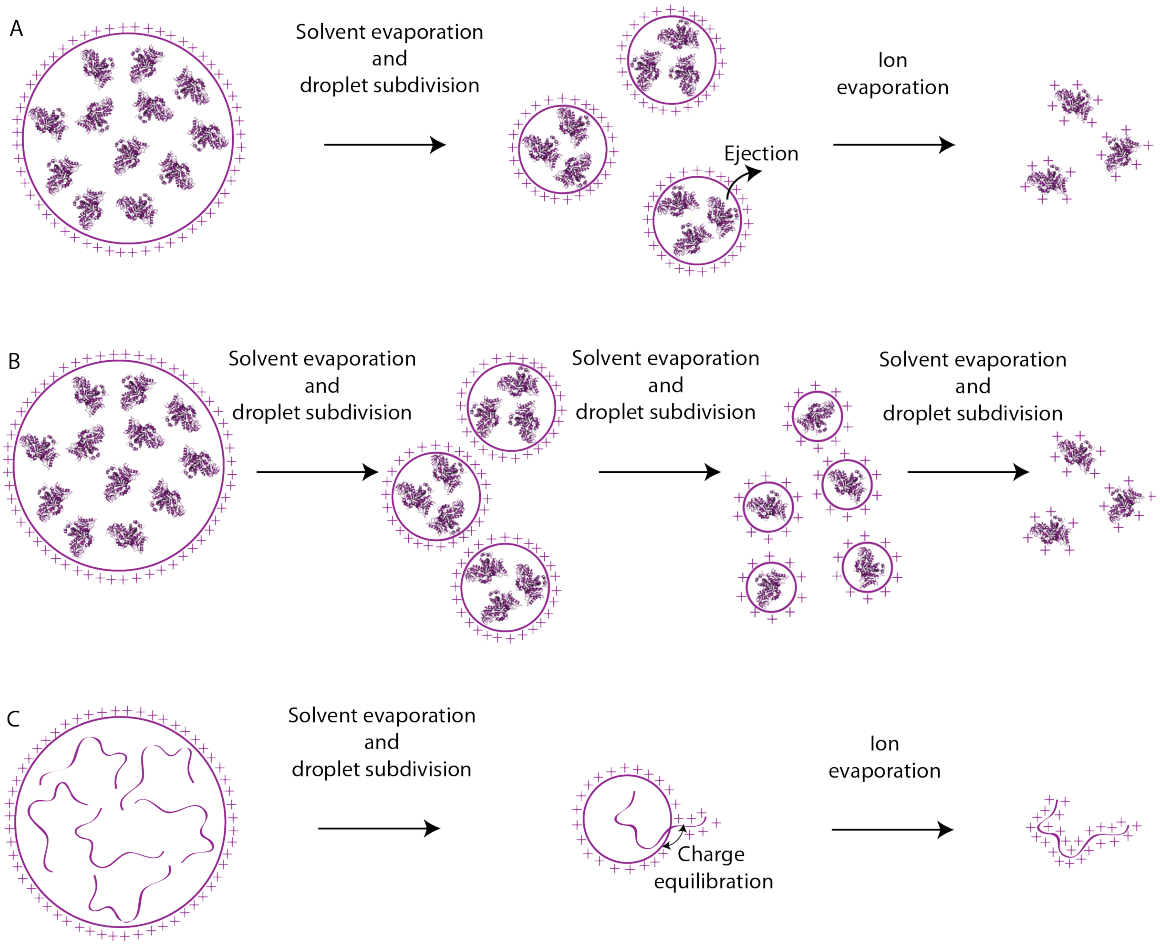


Figure 2.2: Models of ionisation in electrospray ionisation. (A) The ion evaporation model (IEM). (B) The charged residue model (CRM) of ionisation. (C) The chain ejection model (CEM). Protein illustrated by PDB ID: 1PKN [53].

the droplet shrinks through solvent evaporation, the charge density of the droplet increases until the field strength at the droplet surface overcomes solvation forces, ejecting the solvated analyte from the droplet [52].

The second mechanism proposed for ion formation is the charged residue model (CRM, Figure 2.2B), proposed by Dole *et al.* in 1968 [46], and thought to apply to large globular analytes such as folded proteins [52]. In this model it is assumed that a single analyte is present in each droplet, which undergoes repeated cycles of solvent evaporation and droplet fission. This results in the generation of a

near-desolvated droplet containing a single analyte. The charge of globular proteins is closely related to the Rayleigh limit of the droplet immediately before desolvation [54], and not dependent on the analyte's charge in solution [52], supporting the CRM as the mechanism of ion formation for globular proteins.

Molecular dynamics simulations have indicated that unfolded proteins follow a third mechanism rather than the IEM or CRM, called the chain ejection model (CEM, Figure 2.2C), which applies to chargeable, disordered polymers with some hydrophobic character as is the case in unfolded proteins [52]. The use of acidic solution conditions, such as those used in liquid chromatography MS (LC-MS), can trigger protein unfolding, with non-polar hydrophobic residues becoming solvent accessible. In the CEM, it is not favourable for the hydrophobic protein residues to remain within the charged droplet, and the protein termini migrate to the charged droplet surface. The protein terminus is ejected from the droplet, with further ejection continuing until the entire protein chain is expelled from the droplet [52].

2.1.1 Nanoelectrospray Ionisation

The development of nanoelectrospray ionisation (nESI) by Wilm and Mann in 1994 [55] was a major step towards the advancement of non-covalent MS. The difference between standard ESI and nESI is defined by the size of the capillary needle orifice, with an ESI needle orifice of ≈ 0.5 mm compared to a nESI orifice of 1–10 μm. The smaller orifice results in a smaller electrospray droplet, in the 200 nm range, which gives rise to a number of favourable attributes for nESI. Firstly, ESI often requires the use of organic volatile solvents to maintain the electrospray, however this is not necessary for nESI, allowing samples to be introduced in the native-like aqueous buffers, required to maintain protein non-covalent interactions.

Secondly, the lower flow rate of nESI necessitates μL sample volumes whereas for ESI sample volume requirements are in the mL range [21]. This lower sample volume requirement is particularly advantageous when sample amount is limited, as is often the case with biological samples.

Thirdly, the smaller droplets formed with nESI mean that less solvent evaporation is required to generate a desolvated analyte, allowing gentler conditions for solvent evaporation than are necessitated with ESI [21]. This, together with native-like aqueous buffers being permissible as the sample's solvent, means that conditions can be controlled in a way which allows the transmission of protein complexes maintaining non-covalent interactions which would otherwise be broken in ESI with denaturing solvents and harsh desolvation conditions [56]. A second effect of the smaller droplet size is that fewer fission events are required to create a desolvated analyte ion. Salt concentration within the droplet increases with each fission event and salt adducts lead to poor resolution and signal intensity. The fewer fission events required in nESI means that nESI is more tolerant to salt leading to better resolution and signal intensity [55]. Finally, a third advantage to the lower droplet size is the reduced incidence of non-specific interactions observed. When the initial droplets are formed in the electrospray process, it is possible that more than one analyte is resident in each drop. As solvent evaporates, by the CRM, proteins can associate into non-specific complexes leading to artefactual oligomerisation to be observed in the mass spectrum. The smaller the initial droplet size, the lower the chance of trapping more than one analyte in a droplet. In nESI, these non-specific interactions become less problematic with fewer incidences of non-specific complex formation occurring [21], a feature which becomes imperative when using MS to investigate protein-protein and protein-ligand interactions. Although nESI allows gentler desolvation conditions, the conditions required to achieve complete

desolvation may still be harsher than those needed to maintain non-covalent complexes, particularly in high-mass complexes. A frequently-observed phenomenon therefore, is the presence of buffer adducts to the detected complexes, leading to peak broadening and elevated mass measurements, up to 1 % higher than expected in high-mass complexes [57].

2.2 Ion Transmission

In order to transmit ions from the atmospheric pressure electrospray ion source throughout the various stages of the mass spectrometer to the detector, many factors must be carefully controlled with tuning of vacuum pressures and use of focussing lenses among many of the techniques employed [58]. The transmission of larger ions can be difficult to achieve as their high kinetic energy can lead to defocussing of the ion beam. In native MS the low charge states of analytes, together with their high masses, necessitates high m/z ranges. One technique to assist in overcoming this challenge is termed collisional cooling or collisional focussing and is effected by increasing the pressure of the first vacuum region of the mass spectrometer. The higher pressure increases the number of collisions that analytes undergo with neutral gas molecules, collisions which convert a fraction of the analyte's kinetic energy to internal vibrational energy, thus reducing defocussing and additionally aiding in complete desolvation both increasing transmission and resolution [59].

2.3 Mass Analysers

Mass spectrometers employ a range of mass analysers to analyse ions based on their mass to charge ratio (m/z). Different mass analysers have advantages and disadvantages with resolution, mass accuracy, m/z range, and cost being some of

the most important factors to consider.

2.3.1 The Quadrupole Mass Analyser

The quadrupole mass analyser, described in 1953 [60] consists of four equidistant rods of elliptical cross-section to which are applied fixed DC or oscillating AC voltages (Figure 2.3). Voltages are applied to pairs of opposite rods; $\pm(U + Vcos(2\pi ft))$ where U is a DC voltage and V is the magnitude of an oscillating AC voltage with frequency f [61]. Within the quadrupole, this spatial configuration of voltages leads to a quadrupolar electric field. The changing potential with respect to time t gives specific conditions for stable ion trajectories through the quadrupole, with the equations of motion being described by the Mathieu equation [61]. In the solutions for stable trajectories, the stability is dependent on the m/z ratio and the magnitude of the AC and DC voltages. By selecting the appropriate AC and DC voltages, ions with a particular range of m/z

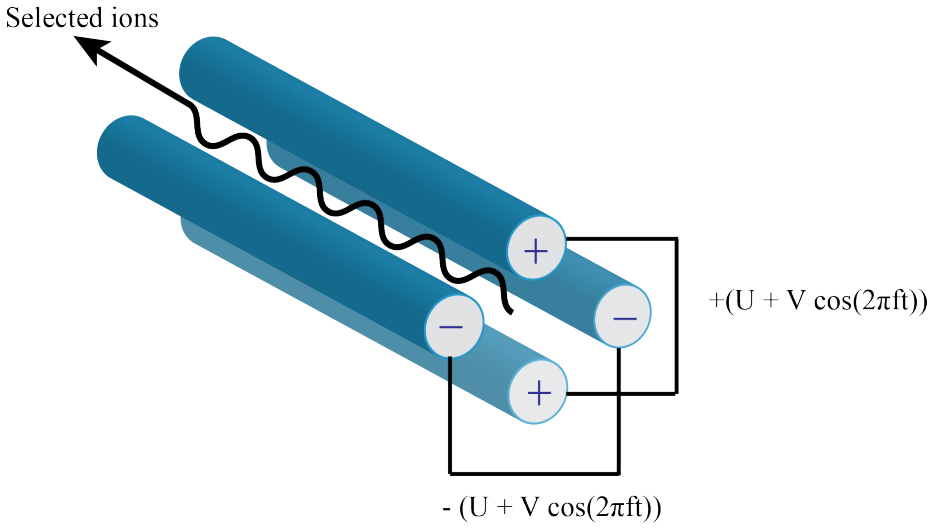


Figure 2.3: Schematic of the quadrupole mass analyser. Opposing rods are electrically connected, with applied voltages $\pm(U + V\cos(2\pi ft))$. Ions with stable trajectories are transmitted, with tuning of the AC and DC voltages allowing a particular range of m/z values to be selected for transmission.

values can be selected for transmission. The quadrupole can also be used as an ion guide to focus ions rather than select them. This is achieved by operating the quadrupole in RF-only mode which allows ions with a wide range of m/z values to be transmitted with stable trajectories.

The maximum m/z transmitted by the quadrupole is determined by

$$\left(\frac{m}{z}\right)_{max} = \frac{4 \times 10^6 V}{f^2 r_o^2} \tag{2.4}$$

where V is the RF amplitude, f is the frequency, and r_o is the inner radius of opposing rods. At lower operating frequency f, an ion experiences fewer RF cycles which reduces the spectral resolution while also leading to transmission of higher m/z ions. Increasing the mass range for higher m/z species therefore comes with the penalty of decreased resolution. A quadrupole can also be transformed into a linear ion trap by the use of end electrodes, to give a confining potential within the axial direction [59].

2.3.2 The Time-of-Flight Mass Analyser

The time-of-flight analyser (ToF) is one of the conceptually-simplest mass analysers. Ions are initially accelerated in the source region, before entering the ToF analyser, which contains a field-free drift region [62]. In the idealised model of a charged ion accelerated from rest, the kinetic energy of the ion entering the drift region is given by

$$E_k = zV, \tag{2.5}$$

where z is the charge of the ion and V is the potential difference applied in the source region. As the ions enter the drift region, the velocity is given by

$$v = \sqrt{\frac{2E_k}{m}} = \sqrt{\frac{2zV}{m}}, \tag{2.6}$$

which is dependent on the mass. As a result, the time of flight through the drift region of length D is given by

$$t = \frac{D}{v} = D\sqrt{\frac{m}{2zV}}, \tag{2.7}$$

and therefore is directly related to $\sqrt{m/z}$. As a result, the time spectrum on the detector can be converted to an m/z spectrum. This requires knowing the time at which the ion was accelerated in the source region, hence a pulsed ion source is used, creating packets of charged ions.

In practice however, ions are not accelerated from rest and have a distribution of kinetic energies. Consequently, time of flight is not solely dependent on the ions' m/z. To account for the variable initial kinetic energy of ions, the ToF can be coupled with a reflectron (Figure 2.4). A 'pusher' creates a potential gradient which

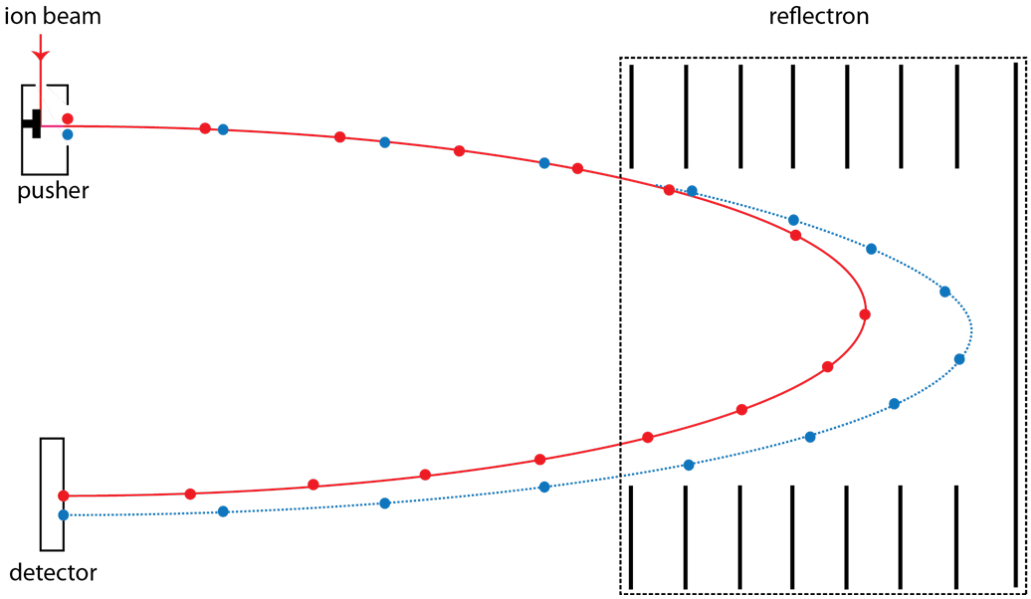


Figure 2.4: Schematic of the ToF mass analyser, incorporating a reflectron. Ions are accelerated orthogonal to their arrival trajectory into the reflectron. Ions with a high initial kinetic energy penetrate further into the reflectron, resulting in a longer path length, correcting their time of flight for the initial kinetic energy distribution.

accelerates ions towards the reflectron orthogonal to the beam axis, such that initial kinetic energy is independent of orthogonal acceleration. The reflectron consists of a series of lenses with a fixed electrostatic field. Positively-charged ions quickly decelerate before being accelerated back out towards the detector. Ions with a greater amount of kinetic energy penetrate further into the reflectron and thus have a longer time of flight. This results in the focussing of isobaric ions by time, reducing the problem of ions' initial kinetic energy distribution and increasing the resolution of mass spectra.

2.3.3 The Orbitrap Mass Analyser

The Orbitrap (Figure 2.5) is a pulsed-injection Fourier-transform mass analyser invented by Makarov [63], using the principle of orbital trapping. The Orbitrap is a development on the earlier Kingdon trap, described by Kingdon in 1923 [64]. The Kingdon trap consists of a stretched wire and an outer cylinder electrode aligned coaxially, between which a DC potential difference is applied, generating a radial electrostatic field of the form

$$U(r) = A \ln(r) + B. \tag{2.8}$$

For ions that are introduced into the trap with a range of velocities perpendicular to the central wire, the radial field causes only certain initial velocities to move in stable orbits within the trap. In the Orbitrap mass analyser, the inner wire of the Kingdon trap is replaced with an axially-symmetric spindle-shaped electrode, and the outer cylinder with a complementary-shaped electrode, across which a potential difference is applied [65]. A key difference to other mass analysers is the use of a static electric field, as opposed to the dynamic oscillating fields found in other popular analysers such as the quadrupole.

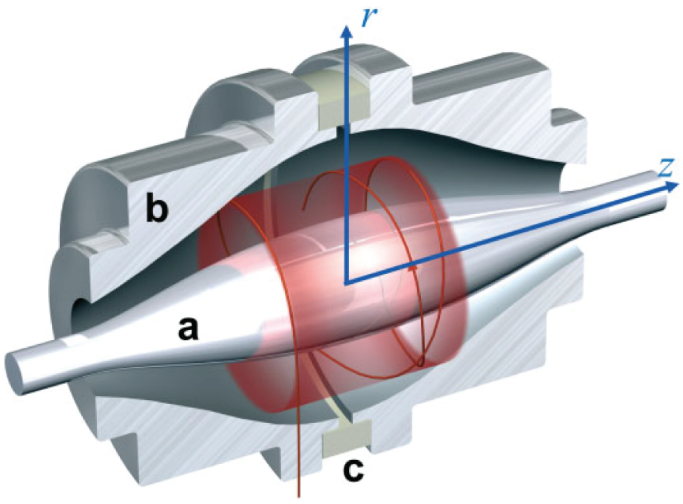


Figure 2.5: Cross-section of the Orbitrap mass analyser. Ions are injected into the Orbitrap with a displacement along the z axis. Ions oscillate around the central electrode (a), with motion along z independent of initial kinetic energy. The outer electrode (b) is split into two by an insulating ring (c), allowing image current detection. Figure from Scigelova and Makarov [66].

The geometry of the Orbitrap is such that a quadrupolar term is introduced into the electrostatic potential, which is described by

$$U(r,z) = \frac{kz^2}{2} - \frac{kr^2}{4} + \frac{k}{2}(R_m)^2 \ln\left[\frac{r}{R_m}\right] + C \tag{2.9}$$

where k is the field curvature and R_m is the characteristic radius. The absence of cross-terms in the electrostatic potential means that the axial and radial/angular motions are independent, and the quadratic potential in the axial z direction causes simple harmonic motion. As such, stable trajectories include oscillations along the axial direction, in addition to the required orbital motion as in the case of a Kingdon trap, with a frequency of axial oscillation

$$\omega = \sqrt{k(z/m)}. \tag{2.10}$$

Therefore, ions are introduced into the trap tangentially to the axial direction, but are displaced from the centre of the trap (i.e. $z \neq 0$) in order to induce the simple harmonic motion along z, as well as to select given ions based on the initial velocities and applied potential difference.

In order to achieve high resolution and signal, ions are injected into the Orbitrap in packets so as to achieve coherence of motion for ions of particular m/z. One of the notable features of the Orbitrap is its high resolution and high mass accuracy, relative to mass analysers such as the ToF. In the Orbitrap the axial frequency is measured, which is independent of energy and spatial distribution of ions, which allows m/z to be determined precisely and accurately, with high mass resolution. In comparison to another high-accuracy mass analyser, Fourier-transform ion cyclotron resonance (FT-ICR), the Orbitrap is much smaller in size and lower in cost, both of which make it a highly-popular mass analyser. These features of high mass accuracy and precision, together with low cost and size make the Orbitrap an excellent mass analyser for both native MS and proteomics applications.

2.3.4 Tandem Mass Spectrometry

Tandem MS uses dissociation of non-covalent complexes to interrogate their composition, by placing two or more mass analysers in series separated by a collision cell [67]. Some mass analyser combinations for tandem MS include quadrupole/ToF and linear ion trap/Orbitrap. Selection of a particular mass in the first analyser, followed by dissociation in the collision cell and detection in the second mass analyser can give information on subunit composition. Many different methods for dissociation exist including electron-capture dissociation (ECD) [68], electron-transfer dissociation (ETD) [69], surface-induced dissociation (SID) [70], and the most commonly used; collision-induced dissociation (CID) [71] and its

variant higher-energy collisional dissociation (HCD) [72]. In CID, which is the method used throughout this thesis, the analyte undergoes numerous collisions with inert gas molecules. During each collision, some of the ion's initial kinetic energy is converted into internal vibrational energy. For multi-protein complexes, a sufficient increase in internal energy can lead to breaking of the non-covalent bonds that bind the subunits, leading to complex dissociation. The number of collisions experienced by a protein complex is proportional to the gas pressure and protein collisional cross-section, which increases with complex size (and hence mass) [21]. The fraction of kinetic energy transferred to internal energy in one collision is greater for smaller proteins. As such, it can be possible for covalent bonds to be broken and dissociation to occur within single peptides, while for large complexes only complex dissociation occurs.

2.3.4.1 Peptide Fragmentation

In proteomics, fragmentation of the peptide backbone allows peptide sequencing for protein identification and also the accurate determination of PTM sites, with use of HCD allowing efficient peptide fragmentation [72]. Fragment ions are named according to the fragmentation site along the peptide backbone (Figure 2.6. [73, 74]), with b and y ions being the most common. To allow efficient MS analysis, proteins are hydrolysed to peptides using various enzymes, the most popular of which is trypsin as it is highly specific, cleaving the protein at the C-terminus of arginine or lysine residues. This generates peptides with a basic residue at the C-terminus, which is beneficial for protonation and as a result y ions are more common. Tryptic peptides are generally between 10–25 amino acids, resulting in masses amenable to MS analysis with mass analysers including Orbitrap mass analysers, now ubiquitous in the field of proteomics.

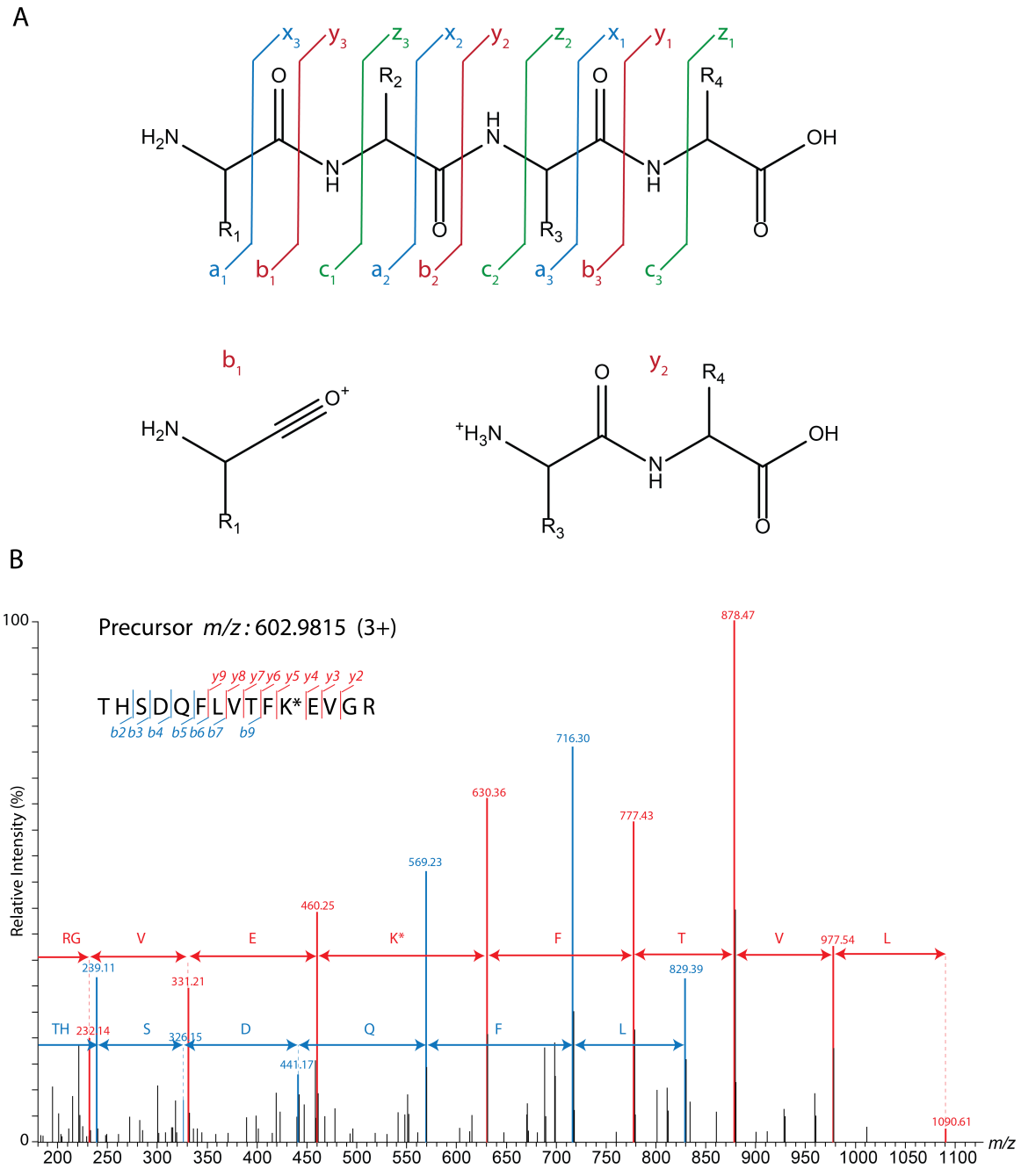


Figure 2.6: Fragmentation of the peptide backbone generates fragment ions. (A) The fragment ions are classified according to the fragmentation site along the peptide backbone [73, 74]. The most common ions formed are b and y ions, which are shown below the peptide. (B) An MS/MS spectrum with b and y ions labelled allows identification of the peptide sequence and the presence of an acetylated lysine residue (K*).

2.4 Ion Detection

The final step in acquisition of mass spectra, before data processing, is ion detection. The Orbitrap functions as both a mass analyser and a detector [63], removing the need for a separate detector. The Orbitrap detects ions using image current detection, with a Fourier transform required to obtain an m/z spectrum. As the ions oscillate along the axial direction, an image current is induced in the electrodes. As the frequency of motion is dependent on m/z (Equation 2.10), the coherent oscillations of ions with a series of m/z values will cause a superposition of oscillating image currents with a number of frequencies. By taking a Fourier transform of the axial image current, the frequencies of all oscillations within the Orbitrap can be extracted and converted into a mass spectrum [65]. Although the Orbitrap can function as a detector, this is not the case with many other mass analysers such as the quadrupole, linear ion trap, and ToF analysers. For this function, one of the most commonly-used detectors is the multi-channel plate (MCP) detector, which is a continuous dynode electron multiplier. An MCP is a disc constructed of resistive material, through which electron multiplier channels run. When operated in positive ion mode, an electrode is held at a negative voltage to which positive ions accelerate. The impact of the positive ions on the detector effects the emission of secondary particles, in this case electrons. The electrons hit the walls of the multiplier channel, releasing more electrons and amplifying the signal [75].

An important point to note in relation to ion transmission efficiency is that MS is not inherently quantitative [76]. The differing physicochemical properties of analytes mean that the ionisation efficiency of analytes can be highly variable, meaning there are large differences in the proportion of analytes producing desolvated gas-phase ions. This, together with variations in ion transmission

efficiency, with potentially large losses of ions at multiple points in the mass spectrometer, means that the overall response factors of analytes can vary significantly resulting in peak intensities that do not necessarily reflect the abundance of analytes in solution. Yet, often quantitative information is required, for example to measure changes in protein expression levels, to compare the formation of protein complex under differing conditions, and to accurately determine protein complex stoichiometries [77].

The use of isotopic labelling has become an effective tool to extract quantitative information from MS. Isotopes share similar chemical and physical properties, resulting in similar behaviour during sample analysis therefore resulting in similar response factors. Therefore comparison of signal intensities of labelled and non-labelled proteins can be compared to produce quantitative information. A number of different approaches to isotopic labelling have been developed including the incorporation of labels into proteins chemically, metabolically, and enzymatically [76].

Spiking of samples with known quantities of isotopically-labelled peptides in an absolute quantification (AQUA) strategy allows the accurate determination of the absolute amounts of proteins in samples by acting as an internal standard [78, 79]. Alternatively, metabolic labelling can be utilised to incorporate isotopic labels during cell culture, with ^{13}C- and ^{15}N-labelling and incorporation of isotopically-labelled essential amino acids in stable isotope labelling by amino acids in cell culture (SILAC) all popular approaches allowing the relative quantification of proteins under variable conditions [80, 81]. While multiple methods are available to extract quantitative information using MS, it is necessary to avoid interpreting signal intensities quantitatively in the absence of such strategies.

2.5 Instrumentation

2.5.1 Micromass Quadrupole-Time-of-Flight II Mass Spectrometer

The Quadrupole Time-of-Flight Mass Spectrometer (Q-ToF II, Micromass, Figure 2.7), modified for high m/z transmission [58], is one of the instruments used in this thesis for native MS measurements. Analytes are ionised in the source region equipped with a Z-spray source configuration such that the cone is positioned orthogonal to the electrospray capillary to reduce transmission of neutral species. Ions are focussed using multipole lenses to create a well-collimated ion beam and the pressure of various pumping stages can be controlled and fine-tuned to enable transmission of high m/z complexes by collisional cooling, as discussed earlier. Ions then enter the quadrupole, which can either be operated in RF-only mode, to allow

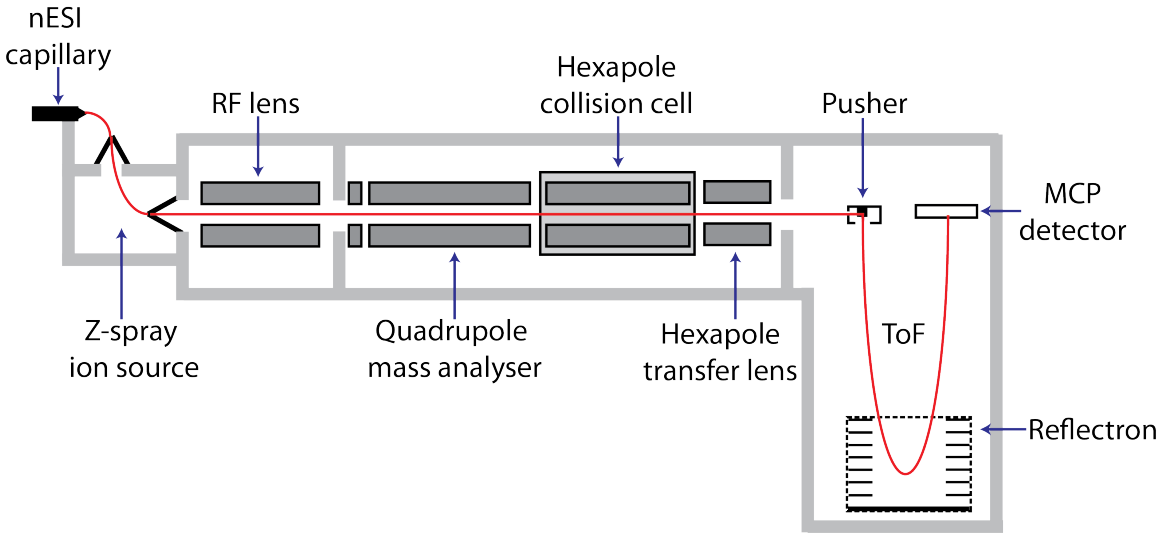


Figure 2.7: Schematic of the Micromass Q-ToF II Mass Spectrometer. Ions, generated by ESI, enter the mass spectrometer via a source in a Z configuration to reduce the transmission of neutral species. The first mass analyser, the quadrupole, can be used to select a particular range of masses before collisions in the hexapole collision cell. The instrument includes a second mass analyser, a ToF incorporating a reflectron, after which ions are detected by the MCP detector.

the transmission of ions with a wide range of m/z values, or alternatively the RF and DC voltages can be tuned to select a particular range of m/z values, for tandem MS analysis. The instrument used in this thesis has been modified by reducing the RF to 300 kHz to increase the upper m/z transmission range to 32,000 m/z [58]. After traversing the quadrupole, ions enter the hexapole collision cell where they may undergo a series of low-energy collisions with inert gas molecules to effect CID. Then, ions are analysed by the ToF reflectron mass analyser, equipped with a pusher that can be operated at frequencies up to 20 kHz allowing a full mass spectrum to be recorded every 50 μs. Finally, ions are detected and the signal amplified by an MCP detector.

2.5.2 Thermo Fisher Scientific LTQ Orbitrap XL Hybrid Ion Trap-Orbitrap Mass Spectrometer

The LTQ Orbitrap XL Hybrid Ion Trap-Orbitrap Mass Spectrometer (LTQ Orbitrap XL, Thermo Fisher Scientific, Figure 2.8) is one of the mass spectrometers used in this thesis for proteomics analysis and denaturing MS. Ions are generated by an electrospray ionisation source and transmitted to the first mass analyser, a linear quadrupole ion trap where ions can be selected and fragmented by CID, with detection enabled by two dynode electron multipliers located off-axis from the linear ion trap. Ions are transmitted to the C-trap, where they are stored for pulsed injection into the Orbitrap for analysis and detection.

2.5.3 Thermo Fisher Scientific Q Exactive Plus Hybrid Quadrupole-Orbitrap Mass Spectrometer

The Q Exactive Plus Hybrid Quadrupole-Orbitrap Mass Spectrometer (Q Exactive, Thermo Fisher Scientific, Figure 2.9) is similar to the LTQ Orbitrap XL in that

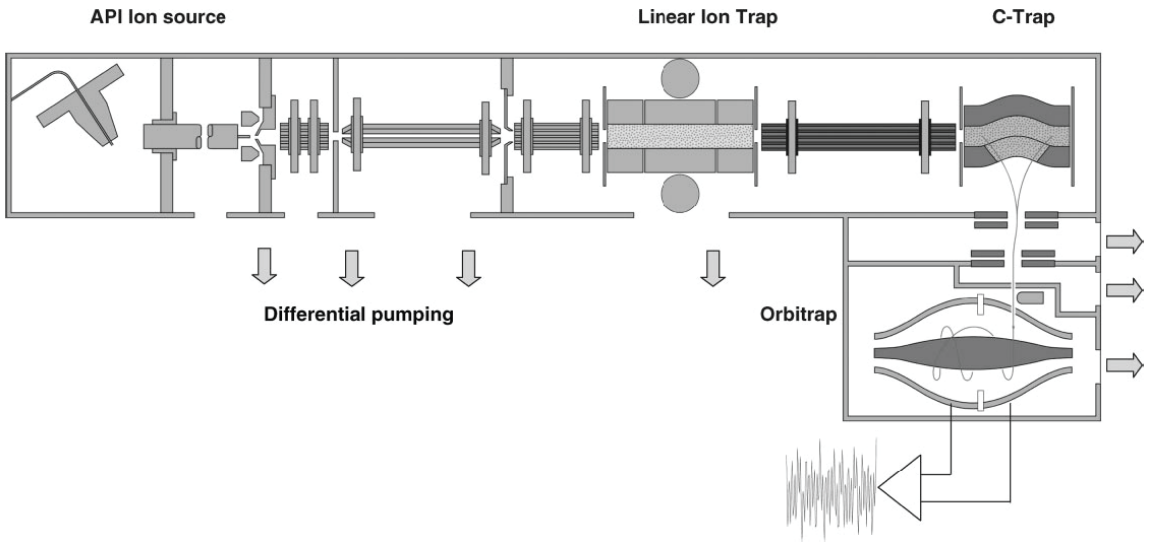


Figure 2.8: Schematic of the Thermo Fisher Scientific LTQ Orbitrap XL Hybrid Ion Trap-Orbitrap Mass Spectrometer. The instrument configuration incorporates an ESI source and an LTQ linear ion trap. A C-trap allows storage of ions for pulsed injection into the Orbitrap for mass analysis and detection. Figure from Thevis *et al.* [82].

they both use an Orbitrap mass analyser. After the ESI source, ion are accelerated in the bent flatapole designed to prevent transmission of neutral species. The first mass analyser is a segmented quadrupole, where mass selection can take place for tandem MS analysis. In an arrangement similar to the LTQ Orbitrap XL, the Q Exactive contains a C-trap for ion storage before injection into the HCD cell for fragmentation and then back into the Orbitrap for mass analysis and detection. The HCD cell of the Q Exactive Plus, utilised in this thesis, allows better fragmentation of peptides, including side-chain fragmentation of modified peptides, which is useful in PTM identification. A Q Exactive Plus Hybrid Quadrupole-Orbitrap Extended Mass Range (EMR) Mass Spectrometer with modifications to enable ultra-high mass range (UHMR) access [44], are also used for native MS experiments in this thesis. When used for native MS experiments, vacuum pressures in the various pumping stages can be tuned to enable transmission of high m/z analytes. Notable UHMR modifications include lowering of the RF value of the ion guides, enabling

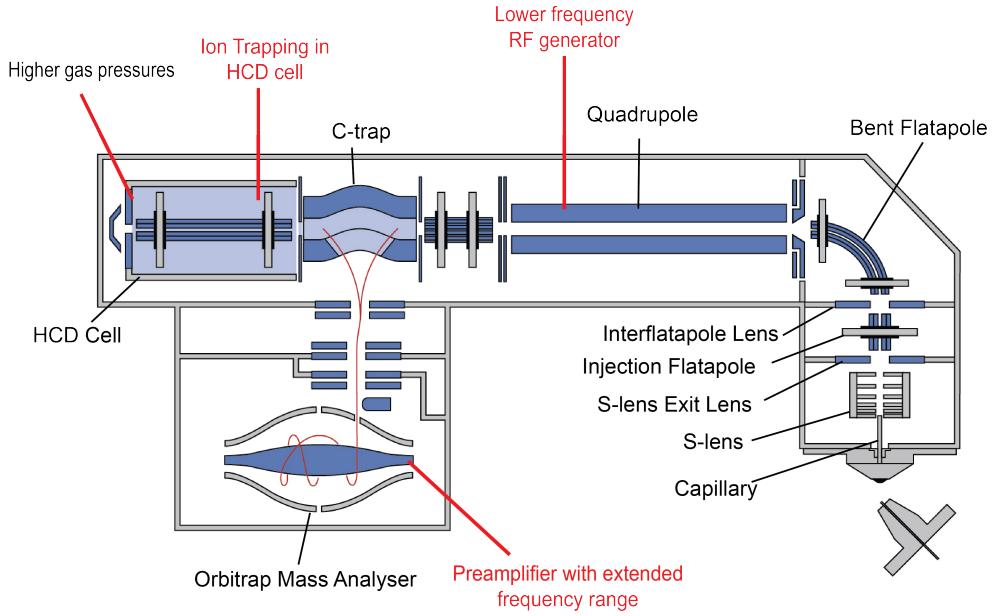


Figure 2.9: Schematic of the Thermo Fisher Scientific Q Exactive Plus Hybrid Quadrupole-Orbitrap Mass Spectrometer. Ions are generated in the ESI source and transmitted via a bent flatapole to remove neutral species. The first mass analyser, the segmented quadrupole allows selection of user-defined m/z ranges, for tandem MS. Ions are stored in the C-trap for pulsed injection into the HCD cell for fragmentation or the Orbitrap for mass analysis and detection. Figure from Gault *et al.* [83].

better focussing of high-mass ions, while better optimisation of the transport of high m/z ions from the C-trap to the Orbitrap enables higher ion transmission and together extend the upper mass range from 6,000 to 20,000 m/z [44].

2.6 Interpretation of Electrospray Mass Spectra

Mass spectra generated from electrospray ionisation exhibit several charge states resulting from the multiple basic sites available for protonation. The multiple charges observed are termed the charge state distribution and usually take the form of a Gaussian distribution. In native MS, protein folding reduces the surface area of the protein and as such reduces the number of sites available for protonation [84]. Folded proteins therefore exhibit narrow charge state distributions at lower charge

than their denatured, unfolded counterparts and the charge state can thus be used as an indication of unfolding. The observed charge state can also be influenced by many other factors including instrument parameters such as applied voltages and solution conditions such as pH and salts present [85–88].

The multiple charge states observed with ESI may seem to present a challenge in terms of mass determination, however the relation of m/z of neighbouring peaks allows the mass and charge to determined (Figure 2.10). Taking the m/z of two adjacent peaks in protonated spectra and expressing their m/z in terms of the molecular mass M and charge z gives the equations

$$x_1 = \frac{M + z_1}{z_1} \tag{2.11}$$

where x_1 is the m/z of the first peak, and

$$x_2 = \frac{M + z_2}{z_2} \tag{2.12}$$

where x_2 is the m/z of the second adjacent peak. The charges of adjacent peaks are related

$$z_2 = z_1 + 1. \tag{2.13}$$

One can eliminate M by relating Equations 2.11 & 2.12 to get

$$z_2 = z_1 \frac{x_1 - 1}{x_2 - 1}. \tag{2.14}$$

Substituting in Equation 2.13 gives

$$z_1 = \frac{x_2 - 1}{x_1 - x_2}, \tag{2.15}$$

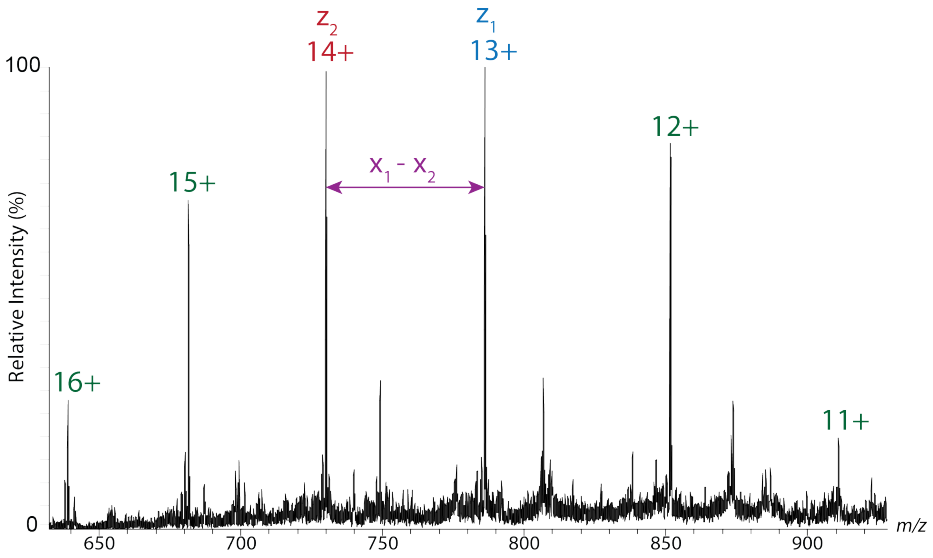


Figure 2.10: Example mass spectrum illustrating the calculation of molecular mass M from two adjacent charge states. Taking the m/z values of two adjacent peaks (x_1=786.19, x_2=730.11) allows the charges (z_1=13+, z_2=14+) and the molecular mass M (10,208 Da) to be calculated.

hence from Equation 2.11 we get

$$M = \frac{(x_2 - 1)(x_1 - 1)}{x_1 - x_2} \tag{2.16}$$

allowing M to be solved by knowing only the m/z of two adjacent peaks [15].

2.7 Liquid Chromatography

In proteomics studies, direct infusion of peptide mixtures into the mass spectrometer leads to complicated spectra with ion suppression due to competition between multiple peptides. Temporal separation of peptides using liquid chromatography (LC) therefore allows a greater number of components to be analysed and detected, without the most intense mass peaks dominating all mass spectra. As ESI-MS requires a liquid analyte, coupling to LC is an excellent way to achieve sample separation. LC of proteins is a technique in which proteins (or

peptides) in solution (the mobile phase) are applied to a column (the stationary phase) and differentially interact with the column surface. There are multiple different types of LC, including gel filtration, ion-exchange, and affinity-based techniques such as immunoaffinity and metal affinity chromatography. Therefore, LC is a widely-used technique, particularly useful in protein purification and proteomics studies.

The partitioning of analyte between the mobile phase and stationary phase can be described using the theoretical plate model [89]. The chromatographic column can be considered as a large number of small segments called theoretical plates. In this theory, immediate equilibration occurs between mobile and stationary phases, with this equilibration occurring at each theoretical plate along the length of the column. In this model, the best results in terms of separation are obtained with a high number of theoretical plates N, or a low height equivalent to a theoretical plate ($HETP$) which is described

$$HETP = \frac{L}{N}, \tag{2.17}$$

where L is the length of the column [90].

In reality, equilibration between stationary and mobile phases is not immediate and in addition multiple paths are available to analytes within the column, both leading to some degree of broadening of the analyte's elution profile (band broadening). A more accurate depiction of analyte partitioning can therefore be described by the van Deemter equation [91]

$$HETP = A + \frac{B}{u} + Cu \tag{2.18}$$

where u is the average velocity of the mobile phase and A, B, and C are band broadening factors; A being Eddy diffusion on account of the multiple paths available to analytes, B being longitudinal diffusion on account of analyte diffusion from the centre of bands to the edges, and C being resistance to mass transfer which reflects the time taken for equilibration between the mobile and stationary phase [90].

Different modes of chromatography are available, with Normal Phase LC employing a polar stationary phase and a less polar mobile phase, while in Reversed Phase LC (RPLC) the mobile phase is more polar than the stationary phase. In proteomics, RPLC is commonly used and is coupled directly to the mass spectrometer allowing separation of peptides before MS analysis, based on their polarity, thus reducing the complexity of the resulting mass spectra. Typically, a hydrophobic stationary phase such as C18 is covalently bound to a silica support. The mobile phase is composed of water and a less polar solvent, such as acetonitrile (ACN) which are mixed on a gradient to create a decreasingly-polar mobile phase through the course of the LC separation. Often a trap column and analytical column are employed in tandem; in this way the trap column allows the loading of a high volume of dilute sample which can be concentrated and washed to remove contaminants such as salts before analytical separation. Further types of chromatography can be used together with HPLC to enrich and purify samples. One example is the use of TiO_2 columns in the selective enrichment of phosphopeptides which are usually present at lower abundancy than their non-phosphorylated counterparts, hindering identification [92].

2.8 Cross-linking: Structural Biology Approaches at the Peptide Level

Chemical cross-linking allows the three-dimensional structure of proteins to be probed via linkage of amino acid residues in close proximity by covalent bonds. The protein can then be hydrolysed to peptides using enzymes, as used in bottom-up proteomics, and analysed by LC-MS with cross-linked peptides indicating amino acid residues in close proximity in the native conformation. Many different types of cross-linking reagents have been developed but some of the most popular are the homo-bifunctional amine-reactive N-hydrosuccinimide (NHS) esters, which target primary amines such as ϵ-amino groups on the side chains of lysine residues and the protein N-terminus. Manipulation of the structures of NHS esters can produce cross-linking reagents which are hydrophobic, such as disuccinimidyl suberate (DSS) (Figure 2.11A) or more water-soluble, such as bis[sulfosuccinimidyl] suberate (BS3) (Figure 2.11B) allowing both soluble and membrane protein structure to be interrogated. Furthermore, cross-linkers with alternative side-chain reactivity and spacer arm lengths [93] have been developed allowing a greater breadth of experiments to be accessed.

As the more hydrophobic NHS esters, such as DSS, are insoluble in the aqueous buffers used to maintain native protein structure, the more water-soluble sulfo-NHS esters (i.e. BS3) are more routinely employed. Mixtures of deuterated and non-deuterated cross-linkers, such as BS3-d0/d4 are often used to aid in the validation of cross-linked peptide mass spectra as a result of the characteristic mass shift that can be observed due to deuteration. The resulting space restraints created by cross-linking analysis can be used to determine the structure and conformation of proteins [94], to ascertain the location of protein subunits in protein complexes [95],

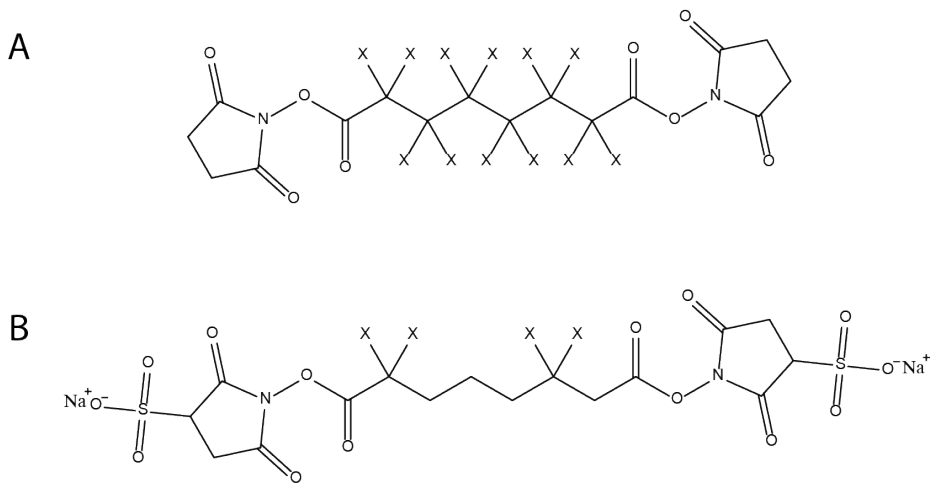


Figure 2.11: The structure of two commonly-used NHS ester cross-linking reagents. (A) DSS cross-linker, with a spacer arm length of 11.4 Å. X denotes either hydrogen, for DSS-d0, or deuterium for DSS-d12. (B) The water-soluble sulfo-NHS ester, BS3. X denotes hydrogen for BS3-d0 and deuterium for BS3-d4.

and to consider the solvent accessibility of specific amino acid residues [96].

Cross-linking offers many advantages over other methods in the determination of protein structure and interactions. These advantages include low sample requirements and the lack of upper bounds on complex size. Furthermore, cross-linking is a solution-phase technique, allowing proteins to remain in native-like buffers where their native structures and interactions can be best preserved. In addition, cross-linking is an ensemble technique, providing a sample of the protein conformations present in solution. This heterogeneity, along with the dynamics of protein complexes, which can be problematic with many techniques, does not preclude cross-linking interpretation, notwithstanding the additional complexity they bring to data analysis [17].

2.8.1 Comparative Cross-linking Mass Spectrometry

The availability of deuterated cross-linkers has allowed the development of further cross-linking strategies which can be employed to quantify changes in protein interactions and structure in response to various conditions. Comparative cross-linking consists of cross-linking a control protein sample with non-deuterated 'light' cross-linker and a treated protein sample (i.e. dephosphorylated) with the deuterated 'heavy' cross-linker [97, 98]. The cross-linking reactions can be quenched and the samples mixed 1:1 and processed together by bottom-up MS. Integration of the resulting MS peaks allows direct comparison between the abundance of individual cross-links in the control and the treated samples. This method has been used successfully to assess the effect of dephosphorylation on the spinach F-type ATP synthase [99], the effect of nucleotide on Hsp70 conformation [100] and to further our understanding of the final maturation step of the proteasome lid [98].

As with many techniques, in addition to its many advantageous features, chemical cross-linking has limitations. The presence of multiple interconverting protein conformations along with the dynamic nature of protein interactions and flexible protein regions can lead to complicated data analysis and creation of unphysical models. Although it has been speculated that many structural artefacts could form during the course of chemical cross-linking experiments, including the formation of non-physiological protein conformations induced by the cross-linker as well as cross-linking of two randomly passing but non-interaction protein complexes, no experimental evidence has shown any of these to occur [17]. Data analysis can be a significant impediment with analysis proving time-consuming and requiring a high level of user input in the form of visual inspection of mass spectra. This becomes an even greater hindrance in the case of comparative cross-linking where further work is required for peak integration. Many data analysis programmes have been

developed to assist in cross-linking data analysis and automate much of this process [98, 101–103], yet significant user input is still required to ensure high quality results and prevent false positives, particularly during comparative cross-linking analysis.

Chapter 3

The Role of Post-Translational Modifications in the Dimerisation of the 70 kDa Heat Shock Protein

3.1 Introduction

In order to function properly, most proteins must fold correctly into a three-dimensional structure known as the native state. This native state is encoded in the amino acid sequence of the protein. For small proteins this can occur spontaneously through the quick burial of hydrophobic segments, as demonstrated by their refolding after removal from denaturing agents *in vitro* [105]. On the contrary, larger multi-domain proteins often form stable intermediate states, which are both non-functional and leave disordered or hydrophobic residues exposed,

Parts of this chapter have been published in: Nina Morgner*, Carla Schmidt*, Victoria Beilsten-Edmands*, Ima-obong Ebong, Nisha A. Patel, Eugenia Clerico, Elaine Kirschke, Soumya Daturpalli, Sophie E. Jackson, David Agard, and Carol V. Robinson, *Cell Reports*, 2015 [100]. *Denotes equal contribution. Parts of this chapter have been summarised in Carla Schmidt, Victoria Beilsten-Edmands, and Carol V. Robinson, *Oncotarget*, 2015 [104]. Purified $Hsp70_{Sf9}$ was provided by Dr Elaine Kirschke, Zygy Roe-Zurz, and Professor David Agard (The University of California, San Francisco) and phosphosite analysis was performed by Professor Carla Schmidt (The Martin Luther University of Halle-Wittenberg).

leading to aggregation. This aggregation can be irreversible, preventing proteins from accessing functional states and increasing aggregation of other protein species. Therefore, it is necessary for proteins to receive assistance, in the form of molecular chaperones, to correctly fold to their native states, in order to avoid misfolding and aggregation. Despite the presence of molecular chaperones to assist in the folding of nascent polypeptide chains, some aggregation and misfolding is present in cells. This can be elevated under stress conditions, such as with elevated temperature. As a consequence, molecular chaperones are additionally required to assist non-nascent proteins in refolding [106].

3.2 Protein Folding

A protein chain is able to adopt a very large number of conformations. In 1968, Cyrus Levinthal remarked that given the number of possible conformations available to a polypeptide chain, it would not be possible for proteins to fold by random sampling of each state on biological timescales [107]. The theoretical timescales for protein folding based on such a model do not match what is observed experimentally [108, 109], with many proteins folding on timescales between milliseconds and minutes [110]. These timescales for protein folding obtained experimentally are realised on account of the relatively small configurational space sampled, compared to the possible search area. Mechanisms of protein folding have been suggested which seek to explain this discrepancy with Bagchi and co-workers proposing that energetic factors introduce small biases in the folding trajectory, such that the protein passes through a relatively small number of protein configurations on its way to the native state [111, 112].

In the current theory of protein folding, proteins explore a potential energy surface

reaching the native state by progression down a potential energy funnel, via folding intermediates, albeit with decreasing configurational entropy, forming increasing numbers of native interactions (Figure 3.1). The potential energy landscape is not smooth and proteins must therefore traverse a number of kinetic barriers on their way to the native state. Breaking these interactions can be rate-limiting. Whilst the folding of small proteins can occur spontaneously on very short time-scales *in vitro*, the folding of large proteins to a native conformation can become a long, complicated process with folding times of hours and in some cases proteins do not arrive at their native folded state [106, 113].

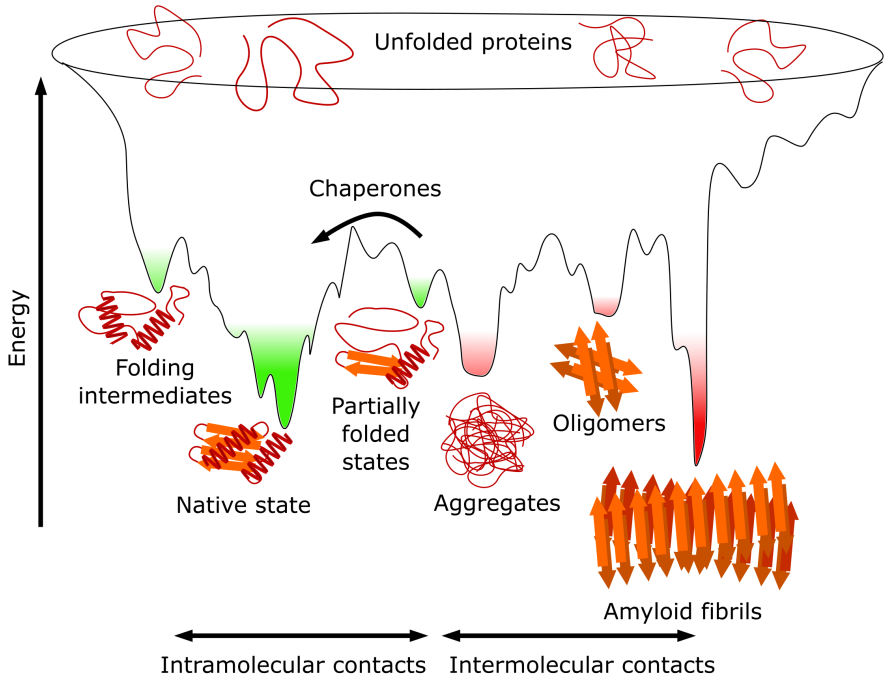


Figure 3.1: Free-energy surface of protein folding and aggregation. Proteins explore different conformations on the energy landscape. Favourable interactions bias the search of configurational space towards the native state, with decreasing potential energy but reduced configurational entropy. Proteins can become trapped in non-native conformations, occupying potential energy wells. Molecular chaperones assist in the correct folding and unfolding of misfolded states to avoid the toxic aggregation of proteins. Figure based on Hartl *et al.* [106].

The tendency of large proteins to undergo hydrophobic collapse in particular presents a problem as this can lead to the formation of large globules or transiently-stable, partially-folded states [114]. These misfolded proteins may contain non-native interactions which must be broken to progress further to a native conformation. Moreover, these misfolded and partially-folded proteins can aggregate, increasingly so at higher concentrations, on account of their exposed hydrophobic residues [115]. Furthermore, the highly-crowded environment *in vivo*, containing macromolecular concentrations of 300–400 mg mL^{-1} [116, 117], compounds the problem by increasing the likelihood of protein aggregation though excluded-volume effects and increasing the affinities of protein–protein interactions [116, 118]. These aggregates are usually amorphous but can form amyloid fibrils which are characterised by a cross-β-structure and are associated with numerous diseases such as Alzheimer's disease (amyloid–β), Huntington's disease (aggregated huntingtin), and Parkinson's disease (aggregated α-synuclein) [115]. With increasing age comes a decreased efficiency in protein homeostasis (proteostasis) machinery [119] and as such the occurrence of aggregation diseases such as Alzheimer's are a significant health issue in ageing populations [120].

3.3 Molecular Chaperones

In order to maintain proteostasis, cells employ a system of protein quality control in which a family of proteins known as the molecular chaperones acts to assist in the correct folding, unfolding, and refolding of nascent polypeptide chains, misfolded, and partially-folded proteins. Molecular chaperones, often described as heat shock proteins, are subdivided into different classes named according to molecular weight, for example the 90 kDa heat shock protein (Hsp90). Heat shock proteins are

upregulated in response to stress, such as elevated temperature, when the prevalence of misfolding and aggregation increase [106]. In addition to their role in the folding of nascent polypeptides and the unfolding/refolding of misfolded states, molecular chaperones are also involved in the assembly of large complexes, protein transportation, and degradation [121]. The molecular chaperones that assist in *de novo* protein folding and refolding are ATP-dependent chaperones. They operate by a kinetic partitioning mechanism (Figure 3.2), in which the chaperone binds to exposed hydrophobic residues of the client (substrate) protein, the presence of which is indicative of non-native proteins. The chaperones then assist in folding via repeated cycles of hydrophobic residue binding and release. This repetitive binding mechanism hinders aggregation through protection of exposed hydrophobic segments and a reduction in the concentration of aggregation-prone material [121]. Substrate release allows fast-folding molecules to refold and bury hydrophobic peptide segments and consequently these chaperones are referred to as foldases. Proteins that require more time to fold may be handed over to the chaperonins for

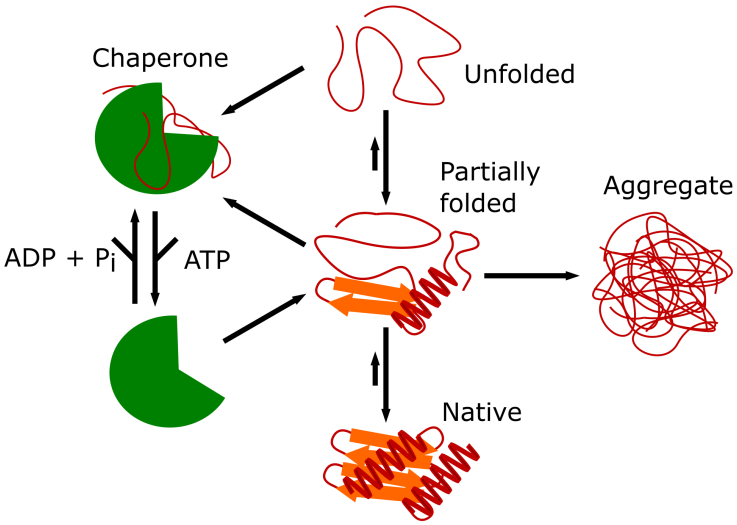


Figure 3.2: Kinetic partitioning of non-native states prevents aggregation and assists in correct folding. Figure based on Kim *et al.* [121].

folding or sent down a different pathway to degradation. In this chapter I will focus on one of these foldase chaperones, the 70 kDa heat shock protein (Hsp70).

3.3.1 70 kDa Heat Shock Protein

The Hsp70 family is an essential component in a range of cellular protein-folding and chaperoning events and in the cytosol includes the constitutively-expressed (Hsc70) and the stress-inducible forms (Hsp70). Thirteen different Hsp70 proteins have been identified in humans, in addition to four related Hsp110 proteins [122]. Hsp70 (Figure 3.3) contains a nucleotide-binding domain (NBD), so called on account of its cleft for nucleotide binding, which is coupled to a substrate-binding domain (SBD) via a flexible linker. The SBD is comprised of a β-substrate-binding pocket and an α-helical lid which together bind, trap, and stabilise a substrate in the binding pocket [123]. The allosteric coupling of the NBD with the SBD is integral to the function of Hsp70. Binding of ATP to the NBD closes the nucleotide-binding cleft which opens the helical lid of the SBD, with separate docking of the SBD and lid to the NBD. In this ATP-bound conformation (Figure 3.3A), the open lid leaves hydrophobic segments of the substrate-binding pocket accessible to the substrate. Exposed hydrophobic regions of substrate are able to bind transiently to the substrate-binding pocket before being quickly released, as the ATP-bound conformation has low affinity for the substrate [123]. Substrate binding, along with the action of co-chaperones such as Hsp40, stimulate the ATPase activity of Hsp70 [124], catalysing conversion of ATP to ADP. Hydrolysis of ATP to ADP induces a conformational change, with narrowing of the nucleotide-binding cleft and a concomitant undocking of the SBD and lid to form a closed-lid conformation [125]. This closed-lid conformation results in a high affinity for bound substrate in the Hsp70 ADP-bound conformation (Figure 3.3B). The tighter binding of substrate seems to assist in

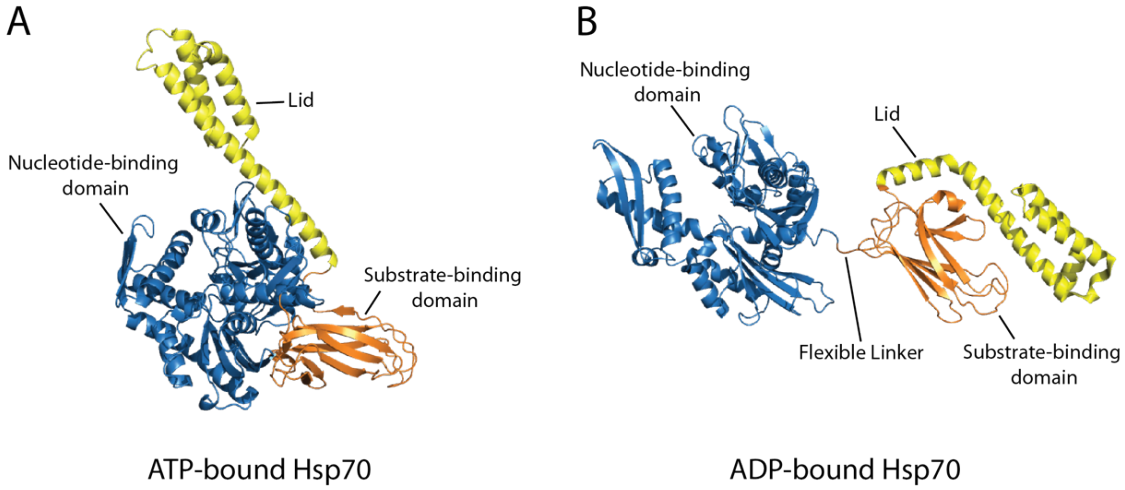


Figure 3.3: Hsp70 interconverts between two conformations as nucleotide binds to the nucleotide-binding domain (blue). (A) The ATP-bound state exposes the hydrophobic residues of the substrate-binding cleft (orange) allowing hydrophobic segments of client proteins to bind transiently. The low affinity for substrate leads to rapid substrate release in the ATP-bound state. However, ATP hydrolysis to ADP induces a conformational change to the ADP-bound structure (B) which exhibits a high affinity for substrate and closes the lid of the binding pocket (yellow) trapping the substrate. Nucleotide exchange of ADP for ATP opens the lid, releasing the substrate.

protein folding and prevention of aggregation. The rate of ATP hydrolysis, and therefore transition to the high-substrate-affinity state, is rapidly increased by interaction with Hsp40, the J-domain of which also functions to recruit substrates for Hsp70 [123]. Cytosolic eukaryotic Hsp70 additionally contains a glycine/proline-rich C-terminal region with an EEVD motif that is able to bind to other proteins, including tetratricopeptide repeat (TPR) domain-containing proteins such as Hsp organising protein (HOP) and C-terminal Hsp70-interacting protein (CHIP) [106]. This EEVD domain has additionally been shown to be important in maintaining cooperation between Hsp40 and Hsp70 in eukaryotic cells [126].

These interactions between Hsp70 and its many co-chaperones and binding partners

mean that Hsp70 takes part in a plethora of cellular events. Although Hsp70 plays a role in *de novo* protein folding and mediation of proper folding, Hsp70 is also known to assist in translocation of cargo proteins [127, 128], the targeting of client proteins to the ubiquitin-proteasome system for degradation in conjunction with CHIP [129], and to facilitate activation of a number of receptors and transcription factors [130]. One of the best known examples of Hsp70's involvement in receptor activation is its role in the maturation of the steroid hormone glucocorticoid receptor (GR), where Hsp70 operates in conjunction with multiple other proteins including the aforementioned Hsp90, Hsp40, and HOP [131, 132].

While Hsp70 forms interactions with many of its co-chaperones and client proteins, it also self-associates to form Hsp70 oligomers. Hsp70 and its homologues from across many species have been found to oligomerise into dimers and other higher-order oligomers [133], including inducible human Hsp70 [133–135] and constitutive Hsc70 in spinach plants [136], rat [137–141], murine [142], and bovine [143], in addition to BiP/Grp78 in murine [144], hamster [145–147], and bovine [148], and the bacterial DnaK [149–151]. The incidence of Hsp70 oligomerisation across many species indicate that this may in fact be a conserved feature across the whole Hsp70 family. A functional role for the formation of Hsp70 oligomers is not fully understood but studies have provided insight in this area, both with indications that oligomerisation serves to regulate the chaperone cycle by altering the level of monomeric Hsp70 available for foldase activity [134, 138] and that dimer formation is necessary for Hsp70's interactions with the co-chaperone Hsp40 [151]. Not only is the function of Hsp70 oligomers unclear, but the mechanism and structure are also uncertain, with various studies reporting alternative options for both specific dimers [151, 152] and higher-order oligomers [133].

Hsp70 has been shown to have a protective effect against protein-based aggregation diseases such as Huntington's and Parkinson's disease [153, 154] with high levels of Hsp70 associated with longevity in humans [155]. While Hsp70 seems to play a positive role in maintaining good health in these cases, Hsp70 also appears to maintain viability in cancer cells by protecting against apoptosis [156–158]. These seemingly contradictory behaviours of Hsp70 in supporting good health mean that although Hsp70 is a target for clinical therapies, these therapies must be carefully guided to strike a balance between reducing its promotion of cancer cell viability and supporting its role in protection against aggregation diseases. As a result, there is great interest in understanding Hsp70's interactions with other proteins as a means to finding ways to regulate and fine-tune these interactions.

The protein-protein interactions of Hsp70 are in many cases modulated by PTMs [159]. Studies have repeatedly shown that Hsp70 can display a broad array of PTMs, yet the effect of many of these PTMs has not been studied in detail. Despite this, some studies have provided some insight, in particular into the role of Hsp70 phosphorylation. By careful use of site-directed mutation to produce phosphomimic variants, phosphorylation of Tyr525 has been shown to increase cell survival and promote relocalisation of Hsp70 to the nucleus, in response to heat shock [160]. A second study showed that incubation of Hsp70 with kinases *in vitro* resulted in phosphorylation of the C-terminus which altered the binding affinities of many of Hsp70's co-chaperones [161]. Furthermore, this study also showed an increase in phosphorylation in cancer cells which were bound preferentially to HOP rather than CHIP, supporting the view that cancer cells require an emphasis on client folding mediated by Hsp70, Hsp90, and HOP, rather than ubiquitin-mediated degradation via CHIP [161, 162]. Besides phosphorylation, other PTMs have been reported for Hsp70 including methylation, which has been shown to influence Hsp70

translocation and is upregulated in many cancer cells [163], and acetylation, which appears to control degradation of oncogenic proteins [164].

3.4 Research Aims

Hsp70 plays a vital role in proteostasis by ensuring the correct folding of proteins. Multiple studies of Hsp70 have reported the formation of dimers and higher-order oligomers, along with reporting an important role for Hsp70 dimerisation in efficient interaction with Hsp40 and suggesting a regulatory role for Hsp70 oligomers. Despite these advances, the Hsp70-Hsp70 interactions that promote this oligomerisation, along with the functional significance of such oligomers, are currently unclear. Furthermore, despite the awareness surrounding Hsp70's many PTMs, studies into their functional roles are scarce. Yet, there is great potential for these PTMs to have important functional significance, especially with respect to Hsp70's interactions with other proteins and possibly even between Hsp70 and itself.

Therefore, in this chapter the oligomerisation state of human Hsp70 will be examined. A main aim is to identify PTMs present on Hsp70 and explore the influence these can have on protein-protein interactions within the context of Hsp70 self-association. On account of previous reports identifying key functional roles for Hsp70 phosphorylation, a particular emphasis here will be on phosphorylation, without neglecting other PTMs of known importance such as acetylation. Overall, by considering PTMs and their role in regulating Hsp70 interactions, this research will provide insights into the fine-tuning that takes place within the Hsp70 chaperone system and leads to the great diversity of Hsp70's roles *in vivo*.

3.5 Results

3.5.1 The Dependence of Hsp70 Dimerisation on the Expression System

As oligomers of Hsp70s from various species have been previously reported [133–151] but much is unknown about their function and regulation, the oligomeric state of Hsp70 was investigated by native MS. Human Hsp70 was expressed in two expression systems; prokaryotic *E. coli* ($Hsp70_{E.coli}$) and eukaryotic Sf9 insect cells ($Hsp70_{Sf9}$) and analysed by native MS. Purified $Hsp70_{Sf9}$ was provided by Dr Elaine Kirschke, Zygy Roe-Zurz, and Professor David Agard (The University of California, San Francisco). For this comparison, an isotopically-labelled $Hsp70_{E.coli}$ was purified, so that the two variants could be distinguished, thereby allowing direct comparison of $Hsp70_{E.coli}$ and $Hsp70_{Sf9}$ peak intensities in the same spectrum. First, $Hsp70_{E.coli}$ was analysed alone by MS under experimental conditions devised to preserve non-covalent interactions [165]. Under these conditions, $Hsp70_{E.coli}$ was observed predominantly as a monomer, with a minimal population of dimer present (Figure 3.4A). A series of shoulder peaks was attributed to bound nucleotide. Next, isolated $Hsp70_{Sf9}$ was similarly analysed by native MS. Monomeric $Hsp70_{Sf9}$ was observed (Figure 3.4B), with evidence of additional shoulder peaks attributed to nucleotide loss similar to the $Hsp70_{E.coli}$ sample. However, $Hsp70_{Sf9}$ additionally shows the presence of a significant population of $Hsp70_{Sf9}$ dimer.

To directly compare peak intensities in order to verify similar protein concentrations, a 1:1 mixture of $Hsp70_{E.coli}$ labelled with heavy isotopes and $Hsp70_{Sf9}$ with natural abundance isotopes was analysed by native MS. The mass spectrum (Figure 3.4C) reveals that a higher proportion of $Hsp70_{Sf9}$ dimer is

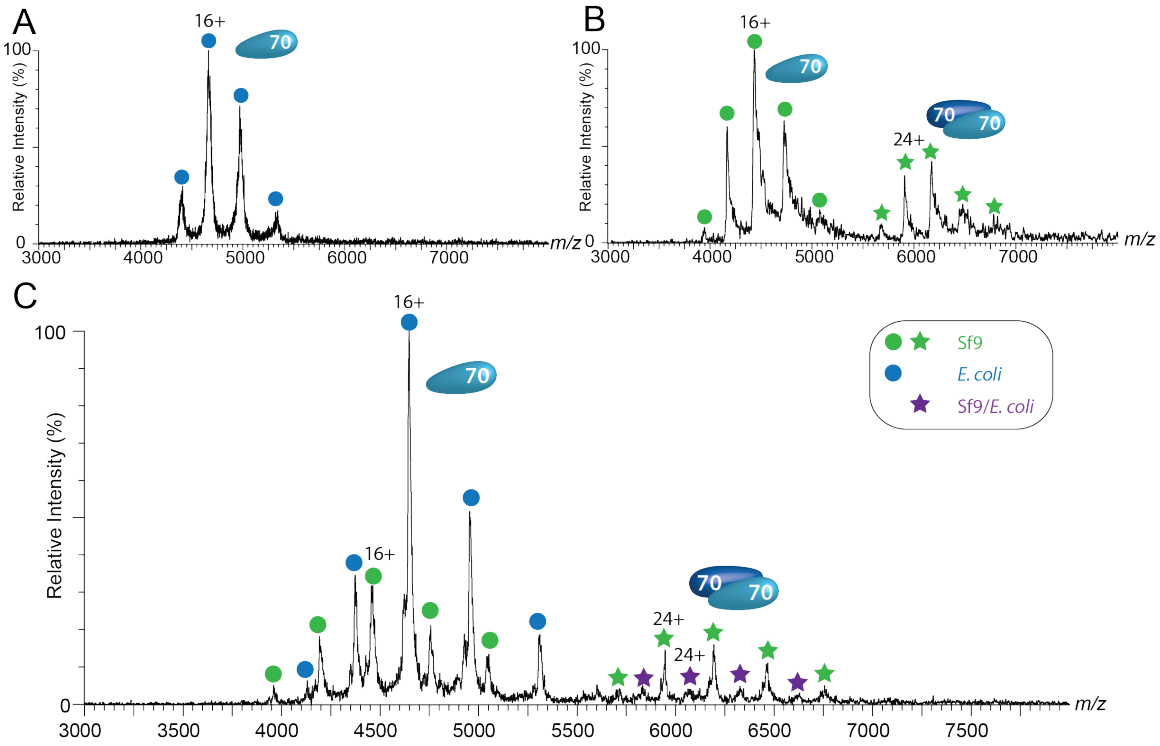


Figure 3.4: Mass spectra of (A) Hsp70 expressed in *E. coli* and (B) Hsp70 expressed in Sf9 cells. Both variants contain monomeric Hsp70 ($Hsp70_{E.coli}$, blue circles, 74,327 ±5 Da and $Hsp70_{Sf9}$, green circles, 71,168 ±5 Da). The Sf9 variant additionally contains a significant population of dimer (green stars, 141,834 ±39 Da), which is absent in the *E. coli* spectrum. (C) When the two variants are mixed in a 1:1 ratio, the Sf9 dimer persists, but a small population of the heterodimer $(Hsp70_{E.coli})_1(Hsp70_{Sf9})_1$ (purple stars, 145,658 ±60 Da) is observed. Overall, a greater proportion of dimeric $Hsp70_{Sf9}$ is observed. The two variants have similar charge distributions.

present than $Hsp70_{E.coli}$ dimer, with each variant displaying similar monomer:dimer ratios when analysed alone (Figures 3.4A and 3.4B) and together (Figure 3.4C). Interestingly, an additional feature was detected in the dimer *m/z* region; the heterodimer $(Hsp70_{E.coli})_1(Hsp70_{Sf9})_1$ consistent with formation by subunit exchange, indicating that $(Hsp70_{E.coli})_1(Hsp70_{Sf9})_1$ heterodimer formation is favoured over $Hsp70_{E.\ coli}$ dimer formation.

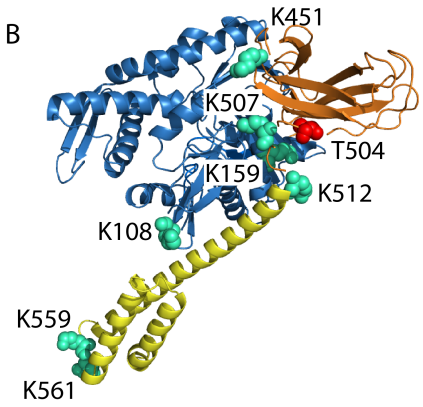
3.5.2 Identification of Post-Translational Modifications on Hsp70

This discrepancy in dimerisation of Hsp70 between the two expression systems was unexpected. As both Hsp70 variants are composed of the human amino acid sequence, one possibility is that post-translational modifications (PTMs) could be involved, on account of their frequent role in regulating protein-protein complex stability [166–168]. In particular, phosphorylation and acetylation, often in combination, can serve to stabilise or destabilise protein-protein interactions [169]. Furthermore, eukaryotic Sf9 cells are able to perform modifications, such as phosphorylation and acetylation, whereas recombinant expression in *E. coli* often does not lead to such PTMs on account of the limited PTM machinery in *E. coli*. In addition, comparison of the expected mass of $Hsp70_{Sf9}$, calculated from this sequence, to the mass observed by MS places $Hsp70_{Sf9}$ at approximately 1,159 Da heavier than expected, indicating some degree of post-translational modification (i.e. phosphorylation = +80 Da and acetylation = +42 Da) in addition to the expected nucleotide binding (Appendix Table A.1). On the other hand, the measured mass of $Hsp70_{E.coli}$ fits well with the expected mass in addition to nucleotide binding.

To assess the PTM status of Hsp70, proteins were digested with trypsin and the tryptic peptides analysed by LC-MS/MS. After database searching and manual validation of spectra, seven lysine acetylation sites were identified (Figure 3.5), five of which had not previously been reported. A phosphopeptide enrichment strategy with TiO_2 was utilised by Professor Carla Schmidt (The Martin Luther University of Halle-Wittenberg), resulting in the identification of a phosphosite at Thr504 (*p*-Thr504). This phosphosite was particularly interesting as a literature search revealed that *p*-Thr504 is highly conserved across many eukaryotes and located in a known phosphorylation 'hot-spot' [170]. Moreover, mutational studies have shown

A

Acetylation Site	Peptide Sequence
108	VQVSYK*GETK
159	QATK*DAGVIAGLNVLR
451	AMTK*DNNLLGR
507	ITITNDK*GR
512	LSK*EEIER
559	SAVEDEGLK*GK
561	GK*ISEADKK



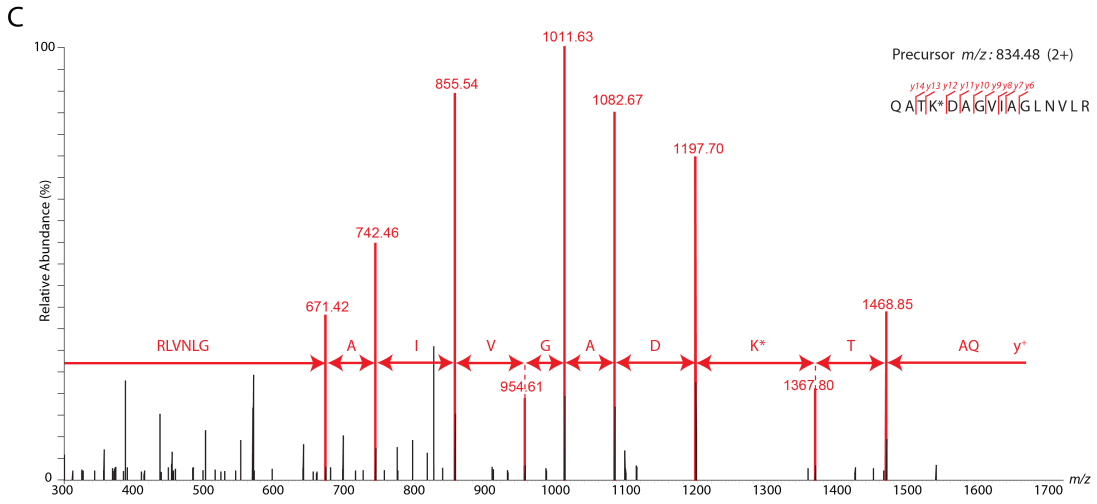


Figure 3.5: (A) Identified acetylation sites in Hsp70$_{Sf9}$. In total, seven acetylation sites and one phosphorylation site were identified, of which five acetylation sites had not previously been identified. (B) Identified phosphorylation and acetylation sites displayed on a high-resolution structure of ATP-bound Hsp70. (C) An example MS/MS spectrum identifying the acetylation site Lys159 with y ions indicated (red). Figure adapted from Morgner *et al.* [100].

that phosphorylation of this 'hot-spot' in yeast is important to the function of Hsp70 [170]. Location of the identified PTMs on a structure of Hsp70 in the ATP state (Figure 3.5B) reveals that the *p*-Thr504 is located near to the hinge of the SBD β-domain and the SBD α-helical lid in a lysine-rich pocket while acetylation sites are located both near this hinge region and also on the α-helical lid near the C-terminus.

3.5.3 The Effect of Hsp70 Dephosphorylation on Dimerisation

To examine the effect of phosphorylation on Hsp70 dimerisation, a phosphatase enzyme was used. A 1:1 mixture of $Hsp70_{E.coli}$ and $Hsp70_{Sf9}$ was treated with a phosphatase enzyme in order to dephosphorylate the protein, and the phosphatase removed before native MS analysis. A control sample was also incubated under the same conditions but without phosphatase. After incubation in the absence of phosphatase, analysis by native MS revealed that both monomer and dimer were still present (Figure 3.6A). An additional peak series was observed and again attributed to *apo*- and nucleotide-bound Hsp70. A heterodimer of $(Hsp70_{E.coli})_1(Hsp70_{Sf9})_1$ formed in the control sample as before (Figure 3.4C) but with greater intensity, consistent with more extensive subunit exchange occurring during longer incubation times. When the proteins are incubated with phosphatase (Figure 3.6B), an additional peak series was again observed with the Hsp70 monomer peaks but significantly, very little dimer was observed with a reduction in both homodimer and heterodimer populations. This shows that dephosphorylation of Hsp70 leads to a dramatic reduction in dimerisation. In addition, the presence of a heterodimer implies that modifications on a single $Hsp70_{Sf9}$ are able to stabilise the dimer interface, albeit not to the same extent as in the case of the $Hsp70_{Sf9}$ homodimer. The heterodimer is also depleted significantly in the phosphatase-treated sample, implying that phosphorylation is necessary to stabilise the dimer interface in both the homodimer $(Hsp70_{Sf9})_2$ and the heterodimer $(Hsp70_{Sf9})_1(Hsp70_{Sf9})_1$.

To ensure that $Hsp70_{E.coli}$ was unaffected by the phosphatase enzyme, as would be expected on account of its lack of phosphosites, samples of $Hsp70_{E.coli}$ were incubated with and without phosphatase under conditions where a small population

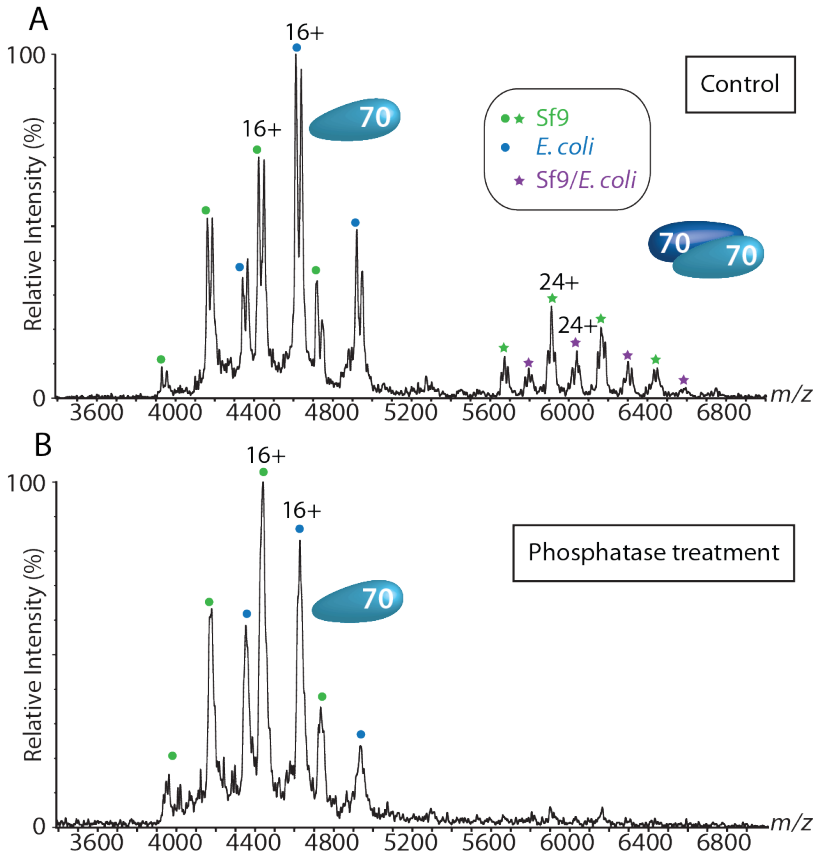


Figure 3.6: Mass spectra of $Hsp70_{E.coli}$ and $Hsp70_{Sf9}$ incubated with and without phosphatase enzyme. (A) After incubation in the absence of phosphatase, nucleotide turnover of monomeric $Hsp70_{Sf9}$ (green circles, 70,795 ±8 Da/71,224 ±6 Da) and $Hsp70_{E.coli}$ (blue circles, 73,861 ±6 Da/74,259 ±5 Da) is observed. Populations of $(Hsp70_{Sf9})_2$ (green stars, 141,955 ±22 Da) and an $(Hsp70_{E.coli})_1(Hsp70_{Sf9})_1$ heterodimer (purple stars, 144,996 ±30 Da) are observed. (B) Incubation in the presence of phosphatase enzyme results in a significant decrease in the population of dimer observed with an increase in the $Hsp70_{Sf9}$ monomer (green circles, 71,127 ±39 Da) relative to the $Hsp70_{E.coli}$ monomer (blue circles, 74,203 ±28 Da). Figure adapted from Morgner *et al.* [100].

of dimer was observed. After incubation, the mass spectra show that the proportion of dimer in the control (Figure 3.7A) and phosphatase-treated samples (Figure 3.7B) were comparable. These results confirm that treatment with the phosphatase enzyme has a dimer-destabilising effect on $Hsp70_{Sf9}$ but not $Hsp70_{E.coli}$.

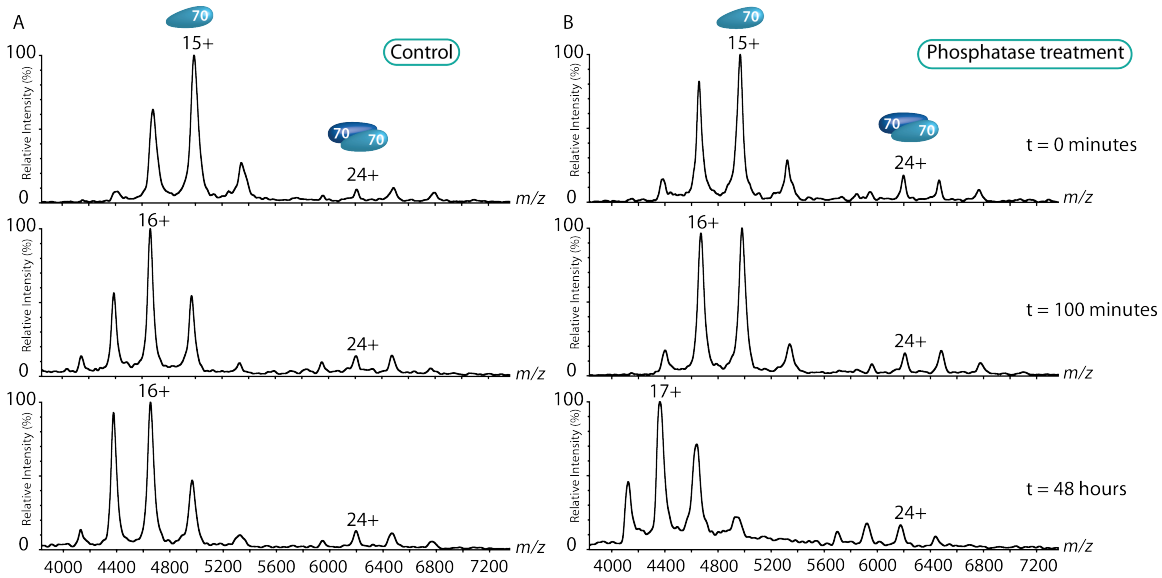


Figure 3.7: Hsp70$_{E.coli}$ incubated in the absence and presence of phosphatase enzyme. (A) In the absence of phosphatase Hsp70 monomer (74,655 ±79 Da) a small population of dimer (148,977 ±31 Da) is observed, with the proportion of dimer not significantly changed after 48 hours incubation. (B) In the presence of phosphatase, Hsp70 monomer (74,511 ±27 Da) and a small population of dimer (148,776 ±80 Da) are observed, with the dimer population also not significantly changed after 48 hours incubation. Figure adapted from Morgner *et al.* [100].

3.5.4 The Effect of Ionic Strength on Hsp70 Dimerisation

Given the location of the PTMs identified (Figure 3.5B) and the effect of dephosphorylation on the Hsp70$_{Sf9}$ dimer, it is plausible that ionic interactions could be playing a role in stabilisation of a protein-protein interface. If ionic interactions are important in stabilising the Hsp70$_{Sf9}$ dimer, then logically increasing the ionic strength of the solution should shield ionic interactions and lead to a decrease in the dimer population present. Aliquots of Hsp70$_{Sf9}$ were therefore incubated in solutions supplemented with increasing concentrations of KCl from 50 mM to 300 mM, thereby increasing the ionic strength. Samples were analysed by MS under instrument parameters designed to favour transmission of higher mass species. In the lowest ionic strength buffer (50 mM KCl) a large population of dimer is observed (Figure 3.8A). On increasing the ionic strength, the population of dimer

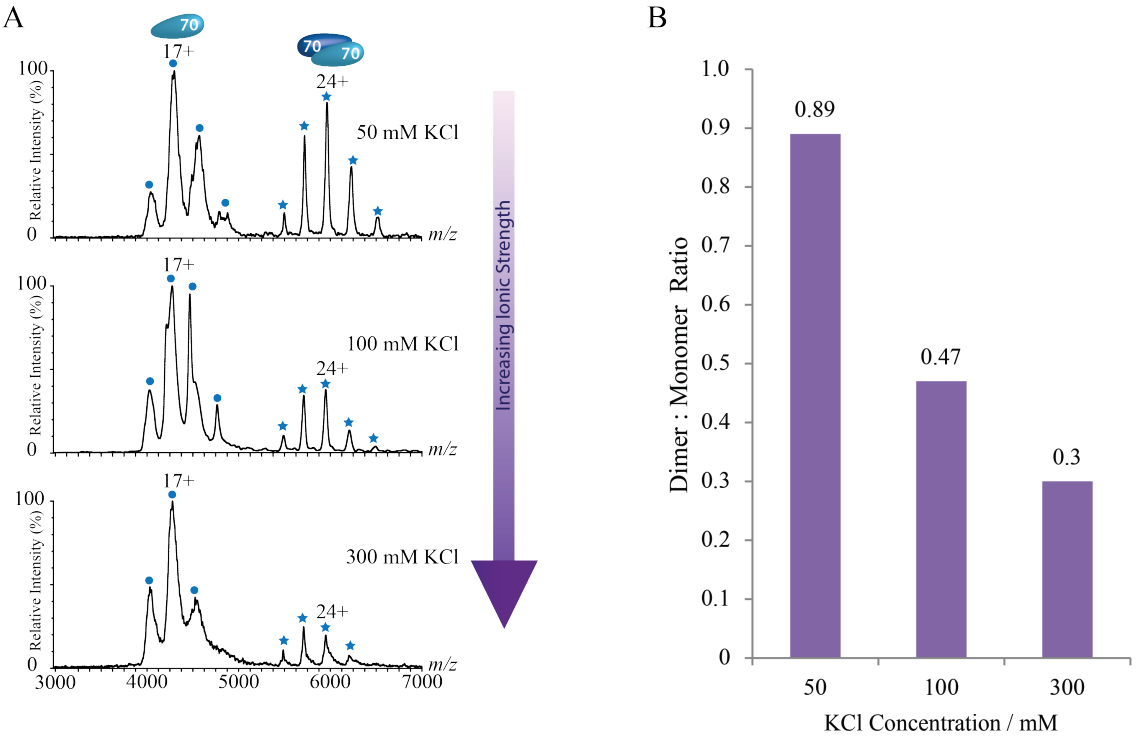


Figure 3.8: (A) Mass spectra of Hsp70$_{Sf9}$ incubated in solutions of increasing ionic strength from solutions containing 50 mM KCl to 300 mM KCl. Monomeric (blue circles, 71,258 ±8 Da) and dimeric (142,952 ±49 Da) Hsp70$_{Sf9}$ are observed. As ionic strength increases, the proportion of dimeric Hsp70$_{Sf9}$ decreases. (B) The ratio of signal intensity of dimer vs. monomer observed in the mass specta calculated by peak fitting with UniDec software [171]. The ratio decreases as the ionic strength of the incubating buffer increases. Figure adapted from Morgner *et al.* [100].

decreases, with the smallest population of dimer observed at the highest ionic strength. The ratio of monomer:dimer peak intensities observed was estimated using UniDec software [171] (Figure 3.8B). The correlation of decreased Hsp70 dimer population with increased ionic strength supports a role for ionic interactions, and potentially phosphorylation, in stabilisation of the dimer interface.

3.5.5 The Dimerisation of an Hsp70 Phosphomimic

As the phosphatase treatment of Hsp70$_{E.\ coli}$ and Hsp70$_{Sf9}$ indicated that global dephosphorylation reduces dimerisation of Hsp70$_{Sf9}$, the role of the identified

Thr504 phosphorylation site located in the phosphorylation 'hot-spot' [170] was investigated. A phosphomimic variant was produced in *E. coli* by replacing the threonine residue Thr504 with glutamic acid via site-directed mutagenesis to produce the $Hsp70_{T504E}$ variant. In order to assess only the effect of the phosphomimic mutation on dimerisation, the (isotopically-labelled) wild type and the phosphomimic T504E, both expressed in *E. coli*, were mixed 1:1 and analysed by native MS, under conditions designed to allow transmission of high mass complexes (Figure 3.9). In the low mass region (3,900–5,200 m/z), monomeric $Hsp70_{E.\ coli}$ and $Hsp70_{T504E}$ were observed with similar intensities. In the higher m/z region, homodimers of $(Hsp70_{E.\ coli})_2$ and $(Hsp70_{T504E})_2$ along with the heterodimer $(Hsp70_{E.coli})_1(Hsp70_{T504E})_1$ were detected, however the theoretical ratio of 1:1:2 expected if the three species formed equally, was not observed. Although

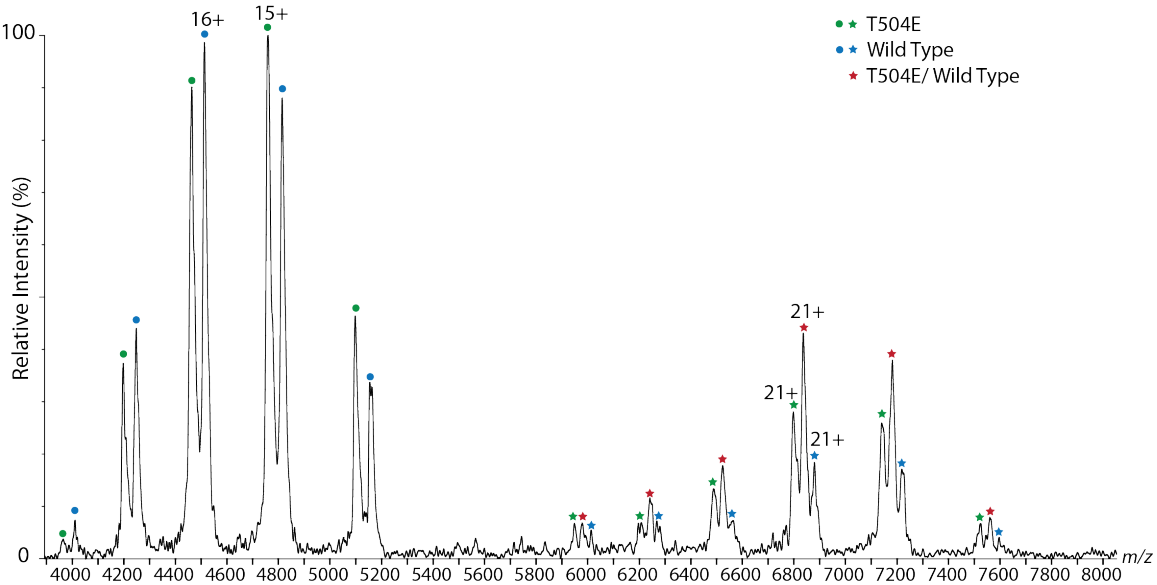


Figure 3.9: $Hsp70_{T504E}$ incubated with isotopically-labelled $Hsp70_{E.coli}$. Populations of $Hsp70_{T504E}$ monomer (green circles, 71,538 ±9 Da) and $Hsp70_{E.coli}$ monomer (blue circles, 72,398 ±5 Da) are similar. A population of the heterodimer $(Hsp70_{T504E})_1(Hsp70_{E.coli})_1$ is observed (red stars, 143,770 ±13 Da). A higher proportion of dimer is observed for the phosphomimic variant $Hsp70_{T504E}$ (green stars, 142,990 ±33 Da) than the isotopically-labelled wild type $Hsp70_{E.coli}$ (blue stars, 144,589 ±41 Da). Figure adapted from Morgner *et al.* [100].

both the wild type $Hsp70_{E.coli}$ and the phosphomimic $Hsp70_{T504E}$ form dimers, the population of dimeric $Hsp70_{T504E}$ was elevated. This supports a role for phosphorylation of Thr504 in stabilising the dimer interface.

3.5.6 The Effect of Mutation of the Substrate-Binding Domain on Dimerisation

Thus far, it has been shown that phosphorylation, including phosphorylation of Thr504, stabilises the Hsp70 dimer. The structure and mode of dimerisation however have not been discussed. The role of Hsp70 *in vivo* is to bind client proteins and therefore the most obvious explanation might be that the dimerisation of Hsp70 observed here is occurring via a substrate-binding model. That is, one Hsp70 monomer acting as the client, binding to the SBD of a second Hsp70 copy (Figure 3.10). A likely candidate for the substrate would be the hydrophobic linker that connects the NBD to the SBD. On the contrary, a recent study reported

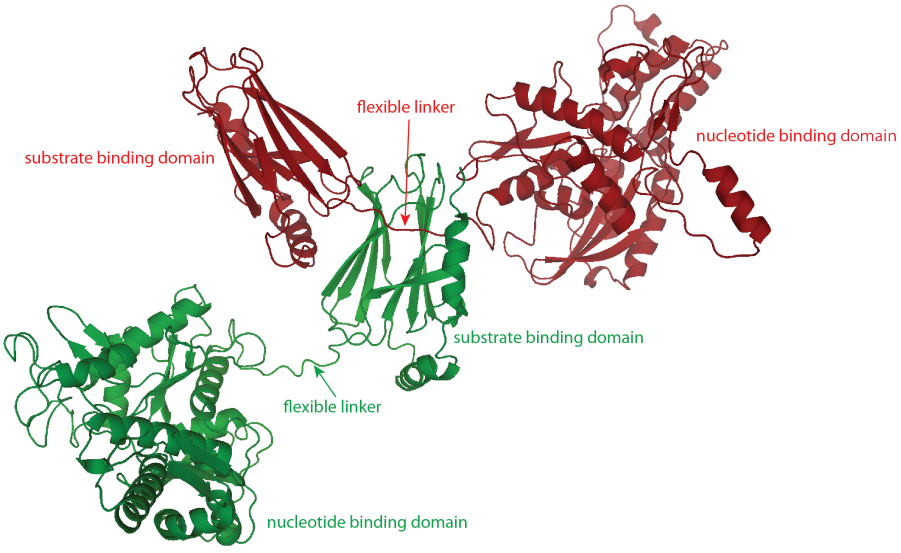


Figure 3.10: Model of Hsp70 dimerisation via a substrate-binding-type interaction. One Hsp70 acts as the substrate (red) and binds via the substrate-binding domain of the second Hsp70 (green). Figure adapted from Morgner *et al.* [100]

binding of two substrate peptides to Hsp70 dimers [133], which would not be possible if Hsp70 dimerisation proceeded via a substrate-binding model. To clarify this and distinguish if the Hsp70 dimers detected by MS in this study are dimerising via the SBD (that is, the substrate-binding model) or via another mechanism, a substrate-binding-deficient mutant was prepared. This mutant was prepared by sterically blocking the hydrophobic substrate-binding pocket via substitution of a bulky phenylalanine residue for the valine in the binding groove (Figure 3.11), a mutation which has been shown to significantly reduce the ability of Hsp70 to bind substrate [172]. This mutation was incorporated into the phosphomimic to produce the double mutant T504E V438F which exhibits both substrate-binding deficiency and the dimer-promoting phosphomimic behaviour.

The isotopically-labelled double mutant T504E V438F ($Hsp70_{T504E\,V438F}$) was incubated with the phosphomimic $Hsp70_{T504E}$ and analysed by MS to directly

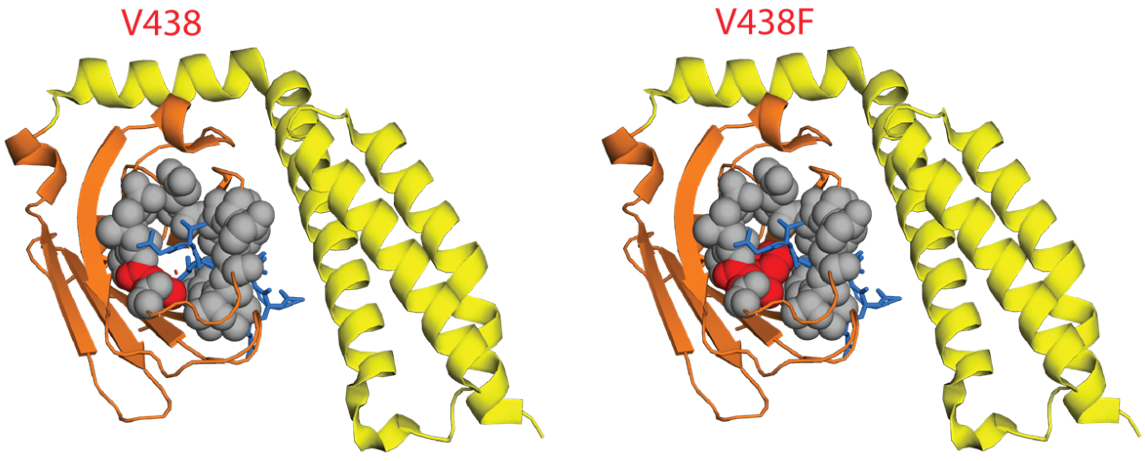


Figure 3.11: The substrate-binding domain in complex with a substrate peptide (PDB ID: 4PO2 [173]) with the β-subdomain (orange) and the α-helical lid (yellow). The substrate peptide (blue) is shown bound within the substrate-binding cavity (grey spheres). Left, the valine residue (red) in the wild type protein allows the substrate to bind in the cavity. Right, the valine residue is mutated to a bulky phenylalanine residue (red) which sterically blocks the substrate-binding pocket. Figure based on Mayer *et al.* [172].

compare the extent of dimer present in each sample (Figure 3.12). Both monomer and dimer were observed in the mass spectrum for both variants with both homodimers $(\text{Hsp70}_{T504E})_2$ and $(\text{Hsp70}_{T504EV438F})_2$, and the heterodimer $(\text{Hsp70}_{T504E})_1(\text{Hsp70}_{T504EV438F})_1$ present. Importantly, similar levels of the two homodimers were observed, with no large decrease in dimerisation for the substrate-binding-deficient mutant. A slight shift in the charge state distribution to lower charges was observed in $\text{Hsp70}_{T504EV438F}$, which may be due to a small change in surface accessibility. These results eliminate the possibility of a substrate-binding model for Hsp70 dimerisation.

3.6 Discussion

The results presented here demonstrate that dimerisation of human Hsp70 can differ depending on the expression system, with Hsp70_{Sf9} showing more dimerisation than $\text{Hsp70}_{E.coli}$. When $\text{Hsp70}_{E.coli}$ and Hsp70_{Sf9} are incubated together, the heterodimer $(\text{Hsp70}_{E.coli})_1(\text{Hsp70}_{Sf9})_1$ forms. It would be expected that statistically a 1:2 ratio of Hsp70_{Sf9} dimer and heterodimer $(\text{Hsp70}_{E.coli})_1(\text{Hsp70}_{Sf9})_1$ would be formed, however this is not the case with more $(\text{Hsp70}_{Sf9})_2$ forming than the heterodimer. Although the presence of heterodimer indicates that a single Hsp70_{Sf9} is able to favour dimer formation, it is not sufficient to favour dimer formation to the levels observed for two Hsp70_{Sf9} molecules.

Extended incubation led to an increase in the population of heterodimer on account of a longer timescale over which subunit exchange could take place. The population of heterodimer is still lower than expected statistically showing that even after extended incubation, the formation of the $(\text{Hsp70}_{Sf9})_2$ homodimer is more favourable than the heterodimer. The change in the ratio of the intensities of

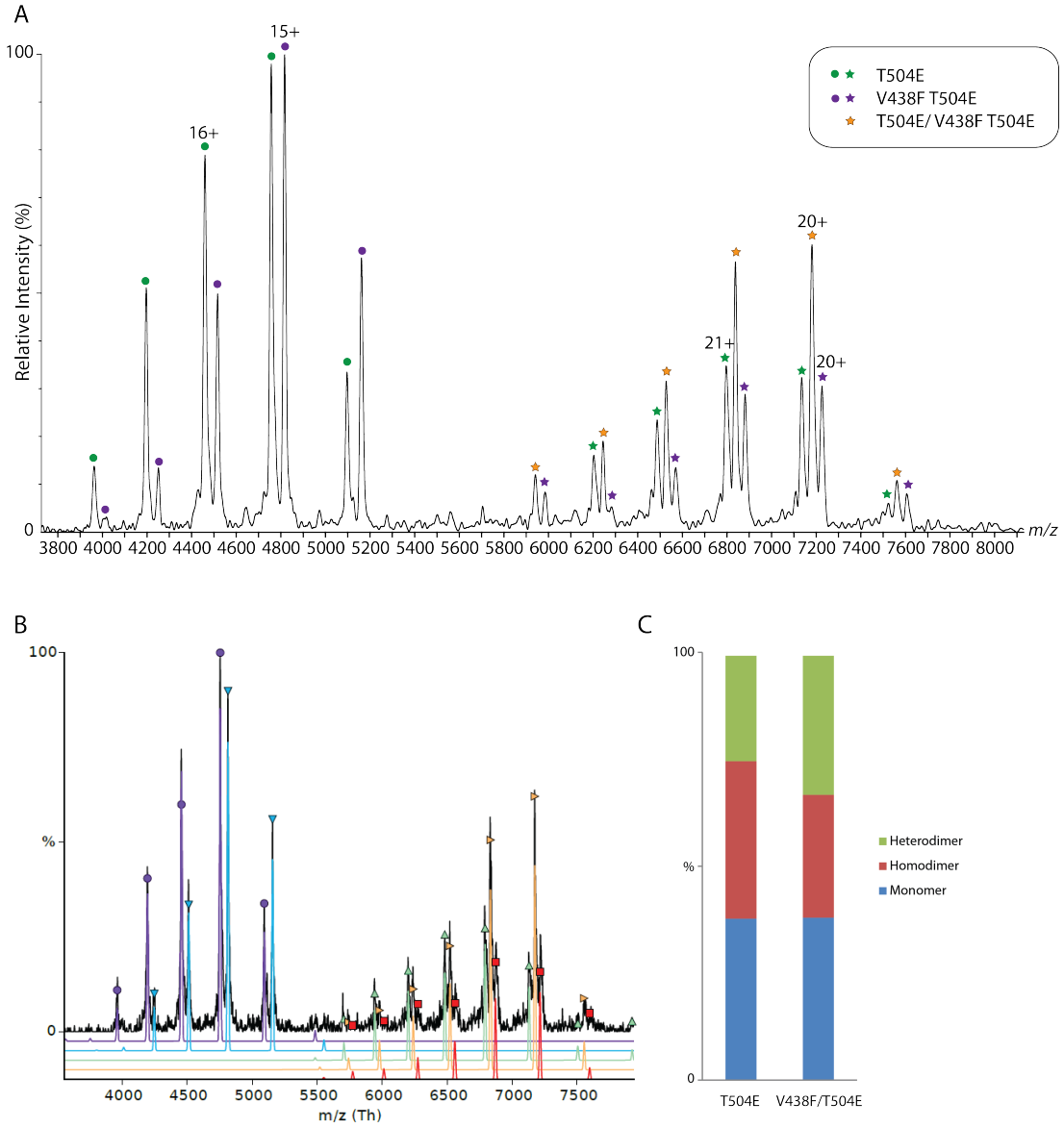


Figure 3.12: (A) Mass spectrum of the substrate-binding-deficient phosphomimic variant Hsp70$_{T504E\,V438F}$ incubated with the phosphomimic Hsp70$_{T504E}$. Populations of monomeric Hsp70$_{T504E}$ (green circles, 71,256 ±7 Da) and Hsp70$_{T504E\,V438F}$ (purple circles, 72,167 ±8 Da) are observed. A population of the heterodimer (Hsp70$_{T504E}$)$_1$(Hsp70$_{T504E\,V438F}$)$_1$ is observed (orange stars, 143,412 ±9 Da). Homodimers are observed for both Hsp70$_{T504E}$ (green stars, 142,538 ±9 Da) and Hsp70$_{T504E\,V438F}$ (purple stars, 144,311 ±11 Da). (B) Peak fitting of the mass spectrum using UniDec software [171] allows quantitation of species. (C) The percentage of each Hsp70 variant (Hsp70$_{T504E}$ and Hsp70$_{T504E\,V438F}$) present in each species was calculated. Comparing the calculated intensities shows that the monomer:dimer ratio of each variant is similar. There is no significant decrease in the population of Hsp70$_{T504E\,V438F}$ dimer in comparison to the Hsp70$_{T504E}$ dimer suggesting that dimerisation does not occur via the substrate-binding pocket. Figure adapted from Morgner *et al.* [100].

monomeric Hsp70$_{E.coli}$ and Hsp70$_{Sf9}$ shifts towards lower intensities of Hsp70$_{E.coli}$, which is consistent with subunit exchange of dimeric Hsp70$_{Sf9}$ with an Hsp70$_{E.coli}$ monomer. Due to the proposed role of PTMs in stabilising the dimer, it would be interesting to consider the distribution of PTMs between the monomer and dimer populations of Hsp70$_{Sf9}$. Unfortunately, this approach is hampered by the presence of multiple nucleotide-bound states (ATP/ADP/nucleotide-free) resulting in multiple peaks which, together with the heterogeneity of PTM states and limited instrument resolution, leads to significant peak overlap and does not yield useful insights into this distribution. Analysis on higher resolution instruments, developed during the course of this investigation [44], may be more fruitful in this regard, however given the high level of heterogeneity inherent in a nucleotide-bound post-translationally modified sample, this may also be unsuccessful. A number of other techniques however could prove useful in determining the PTM status amongst the monomer and dimers, for example native gel electrophoresis coupled with proteomics or a top-down MS approach.

This raises the question of the impact of nucleotide state on dimerisation. Hydrolysis of ATP to ADP results in a significant change in conformation of Hsp70 and therefore might be expected to have a significant effect on dimerisation. Considering the distribution of ADP-bound and nucleotide-free Hsp70$_{Sf9}$ in the phosphatase control mass spectrum (Figure 3.6A) with its clear distribution of approximately 1:1 ADP-bound:nucleotide-free Hsp70$_{Sf9}$ monomer, statistically one would expect a 1:2:1 ratio of $(ADP)_0$:$(ADP)_1$:$(ADP)_2$-bound Hsp70$_{Sf9}$ dimer. This is in fact what is observed, with a slight preference for the $(ADP)_2$-bound dimer. Based on similarity with *E. coli* DnaK, the structures of ADP-bound and nucleotide-free Hsp70 are highly similar [124]. As such, this statistical distribution of homodimer and heterodimer fits well with what might be expected based on

conformational similarity. What is more difficult to determine is the expected distribution of ADP- and ATP-bound Hsp70. Reported Hsp70$_{Sf9}$ cross-linking data carried out in excess ADP, ATP, and non-hydrolysable ATP showed that in solution these nucleotide-bound states interconvert rapidly, yet evidence for dimerisation of both ADP- and ATP-bound Hsp70 was found with this approach [100] indicating dimers form in both conformations.

Incubation with phosphatase significantly reduced dimerisation of Hsp70$_{Sf9}$, but also reduced the heterodimer (Hsp70$_{E.\ coli}$)$_1$(Hsp70$_{Sf9}$)$_1$ population showing that phosphorylation is important in stabilising both the Hsp70$_{Sf9}$ homodimer and the heterodimer. Comparing the ratios of monomeric Hsp70$_{E.coli}$ to Hsp70$_{Sf9}$ before and after phosphatase treatment is consistent with the increase in Hsp70$_{Sf9}$ on account of the dissassembly of Hsp70$_{Sf9}$ dimers. The additional peak series, attributed to nucleotide turnover during incubation, appear different in the presence and absence of phosphatase, with better resolution between the peaks seen in the absence of phosphatase. Close inspection of the peaks reveals that in fact three species are present; ADP-bound, ATP-bound and nucleotide-free Hsp70. In the absence of phosphatase, most of the monomeric Hsp70 is either nucleotide-free or ADP-bound, with only small quantities of ATP-bound monomer observed. After phosphatase treatment, the additional peak series is less well resolved, although the reason for this is not clear. One possibility could be due to less efficient desolvation of the analyte ions in the phosphatase-treated sample, resulting in less well-resolved peaks.

Incubation of Hsp70$_{E.coli}$ with phosphatase did not decrease the population of dimer detected by MS. An interesting point to note is the gradual turnover of nucleotide, with ATP-bound Hsp70 being the predominant species at t=0, and increasing levels of ADP-bound and nucleotide-free species with increasing incubation time. In spite

of this, the ratio of monomer:dimer stays relatively constant at all time points, indicating that the ATP- and ADP-bound Hsp70 dimers form at similar levels. This observation is contradictory to a substrate-binding dimerisation model, as one would expect a significant level of dimerisation in the ADP-bound state and very little if any in the ATP-bound state. In the mass spectra of $Hsp70_{E.coli}$ with and without phosphatase, there was a gradual shift in charge state towards higher charges; from 15+ to 17+ in the monomer and from 23+ to 25+ in the dimer. This may be due to the long incubation times used and could be on account of the larger surface area of the more extended ADP-bound conformation [174], the population of which increases with incubation, relative to the ATP-bound conformation.

Comparison of the phosphomimic $Hsp70_{T504E}$ with wild type $Hsp70_{E.coli}$, both expressed in *E. coli*, showed an increase in dimerisation for the phosphomimic. The relative increase in dimerisation however is not as significant as that seen when comparing $Hsp70_{E.coli}$ with $Hsp70_{Sf9}$. A number of possible factors could explain this discrepancy. Firstly, the phosphomimic is generated by mutation of a threonine residue to glutamic acid, in order to mimic a phosphothreonine residue. Comparison of the structures of these three residues (Thr, Glu, *p*-Thr; Figure 3.13), shows that Glu does not perfectly replicate the *p*-Thr residue, most notably by the difference in charge (2- for *p*-Thr vs. 1- for Glu). If the dimer interface is stabilised by ionic interactions, for example, then this interaction would be weaker for Glu than *p*-Thr. Secondly, the significant decrease in dimerisation observed with dephosphorylation compared to the smaller increase observed for the phosphomimic indicates that there may be more than one phosphorylation site contributing to stabilising the dimer. As Thr504 is located in a highly-conserved phosphorylation 'hot-spot', multiple phosphoryated residues could be present in the vicinity of T504E, or elsewhere on the protein that may contribute to complex stability. Finally, it is

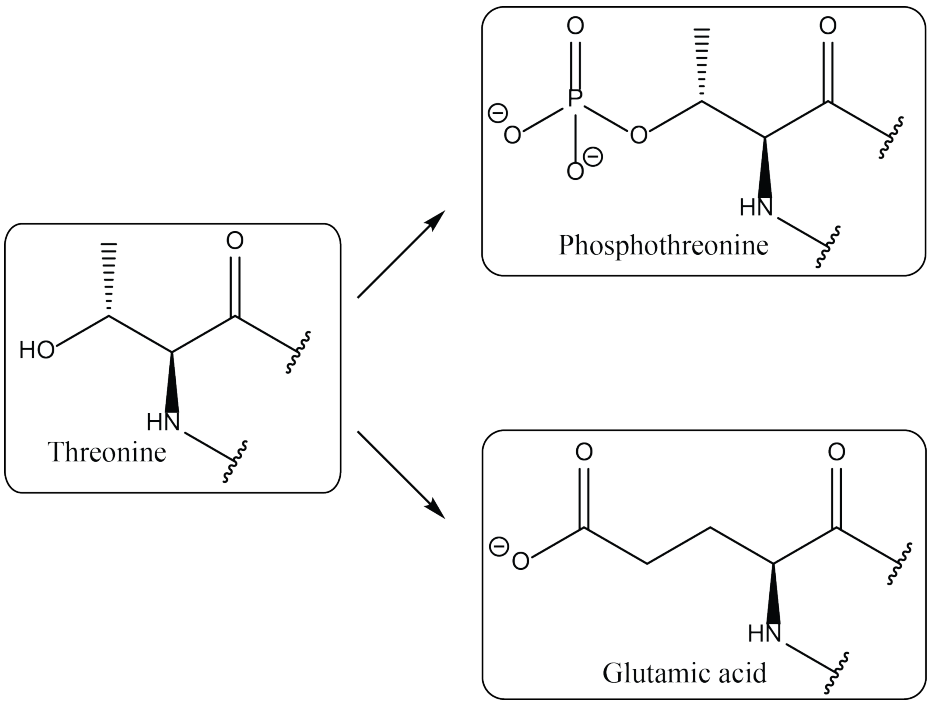


Figure 3.13: Amino acid residue structures. Threonine (left) is phosphorylated *in vivo* to *p*-threonine (top right). In the T504E phosphomimic, threonine is mutated to glutamate (bottom right) to mimic *p*-threonine.

possible that further factors beyond phosphorylation are contributing to the stability of dimeric $Hsp70_{Sf9}$ that are not replicated in the phosphomimic, for example other PTMs such as acetylation.

Mutation of the substrate-binding pocket ruled out the substrate-binding model for dimerisation, however this leaves open the question of how dimerisation occurs if not directly via the substrate-binding domain. Previous studies have proposed various models for dimerisation, including the formation of higher-order oligomers [133]. Higher-order oligomers were not detected under the experimental conditions used in this chapter, even under conditions designed to promote the detection of higher-mass species in the mass spectrometer. Insights in this area are offered by a number of studies reporting Hsp70 dimers which do not involve the substrate-binding pocket. These include structures of antiparallel dimers in both

DnaK [152] and the related yeast protein Hsp110 [175] determined by X-ray crystallography, and of particular relevance here the antiparallel dimer in human $Hsp70_{Sf9}$ deduced from cross-linking data [100]. In this final model, dimerisation occurs via a long interface running the length of the C-terminal helical lid. Taking this final antiparallel model and locating on it the phosphorylation and acetylation sites identified here (Figure 3.14), shows alignment of PTMs along the dimer interface. In this model, *p*-Thr504 is located on the SBD near the hinge of the α-helical lid and the β-domain, comprising the substrate-binding pocket. In the ATP-bound state (Figure 3.14A), *p*-Thr504 is located in a lysine-rich pocket opposite a number of lysine residues on the C-terminal α-helical lid of the second Hsp70, which could contribute to stabilisation of the protein-protein interface. In the ADP-bound state (Figure 3.14B), stabilisation could also be provided by lysine residues located across the proposed dimer interface.

The purpose of Hsp70 dimerisation *in vivo* is still unclear. A single Hsp70 is able to bind to a client [100, 173, 176], indicating that dimerisation is not required for substrate binding. However, studies by mass spectrometry have shown that when looking at the formation of Hsp70 complexes on the HOP-mediated GR maturation pathway, two copies of Hsp70 are incorporated into a stable, hexameric, client complex $(Hsp90)_2(Hsp70)_2(Hop)_1(GR)_1$ [100]. Other studies however have shown the formation of $(Hsp90)_2(Hsp70)_1(Hop)_1(GR)_1$ [177], indicating that both complexes can form. Of particular interest here is the comparison of complex formation using either $Hsp70_{E.coli}$ or $Hsp70_{Sf9}$ which showed an increase in the formation of this hexameric client complex in $Hsp70_{Sf9}$, which quickly disassembled on p23 addition to form the client transfer complex $(Hsp90)_2(GR)_1$ [100], which is necessary for progression of client maturation. When the protein concentration ratios were altered to reflect the stoichiometry of the hexamer, which also reflects

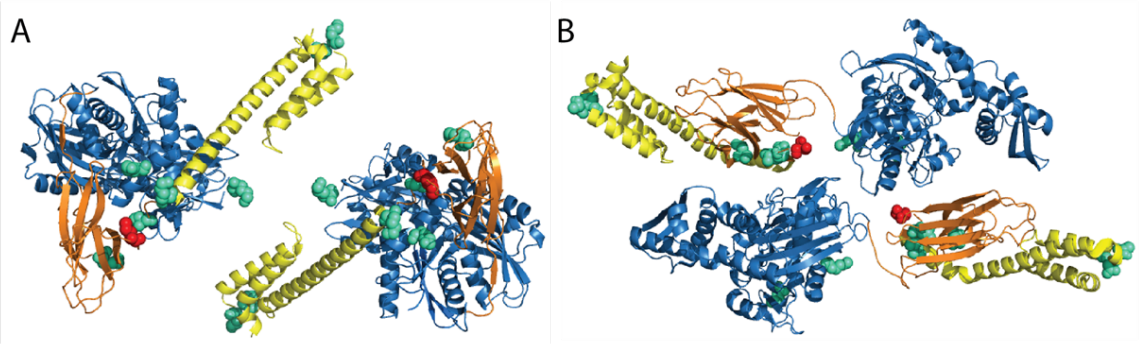


Figure 3.14: Acetylation sites (cyan spheres) and phosphorylation sites (red spheres) identified in this study located on a proposed antiparallel dimer structure [100]. (A) In the ATP-bound state the Thr504 phosphosite and lysine-rich pockets are located adjacent to and across the dimer interface. (B) The ADP-bound state of the proposed Hsp70 antiparallel dimer [100].

the *in vivo* Hsp70:Hsp90 ratio [178], the hexamer dominated with few other species observed [100]. This suggests that enhancing Hsp70 dimerisation may promote GR maturation. Taking these results together with a previous study showing that another Hsp70 phosphorylation site (Tyr525) promotes Hsp70 binding to HOP over CHIP, thereby enhancing protein folding over degradation pathways [161], suggests that a similar regulation of Hsp70 binding to HOP/CHIP could occur here. As such, phosphorylation-mediated dimerisation may target Hsp70 towards the HOP pathway, perhaps over the CHIP/ubiquitin degradation pathway. As activity of the HOP pathway is enhanced in cancer cells, compared to the CHIP pathway, further study of the impact of Hsp70 phosphorylation, including the site identified here in a phosphorylation 'hot-spot', may provide opportunities to target cancer cell viability.

3.7 Conclusions

The results presented in this chapter show an elevated level of dimer present in human $Hsp70_{Sf9}$ when compared to $Hsp70_{E.coli}$. The identification of PTMs

including a phosphorylation site at Thr504 of $Hsp70_{Sf9}$ suggest that phosphorylation may contribute towards this discrepancy in dimerisation. Further investigation revealed that the dimer interface is weakened by increasing ionic strength. Supporting the stabilising role of phosphorylation was the observation that dephosphorylation of $Hsp70_{Sf9}$ markedly decreased dimerisation whereas comparable treatment of the $Hsp70_{E.coli}$ with phosphatase did not. Pinpointing the role of phosphorylation at Thr504, the phosphomimic mutant T504E expressed in *E. coli* displayed enhanced dimerisation relative to the wild type, albeit not to the level seen in $Hsp70_{Sf9}$. Finally, a substrate-binding type interaction with the flexible linker of one Hsp70 monomer acting as the substrate for the second monomer was ruled out by the observation of dimer formation in the substrate-binding-deficient mutant. These results taken together support the view that phosphorylation, including phosphorylation of Thr504, contribute to Hsp70 dimerisation and this dimerisation does not occur through a substrate-binding type interaction. Finally, looking at the enhancement of dimerisation in the broader context of Hsp90/Hsp70 client complex formation, the reported increase in formation of a hexameric client complex [100] suggests a functional role for an Hsp70 dimer in promoting client complex formation.

The role of nucleotide in dimerisation is an area for substantial further study. Recent advances in high-accuracy, high-mass-range mass spectrometers [44] have the potential to accurately differentiate nucleotide and PTM states, making these studies more feasible and this will likely be the next step in studying the dimerisation behaviour of Hsp70. To further investigate the role of nucleotide binding on Hsp70 dimerisation, incubation of Hsp70 with non-hydrolysable ATP analogues such as AMP-PNP and ATP-γS prior to native MS would allow the ATP-bound Hsp70 to be investigated without the complication of ADP turnover.

Alternatively, native MS could be used to compare an ATP hydrolysis-deficient Hsp70, such as the T204A mutant [179], purified in the presence of either ADP or ATP to investigate differing dimerisation behaviour as a result of nucleotide-binding state.

While acetylation sites were identified in this study, further work is required to determine their exact role. Previously, acetylation has not been well-studied with much greater attention placed on phosphorylation. More recently however, the importance of acetylation has been increasingly recognised, in particular in conjunction with phosphorylation [169]. The results reported in this chapter should stimulate further investigation into the role of acetylation in Hsp70 interactions by both highlighting the occurrence of Hsp70 acetylation and showing the importance of Hsp70 PTMs in modulating protein-protein interactions. To further investigate the role of acetylation, site-directed mutagenesis could be used to produce acetylation mimics [180] of the acetylation sites identified in this study (i.e. K108Q, K159Q, etc.) which could then be compared to wild type Hsp70 by native MS, in a similar strategy used for the investigation of the phosphomimic T504E.

The research shown in this chapter leaves open various new avenues for further studies, with the research described in this chapter highlighting the important role PTMs have in Hsp70 interactions. In particular, the role of Hsp70 in cancer cells, where Hsp70 PTMs can regulate the activities of HOP-mediated protein folding and CHIP-mediated protein degradation, mean that understanding Hsp70's interactions and therefore ways to disrupt them, is imperative to finding new Hsp70-targeted treatments. To this end, further investigation into the impact of Hsp70 phosphorylation on HOP/CHIP competition for Hsp70 should be the next step. Native MS could be used to compare the complexes formed when the Hsp70

variants used here (i.e. Hsp70$_{Sf9}$, Hsp70$_{E.coli}$, Hsp70$_{T504E}$) are incubated with HOP and CHIP to investigate these effects, with further understanding of these competing pathways possibly leading to exciting new therapies.

3.8 Methods

3.8.1 Protein Preparation

Proteins were purified as previously described [181]. Plasmids were provided by Dr Elaine Kirschke and Professor David Agard (The University of California, San Francisco).

For *E. coli* expression, pET151 plasmids encoding wild type full-length human Hsp70 and the T504E phosphomimic with N-terminal 6×His-tag and tobacco etch virus (TEV) protease cleavage site (Appendix Figure A.1) were transformed into BL21 (DE3) gold competent cells (Agilent). For proteins expressed with natural abundance isotopes, colonies were inoculated in lysogeny broth (LB), supplemented with ampicillin (50 μg/mL). Cells were grown at 37 ^{o}C with 200 rpm shaking to an optical density (OD) of 0.7–0.8 measured at 600 nm with a UV-vis spectrophotometer (Biomate). Cells were then induced by addition of 1 mM isopropylthiogalactopyranoside (IPTG) for 16 hours at 16 ^{o}C with 200 rpm agitation. Cells were harvested by centrifugation at 5000 rpm for 10 minutes. The cell pellet was resuspended in phosphate-buffered saline (PBS) and centrifuged. The cell pellets were flash-frozen at –80 ^{o}C.

Thawed cells were resuspended in lysis buffer (20 mM Tris (pH 7.5), 20 mM KCl, 1 mM ethylenediaminetetraacetic acid (EDTA), 1 mM dithiothreitol (DTT)). Cells were lysed with a microfluidiser (M-110PS, Microfluidics) and cell debris removed by centrifugation at 25,000 × g for 50 minutes. The supernatant was filtered through a 0.45 μm syringe filter tip and loaded onto a Ni NTA column (GE Healthcare) equilibrated with buffer (30 mM (Tris pH 8.0), 50 mM KCl, 30 mM imidazole). The column was washed with wash buffer (20 mM Tris (pH 8.0), 5 mM $MgCl_2$, 100 mM

NaCl, 2 mM ATP) and the sample eluted with elution buffer (30 mM Tris (pH 8.0), 50 mM KCl, 300 mM imidazole). Fractions containing Hsp70 protein were pooled and supplemented with 1 mM DTT, 0.5 mM EDTA, and 500 μL TEV protease. TEV cleavage was performed at 4 ^{o}C overnight. The sample was concentrated to 10 mL with Amicon Ultra-15 centrifugal filter units (30 kDa molecular weight cut off (MWCO), Millipore) and diluted to a final volume of 100 mL with buffer (20 mM Tris (pH 8.0), 5 mM $MgCl_2$, 100 mM NaCl). The sample was loaded onto a Ni NTA column, equilibrated with buffer (30 mM Tris (pH 8.0), 50 mM KCl, and 30 mM imidazole), and washed with buffer (20 mM Tris (pH 8.0), 5 mM $MgCl_2$, 100 mM NaCl). Uncleaved His-tagged protein was eluted with elution buffer (30 mM Tris (pH 8.0), 50 mM KCl, 300 mM imidazole). Hsp70-containing fractions from the flow-through were pooled, concentrated, and washed in storage buffer (20 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) (pH 7.6), 5 mM $MgCl_2$, 100 mM KCl) with an Amicon Ultra-15 centrifugal filter unit (30 kDa MWCO) until the OD_{260} of the flow-through was equal to 0. Aliquots of Hsp70 protein were flash-frozen and stored at -80^{o}C.

For isotopically-labelled $Hsp70_{E.\ coli}$ variants, protein was expressed in *E. coli* using M9 media, supplemented with either ^{15}N-labelled ammonium chloride or ^{13}C-labelled glucose and then purified as described above. In order to produce the T504E/V438F mutant, the T504E plasmid was modified by site-directed mutagenesis using the QuikChange Lightning kit (Qiagen), according to the manufacturer's directions. The plasmid was extracted by MiniPrep (Qiagen), according to the manufacturer's instructions, was sequenced to confirm its identity (Source BioScience) and protein purified as described above.

Purified $Hsp70_{Sf9}$ was provided by Dr Elaine Kirschke, Zygy Roe-Zurz, and

Professor David Agard (The University of California, San Francisco), purified as follows. Human $Hsp70_{Sf9}$ was purified from pellets of baculovirus-infected Sf9 cells expressing human Hsp70 produced at the Baculovirus/Monoclonal Antibody Facility at Baylor College of Medicine, USA. Cells were lysed in lysis buffer (20 mM Tris (pH 7.5), 20 mM KCl, 1 mM EDTA, 1 mM DTT). Protein was purified by ion exchange using a HiPrep 16/60 Q FF column (GE Healthcare) and eluted with a linear gradient between 20 and 500 mM KCl. Protein was further purified by an ATP agarose-affinity column (Sigma), eluted with 20 mM ATP/$MgCl_2$. Protein was purified by size-exclusion chromatography (SEC) on a Superdex S200 16/600 (GE Healthcare) and the buffer was exchanged to storage buffer (30 mM HEPES (pH 7.5), 50 mM KCl, 1 mM DTT, 5 mM $MgCl_2$, 10 % glycerol) using a HiTrap desalting column (Sigma).

3.8.2 Mass Spectrometry

For MS analysis, the buffer of protein complexes was exchanged to 100 mM ammonium acetate (pH 7.5) using micro Bio-spin columns (Bio-Rad) or Amicon (10 kDa MWCO columns). 2.5 μL protein complex was loaded into a gold-coated borosilicate capillary manufactured in-house [165]. Mass spectra were acquired on a QToF II Mass Spectrometer (Micromass) modified for transmission of high mass complexes [58]. Instrument parameters were optimised for each spectrum, although parameters were kept constant for spectral comparisons. Typical parameters were: capillary voltage 1.7 kV; cone voltage 100–200 V; extractor 15 V; collision voltage 20–60 V; backing pressure 5–6 mbar. Spectra were processed with MassLynx v4.1 and peak fitting performed with UniDec software [171]. Spectra were calibrated externally with cesium iodide (10 mg mL^{-1}). Details of solution conditions for experiments are given below.

3.8.2.1 Hsp70$_{E.\ coli}$/Hsp70$_{Sf9}$ Comparison and Phosphatase Treatment

3 μM Hsp70$_{Sf9}$ and ^{13}C isotopically-labelled-Hsp70$_{E.\ coli}$ were analysed by MS separately and in a 1:1 mixture. They were then incubated in 100 mM ammonium acetate with alkaline calf phosphatase agarose beads (Sigma) for 16 hrs at 4 ^{o}C and a control sample incubated without phosphatase. Phosphatase beads were removed by filtration, before MS analysis. In a second experiment, 6 μM Hsp70$_{E.\ coli}$ was incubated with alkaline calf phosphatase agarose beads at 4 ^{o}C or without phosphatase as a control, in 100 mM ammonium acetate. Sample were analysed at t=0, t=100 minutes and t=48 hours.

3.8.2.2 Ionic Strength Titration

2 μM Hsp70$_{Sf9}$ was incubated for 1 hour at room temperature in binding buffer (30 mM HEPES (pH 7.5), 1 mM DTT, 200 μM ATP-Mg) supplemented with KCl at increasing concentrations: 50 mM, 100 mM and 300 mM KCl. For MS analysis, instrument parameters were as above but with a reduced collision voltage of 20 V and a higher backing pressure of 6 mbar.

3.8.2.3 Native Mass Spectrometry of Hsp70 Mutants

For comparison of phosphomimic, wild type, and substrate-binding-deficient variants, 1:1 mixtures of either 2 μM Hsp70$_{T504E}$ with natural abundance isotopes and ^{15}N Hsp70$_{E.\ coli}$ or 3.5 μM Hsp70$_{T504EV438F}$ with natural abundance isotopes and ^{15}N Hsp70$_{E.\ coli}$ were incubated after buffer exchange into 100 mM ammonium acetate supplemented with 200 μM ATP-Mg.

3.8.2.4 LC-MS/MS Analysis of Post-Translational Modifications

For identification of PTMs, proteins were separated by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and gel bands excised. Proteins

were digested with trypsin in-gel as described previously [182].

For identification of acetylation sites, peptides were separated by nano-flow RPLC (EASY nLC 1000, Thermo Fisher Scientific) with mobile phase A, 0.1 % (v/v) FA/ 5 % (v/v) DMSO and mobile phase B, 100 % (v/v) ACN/ 0.1 % (v/v) FA/ 5 % (v/v) DMSO) coupled to a Q Exactive Orbitrap Mass Spectrometer (Thermo Fisher Scientific). Peptides were loaded onto a trap column (5 mm, PepMap RSLC, C18, 300 μm I.D. particle size 3 μm; Thermo Fisher Scientific) and separated on an analytical C18 capillary column (50 cm, PepMap RSLC, EASY-spray column, C18, 75 μm I.D. particle size 3 μm; Thermo Fisher Scientific) with a flow rate of 200 nL min^{-1} and a gradient of 7-30 % (v/v) mobile phase B over 30 minutes. Eluted peptides were analysed by MS with instrument parameters: spray voltage 2.1 kV; capillary temperature 320 ^{o}C. The Q Exactive Orbitrap was operated in data-dependent mode. Survey full scan MS spectra were acquired in the Orbitrap (m/z 350-1500) with a resolution of 70,000 at an AGC target at 3×10^6. The ten most intense ions were selected for HCD at 30 V at an AGC target of 50,000 and previously-selected ions dynamically excluded for 5 s.

Phosphopeptide analysis was performed by Professor Carla Schmidt (The Martin Luther University of Halle-Wittenberg), as follows. Phosphopeptides were enriched by TiO_2-affinity chromatography as described previously [92]. Briefly, TiO_2 columns were prepared by packing TiO_2 material (GL Sciences) into pipette tips. The columns were washed with 5 % trifluoroacetic acid (TFA)/ 80 % acetonitrile (ACN) and then with 20 % 2,5-dihydroxybenzoic acid (DHB)/5 % TFA/80 % ACN. Peptides were dissolved in 20 % DHB/ 5 % TFA/ 80 % ACN and loaded onto the column. The columns were washed with 20 % DHB/5 % TFA/80 % ACN followed by 5 % TFA/ 80 % ACN. The phosphopeptides were eluted with 0.3 N ammonia

solution (pH > 10.5). Eluted peptides were dried in a vacuum concentrator. Peptides were separated by nanoflow RPLC (DionexUltiMate 3000 RSCL nano System, Thermo Fisher Scientific) with mobile phase A, 0.1 % (v/v) formic acid (FA) and mobile phase B, 80 % (v/v) ACN/ 0.1 % (v/v) FA. Peptides were loaded onto a trap column (HPLC column Acclaim PepMap 100, C18 100 μm I.D. particle size 5 μm; Thermo Fisher Scientific) and separated on a 50 cm analytical column (Acclaim PepMap100, C18, 75 μm I.D. particle size 3 μm; Thermo Fisher Scientific) with a gradient of 5-80 % (v/v) mobile phase B over 74 minutes. Peptides were eluted into a coupled LTQ Orbitrap XL Mass Spectrometer (Thermo Fisher Scientific). MS instrument parameters were as follows: spray voltage 1.8 kV; capillary temperature 180 ^{o}C; normalised collision energy 35 % at an activation q = 0.25 and an activation time of 30 ms. The LTQ Orbitrap XL was operated in data-dependent mode. Survey full scan MS spectra were acquired in the Orbitrap (m/z 300–2000) with a resolution of 30,000 at 400 m/z and an automatic gain control (AGC) target at 1×10^6. The five most intense ions were selected for CID in the linear ion trap at an AGC target of 30,000, with dynamic exclusion of previously-detected ions. For phosphosite identification, multistage activation was used for neutral loss of the mass of one to four phosphosites per peptide for charge states 2+ or 3+. Previously-detected ions were dynamically excluded from detection in the linear ion trap for 30 s. The Orbitrap was calibrated internally with the lock mass option (m/z=445.120025 [183]).

3.8.2.5 Database Search

For identification of phosphorylation sites, raw data were searched against the NCBI non-redundant database with the *Homo sapiens* taxonomy filter using Mascot v2.4.1 search engine (Matrix Science). Parameters were set as follows: precursor mass accuracy 15 ppm; MS/MS fragment mass accuracy 0.5 Da; digestion

enzyme trypsin; up to two missed cleavage sites; variable modifications of carbamidomethylation of cysteine, oxidation of methionine, phosphorylation of serine, threonine, and tyrosine. Spectra were validated by manual inspection. For identification of acetylation sites, raw data was searched against SwissProt database with the *H. sapiens* taxonomy filter (20,353 sequences) using the Mascot v2.4.1 search engine. Parameters were set as follows: precursor mass accuracy 7 ppm; MS/MS fragment mass accuracy 0.005 Da; digestion enzyme trypsin; up to two missed cleavage sites; variable modifications of carbamidomethylation of cysteine, oxidation of methionine, acetylation of lysine and protein N-termini. Spectra were validated by visual inspection to ensure spectral quality.

3.8.3 Homology Modelling

A homology model of ADP-bound human Hsp70 was generated with the Swiss-Model Workspace (http://swissmodel.expasy.org/), using the template PDB ID 4B9Q [184]. For the ATP-bound conformation, the amino acid sequence of human Hsp70 was aligned to PDB ID: 2KHO [124] with Pymol software.

Chapter 4

Regulation of the Interactions of Eukaryotic Translation Initiation Factor 2B

4.1 Introduction

Molecular chaperones, such as Hsp70, assist in protein folding and can be upregulated in response to stress, however cells also possess alternative means to prevent the accumulation of unfolded proteins. When the chaperone machinery is unable to cope with the workload imposed upon it, cells are able to modify the supply of clients, that is newly-synthesised proteins, in order to maintain proteostasis and homeostasis in general. Newly-synthesised proteins are transported to the endoplasmic reticulum

Parts of this chapter have been published in: Victoria Beilsten-Edmands, Yuliya Gordiyenko, Jocky CK Kung, Shabaz Mohammed, Carla Schmidt, and Carol V Robinson, *Cell discovery*, 2015 [185] and Carla Schmidt, Victoria Beilsten-Edmands, and Carol V Robinson, *Journal of Molecular Biology*, 2016 [186]. Dr Yuliya Gordiyenko and Jocky Kung (The University of Oxford, UK) purified eIF2 from *S. cerevisiae* and performed preliminary phosphosite experiments which were analysed in conjunction with my phosphosite experiments. Purified eIF2B from *S. cerevisiae* was provided by Professor Graham Pavitt (The University of Manchester, UK). Purified human eIF2B and eIF2Bα, expressed in *E. coli*, were provided by Jordan Tsai and Professor Peter Walter (The University of California, San Francisco).

(ER) for folding and processing and it is here that another important control point in proteostasis is located. As unfolded proteins accumulate in the ER, pathways are triggered including the integrated stress response (ISR). The ISR is focussed on a critical step in the initiation of protein translation and thus modifies the supply of newly-synthesised proteins to the ER.

4.2 Protein Translation

4.2.1 Translation Initiation

The central dogma of molecular biology asserts that genetic information in the form of DNA is transcribed to RNA which is then translated to a protein sequence [187, 188]. The process of biological protein synthesis proceeds therefore via translation and transcription of genetic information on the ribosome, with translation initiation being the first step on the pathway to protein synthesis.

The process of eukaryotic translation can be broken down into three main stages; initiation, elongation, and termination. The first of these, translation initiation, is the process by which the small (40S) and large (60S) ribosomal subunits, along with initiator methionyl-tRNA (Met-tRNA$_i^{Met}$), assemble into the functional 80S ribosome complex required for the elongation phase [189]. Initiation is a tightly-regulated process, the most highly-regulated stage of protein translation, and necessitates the coordination and interaction of numerous translation initiation factors [190]. The processes of translation initiation in eukaryotes and prokaryotes share many similarities but also have notable differences. Firstly, they have distinct initiation factors while prokaryotic initiation also requires the formylation of Met-tRNA$_i{}^{Met}$ to fMet-tRNA$_i{}^{fMet}$. They also display variation in mRNA binding site on the ribosome, with the Shine-Dalgarno sequence serving as the binding site

for mRNA in bacteria, while in eukaryotes it is the AUG start codon (located downstream of the 5' untranslated region (5' UTR)) that is the site for mRNA binding [191]. Eukaryotic translation initiation and its associated eukaryotic translation initiation factors (eIFs) will be the focus of this chapter.

Translation initiation, a mechanistic overview of which is shown in Figure 4.1, begins with the binding of GTP-eIF2 and Met-tRNA$_i{}^{Met}$ to form the ternary complex (TC) [190]. The binding of eIFs 1, 1A, 3, and 5 together form the multifactor complex (MFC) which binds to the 40S ribosome subunit to assemble the 43S pre-initiation complex (PIC). A pre-activated mRNA attaches to the 43S pre-initiation complex and begins scanning for the AUG start codon, which can be located over a thousand nucleotides away [192]. Scanning and AUG start codon recognition instigate eIF1 dissociation and GTP hydrolysis, with scanning halted, followed by dissociation of GDP-eIF2 bound to eIF5 [193]. The 60S subunit joins the 40S subunit, facilitated by the GTPase eIF5B before the ribosomal initiation complex (IC) proceeds to the elongation phase. GDP-eIF2, which exited the initiation process after AUG start codon recognition while bound to eIF5, must undergo nucleotide exchange to GTP-eIF2 before the next cycle of translation initiation can commence. This nucleotide exchange is facilitated by a guanine nucleotide exchange factor (GEF) eIF2B, which displaces eIF5 [194, 195] and is present at catalytic ratios to eIF2 and eIF5 [193, 196]. The interaction of eIF2 and eIF2B is a key control point for protein translation and will be the focus of this chapter.

4.2.2 The Eukaryotic Translation Initiation Factors

As discussed above, there are many eIFs required for successful translation initiation in eukaryotes, with at least 12 eIFs required [197], but the interaction of

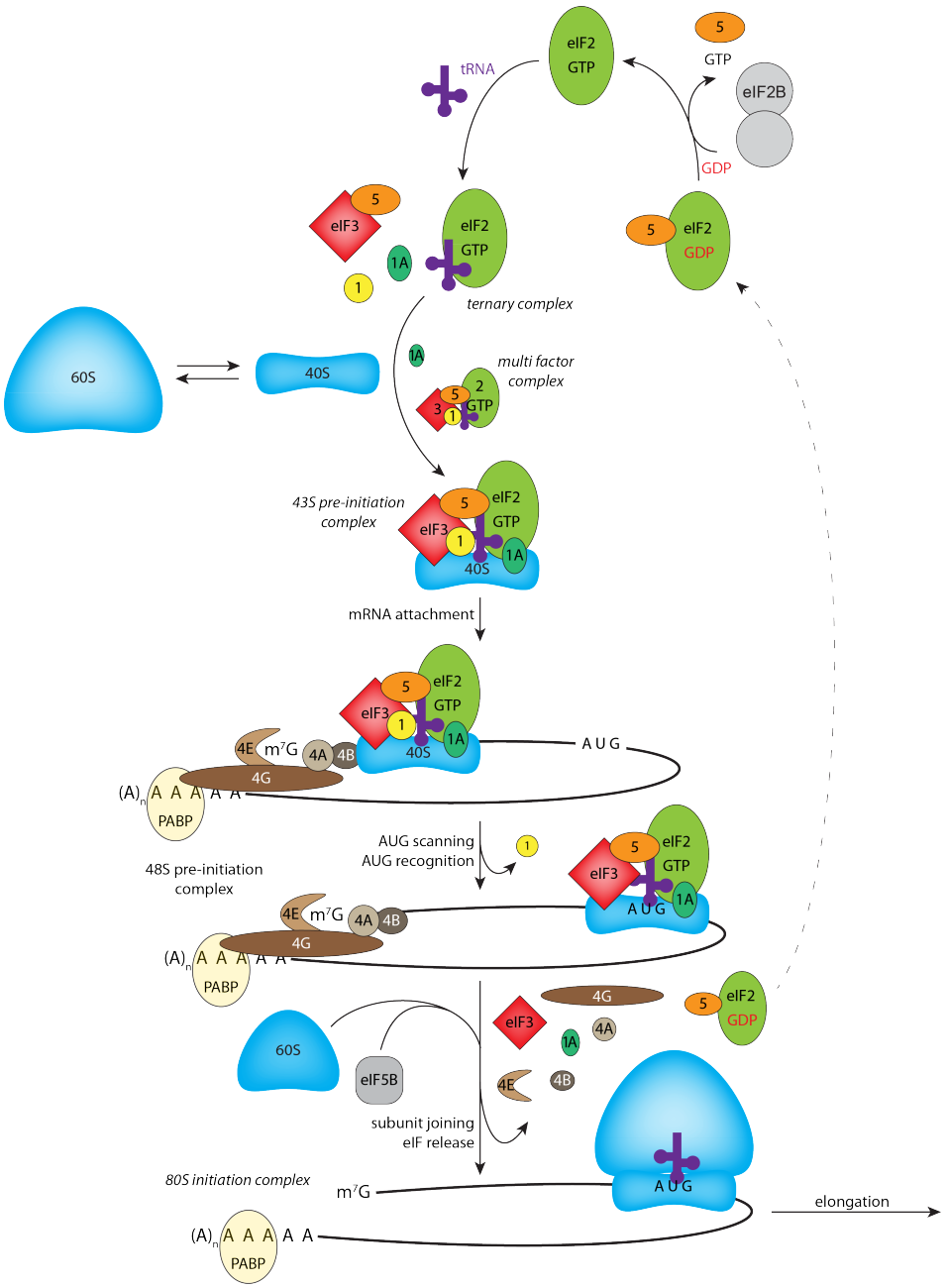


Figure 4.1: Overview of eukaryotic translation initiation. GTP-bound eIF2 binds to tRNA to form the TC. The binding of eIFs 1, 1A, 3, and 5 together form the MFC which binds to the small ribosomal subunit to form the 43S pre-initiation complex. A pre-activated mRNA attaches to the 43S pre-initiation complex and AUG scanning commences. After AUG recognition, GTP on eIF2 is hydrolysed to GDP with dissociation of the eIFs. Ribosomal subunit joining occurs, facilitated by eIF5B before the ribsomal complex proceeds to the elongation phase. GTP-eIF2 is regenerated by eIF2B. The interaction between eIF2 and eIF2B is an important site of translation initiation regulation. Phosphorylation of Ser51 on eIF2α induces tight binding of eIF2 and eIF2B, preventing eIF2B, from further nucleotide exchange. Figure adapted from Schmidt *et al.* [186].

eIF2 and eIF2B is particularly interesting from a regulatory point of view as it serves as a key control point in protein synthesis.

eIF2 is a trimeric G-protein comprising subunits α, β, and γ which is required for Met-tRNA$_i^{Met}$ loading onto the 40S subunit. Formation of the TC is driven by eIF2γ which binds both GTP and Met-tRNA$_i^{Met}$ [198], with Met-tRNA$_i^{Met}$ binding stabilised by subunits eIF2α and β [199]. eIF2α is the main regulatory subunit of eIF2, regulating Met-tRNA$_i^{Met}$ binding allosterically [199] and also containing a key regulatory phosphorylation site [200] which will be discussed in detail later. eIF2β stabilises Met-tRNA$_i^{Met}$ binding and also interacts with multiple eIFs within the PIC, including eIFs 1, 1A, and 5 [201–203]. The structure of the related archaeal initiation factor (aIF2) G-protein from *Solfolobus solfataricus* has been solved by X-ray crystallography at a resolution of 2.8 Å (Figure 4.2) [204]. This structure reveals a high degree of conformational flexibility on subunits eIF2β and eIF2α which both contact the central eIF2γ subunit; flexibility which is constrained in Met-tRNA$_i^{Met}$-bound eIF2, as shown by comparative cross-linking analysis [185]. Ultimately, GDP-eIF2 must undergo eIF2B-mediated nucleotide exchange to regenerate GTP-eIF2 which is required for TC formation and therefore eIF2B is an indispensable part of the pathway.

eIF2B is a GEF comprised of five subunits: $\alpha, \beta, \gamma, \delta$, and ϵ. The catalytic subunit is ϵ, whose C-terminal domain displays GEF activity and interacts with lysine-rich segments of eIF2β [205] and the G domain of the nucleotide-bound eIF2γ [206, 207]. Subunits eIF2Bα, β, and δ, which are homologous to one another [206], form the regulatory subcomplex and interact with eIF2α to stabilise eIF2 binding [208]. Native MS has shown that rather than consisting of a pentamer of the five subunits, eIF2B is in fact decameric with the stoichiometry $\alpha_2\beta_2\gamma_2\delta_2\epsilon_2$ in human [209] and

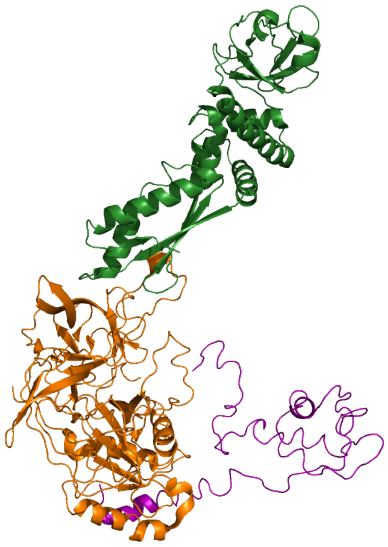


Figure 4.2: The structure of aIF2 from *S. solfataricus* determined by X-ray crystallography (PDB ID: 3CW2 [204]). The structure is trimeric and is comprised of subunits eIF2Bα (green), β (purple), and γ (orange).

yeast [210]. Subunit eIF2Bα has been shown to be non-essential in eIF2 binding and nucleotide exchange [211], and is frequently lost during purification [212]. Subunit eIF2Bα was also shown to dissociate from the remaining assembly by native MS, with eIF2Bα dimers observed both by MS [209] and also by X-ray crystallography [213]. The structure of eIF2B has remained elusive, albeit with multiple studies shedding light on subunit arrangement. Multiple models of eIF2B have been proposed and yet the reported data are somewhat contradictory to these, with the formation of eIF2Bα dimers [209, 213] and belief that eIF2Bα is essential to decamer formation leading some to propose an eIF2Bα_2 core [214] while others have pointed to the formation of eIF2B$\gamma_2\epsilon_2$ complexes and intact eIF2B$\beta_2\gamma_2\delta_2\epsilon_2$ complexes in the absence of eIF2Bα to construct models with an eIF2B$\gamma_2\epsilon_2$ core and non-contacting eIF2Bα subunits [210]. Despite the uncertainty surrounding the structure of eIF2B, it is clear that it serves an important role in translation

initiation and consequently in protein synthesis.

4.2.3 The Integrated Stress Response

Translation initiation has a critical role in maintaining cellular homeostasis and therefore it is not surprising that this process is targeted for regulation in response to stress. Eukaryotes have adapted a shared response called ISR [215–219], which is activated when the cell is subjected to various stresses and cellular changes such as viral attack, oxygen deprivation, and glucose starvation [215, 218–222]. The ISR centres on the interaction of GDP-bound eIF2 with its GEF, eIF2B. In the ISR in metazoans, one of four protein kinases (GCN2, PKR, HRI, PERK) phosphorylate Ser51 on eIF2α [200]. This phosphorylation induces tight binding of eIF2 with eIF2B, preventing regeneration of GTP-eIF2 for a new round of initiation and sequestering eIF2B, present in catalytic amounts [193, 196], from further nucleotide exchange interactions. Due to the high ratio of eIF2:eIF2B, only a small percentage of eIF2 needs phosphorylating to have a significant effect on the pathway. This eIF2-eIF2B interaction has the effect of downregulating 5'Cap-dependent protein synthesis and also induces specific genes, including activating transcription factor 4 (ATF4), which together support cell recovery and restore homeostasis [223, 224]. In the face of significant stress, be that either the intensity or duration of stress, the ISR proceeds to the triggering of cell death. Ultimately, phosphorylation of eIF2α acts as a reversible switch of the ISR and dephosphorylation restores normal levels of protein synthesis [225].

The four protein kinases are activated through dimerisation and auto-phosphorylation. They are all responsible for eIF2α phosphorylation, but despite sharing a common target, they differ in the conditions under which they are activated. Amino acid starvation triggers activation of general control

derepressible 2 kinase (GCN2), while it is heme deprivation and viral infection which trigger activation of heme-regulated eIF2α kinase (HRI) and double-stranded protein kinase (PKR), respectively. ER stress, for example through accumulation of unfolded proteins which also upregulates expression of Hsp70 as discussed in Chapter 3, triggers PKR-like ER kinase (PERK) activation, found in metazoans. Ultimately, the ISR is essential to survival in eukaryotes with the mutation S51A on eIF2α permanently turning off the ISR and proving lethal shortly after birth in mouse models [223].

4.2.3.1 The Integrated Stress Response Inhibitor

While the ISR can be critical to survival, it can also have adverse effects. Memory consolidation is impaired by the ISR and this is particularly apparent in the case of traumatic brain injury (TBI) which is a significant public health issue [226, 227]. In TBI, the ISR becomes permanently switched on [228], with cognition and long term memory significantly affected. Furthermore, TBI has been linked to an increased risk of developing dementia later in life [229, 230] and as such, ways to target the ISR and improve memory, without the adverse effects of inhibiting the important function the ISR has in cellular homeostasis, are a much-sought-after therapeutic intervention.

A significant advance in this area was made in 2013 with the remarkable discovery of the integrated stress response inhibitor (ISRIB, Figure 4.3) using cell-based screening of a library of small molecules in search of a PERK inhibitor [231]. ISRIB reversed the effect of eIF2α phosphorylation (IC_{50} = 5 nM), downstream of the kinase activity without reversing eIF2α phosphorylation itself, the first molecule discovered to do so [231, 232]. Furthermore, when mice afflicted with TBI were treated with ISRIB, they showed enhanced learning abilities in water maze

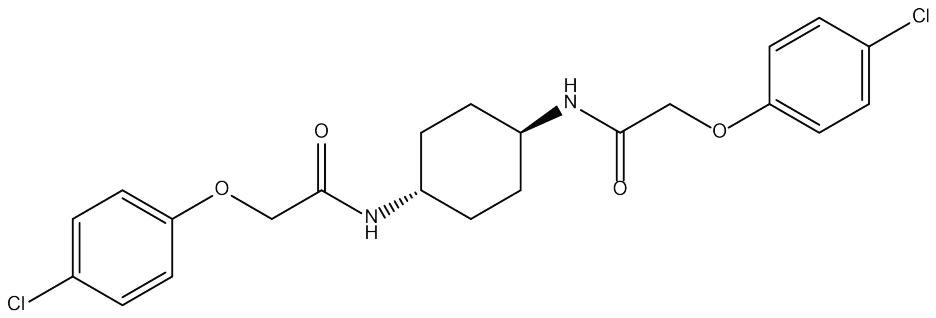


Figure 4.3: The structure of the integrated stress response inhibitor (ISRIB).

experiments [231] indicating that ISRIB is able to reverse the negative effects of the ISR on memory.

The exact mechanism by which ISRIB inhibits the ISR is unclear, but a recent study has shown that ISRIB acts on eIF2B, enhancing its GEF activity [233]. Furthermore, studies have also indicated that eIF2Bδ is the target of ISRIB with a thermal shift assay showing ISRIB-induced stabilisation of eIF2Bδ [233] and mutations to eIF2Bδ reversing the effects of ISRIB [234]. Velocity sedimentation analysis has indicated that ISRIB may stabilise eIF2B decamers [233], thus increasing the proportion of active eIF2B available for guanine nucleotide exchange. How ISRIB enhances the GEF activity of eIF2B and reverses the effect of eIF2α phosphorylation is not yet fully understood and elucidating the mechanism of ISR inhibition by ISRIB is consequently an area of significant interest.

4.3 Research Aims

PTMs play an important role in the regulation of translation initiation, particularly the interaction of eIF2 and eIF2B, notably on Ser51 phosphorylation of eIF2α. The first research aim is therefore to investigate the presence of PTMs on endogenously-expressed eIF2 and eIF2B from *Saccharomyces cerevisiae*, a model

eukaryotic organism, in order to determine their PTM status and consider how this may impact on the regulation of translation initiation. Secondly, the structure of eIF2B, which appears to form a decamer, has been unclear for many years and reconciling the data of many studies with a single structure has been a challenge. Information on the structure of eIF2B will therefore be probed by employing chemical cross-linking of eIF2B from *S. cerevisiae*. The recently-identified inhibitor ISRIB which targets eIF2B has produced significant results with respect to its inhibition of the ISR, yet its mechanism of action is unknown. The final aim of this chapter is therefore to investigate the interaction of ISRIB with human eIF2B in order to address some of the unknown elements of its behaviour. I will first analyse human eIF2B, in the absence of ISRIB, to establish its composition before addition of ISRIB. Then, I will investigate any differences in formation of the decamer, particularly any differences in the formation of eIF2B decameric $\alpha_2\beta_2\gamma_2\delta_2\epsilon_2$ and pentameric $\alpha\beta\gamma\delta\epsilon$ complexes when eIF2B is treated with ISRIB. Finally, I will assess if ISRIB induces any changes in eIF2Bα association to the rest of the eIF2B complex and consider if this has any influence on the ISR.

4.4 Results

4.4.1 Yeast eIF2B

4.4.1.1 Post-Translational Modifications

Phosphorylation is known to be important in the regulation of eIF2 on account of its role in inhibiting guanine nucleotide exchange by eIF2B. Given this, there is potential for further regulation by phosphorylation and as such investigating the phosphoproteome of the eIFs is an important step in exploring eIF regulation. To survey the PTM status of eIFs, samples of purified eIF2 and eIF2B from

S. cerevisiae were digested with trypsin and analysed by LC-MS/MS. Purified eIF2 was provided by Dr Yuliya Gordiyenko and Jocky Kung (The University of Oxford, UK) and purified eIF2B was provided by Professor Graham Pavitt (The University of Manchester, UK). As the PTM status of proteins can change during the cell cycle and in response to various stresses, endogenously-expressed eIF2 and eIF2B, harvested during the log phase of the cell cycle were analysed. As phosphorylation is often substoichiometric and therefore difficult to identify, phosphopeptides were enriched in samples by TiO_2 affinity chromatography, before LC-MS/MS analysis. Raw data from preliminary experiments, performed by Dr Yuliya Gordiyenko and Jocky Kung (The University of Oxford, UK), was also included in the data analysis. After database searching and manual verification, thirteen and seven phosphorylation sites were identified for eIF2 and eIF2B, respectively (Appendix Table B.1). These sites were predominantly located on serine residues, with a smaller number located on threonine and tyrosine residues, as is typically observed [235].

A crosstalk between phosphorylation and acetylation has been described in eukaryotes and bacteria [169], therefore the existence of acetylation sites was also studied. Nine acetylation sites were identified on eIF2 and eIF2B. Of these nine sites, four were located at the N-terminus with the remaining five sites located on lysine residues (Appendix Table B.1, Figure 4.4). Three PTM databases were searched and compared to these results. Of the 20 phosphosites identified here, six were previously unreported and of the nine acetylation sites identified here, eight were previously unreported.

To visualise the location of the identified sites, available structures in the RSCB Protein Data Bank (PDB, https://www.rcsb.org/ [236]) were utilised. However,

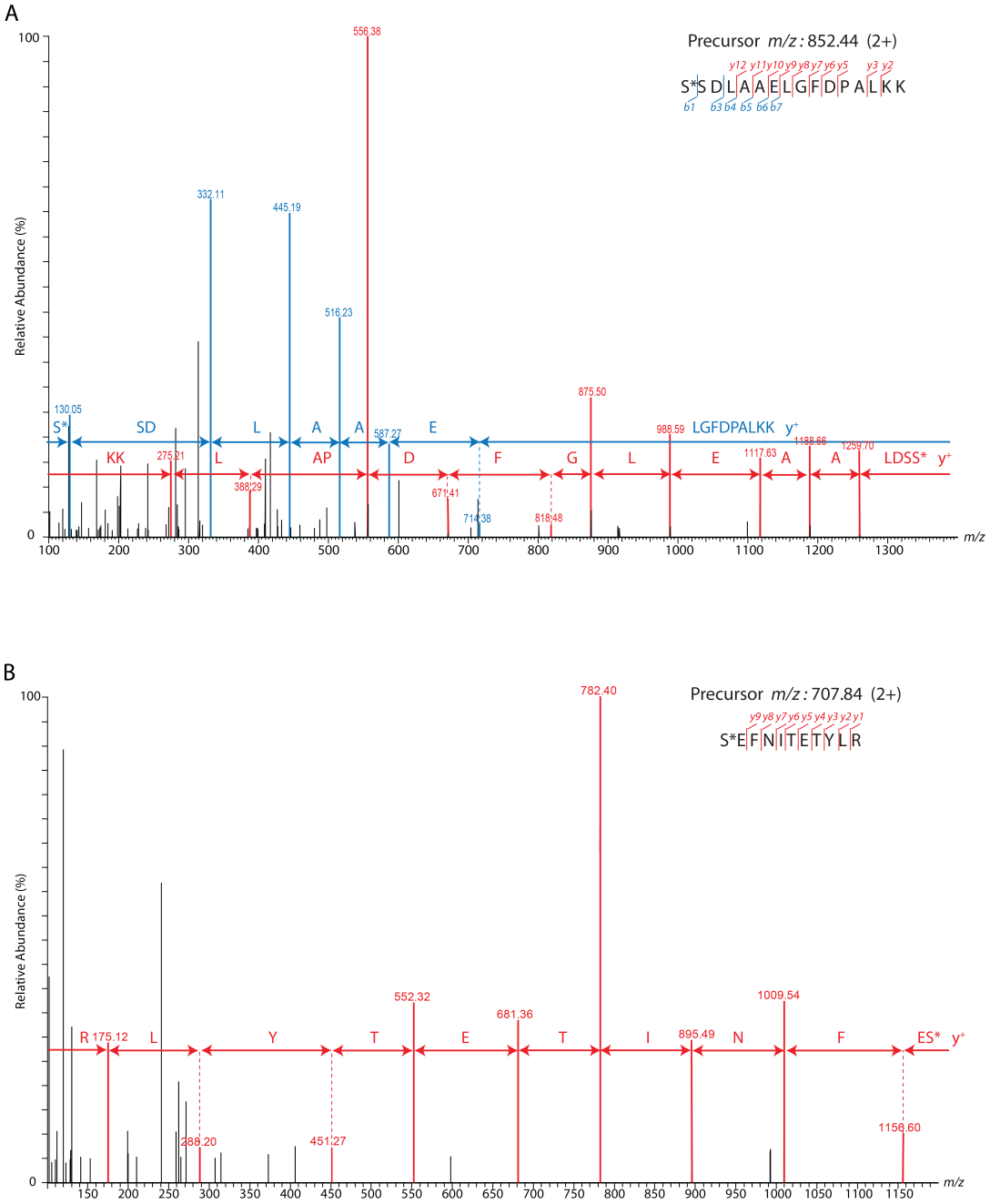


Figure 4.4: MS/MS spectra identifying PTMs, with y and b ions marked (red and blue, respectively). (A) MS/MS spectrum identifying N-terminal acetylation of Ser2 on eIF2β. (B) MS/MS spectrum identifying N-terminal acetylation of Ser2 on eIF2Bα.

many of the sites were located on residues that were not included in the available structures, suggesting that they are located in the highly-flexible regions of the proteins. For visualisation, the proteins are therefore represented as bars onto which the location of the sites identified here and in the databases are shown (Figure 4.5A). It is clear that all of the protein subunits contain modifications and some are highly modified. Phosphorylation sites outnumber acetylation sites, particularly in the databases, possibly as a result of the greater interest in phosphosites. Furthermore, many of the PTMs appear in clusters on the protein chain. Highlighting the location of available structural information on the bars reveals that these clusters are present around the unstructured or uncharacterised parts of the protein chains. For eIF2, most of the phosphorylation occurs on the C-terminus of eIF2α and the N-terminal half of eIF2β, whereas the sites on eIF2γ do not cluster and are situated along the length of the protein chain. In the case of acetylation, both α and β subunits of eIF2 are acetylated at the N-terminus, with the remaining acetylation sites located on lysines on eIF2β and γ. There is no clustering of acetylation sites apparent unlike with the phosphorylation sites, however the acetylation sites are often located near to phosphorylation sites.

Next, the phosphorylation and acetylation sites were located on a homology model of eIF2 generated with the Phyre2 web server [237] (Figure 4.5B). As many of the PTM sites were located in parts of the protein without structural information, these regions of the protein structure were determined by *ab initio* modelling and as such are unreliable. Most of these PTMs were found in the Met-tRNA binding cleft, suggesting a role in tRNA or nucleotide binding.

The PTMs on eIF2B identified in this study and in the literature were visualised in bar diagrams (Figure 4.5C), however a homology modelling approach could not be

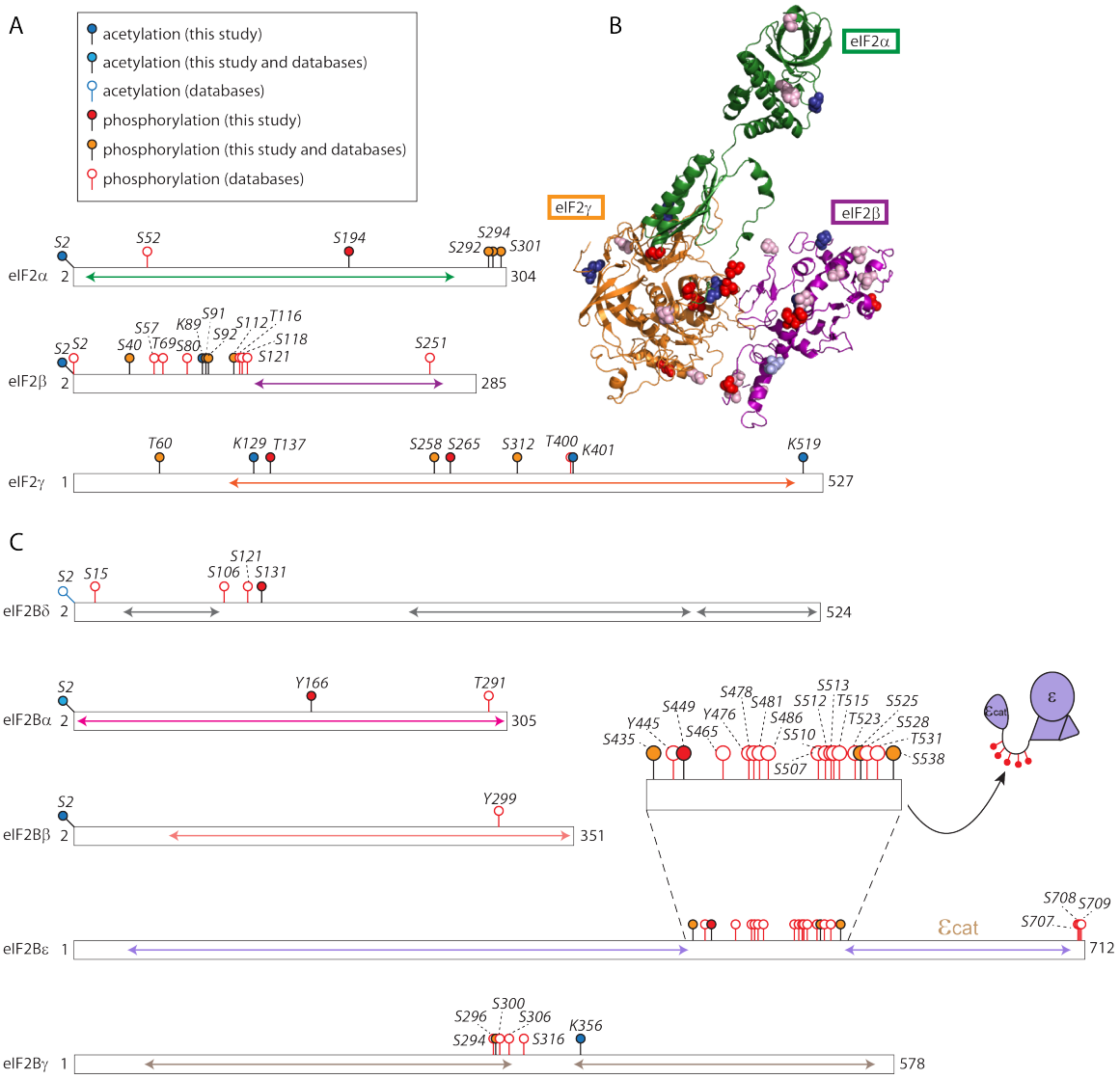


Figure 4.5: Post-translational modifications of eIF2 and eIF2B. The location of PTMs identified in this study and by literature search are indicated. See key for details. (A) eIF2 subunits represented as bar diagrams. The green, purple, and orange arrows indicate regions of the protein for which high-resolution structures are available for eIF2α, β, and γ, respectively. B) The locations of these PTMs are displayed on the homology model of eIF2. Phosphorylation and acetylation sites found in this study (red and blue, respectively) and phosphorylation and acetylation sites reported in the literature (light pink and light blue, respectively) are shown as spheres. C) The location of PTMs on eIF2B identified in this study and by literature search are indicated on the protein sequence. The arrows within the bars indicate regions of the protein for which high-resolution structures are available. The bar region corresponding to the linker adjoining the catalytic domain of eIF2Bϵ is enlarged for greater detail and displayed on a representation of the linker. Adapted from Beilsten-Edmands *et al.* [185].

used here on account of the uncertainty in the subunit organisation of eIF2B. The most heavily-modified subunit of eIF2B is subunit eIF2Bϵ, with a high degree of phosphorylation. It is striking that all of the phosphosites both identified here and listed in the databases occur in protein regions with little structural information. In particular most of these are found in the flexible, unstructured region which links the catalytic and non-catalytic domains of eIF2Bϵ.

Significant changes in the location of the catalytic domain of subunit ϵ are required as part of the functional role of eIF2B in mediating guanine nucleotide exchange on eIF2 [210]. The high degree of modification present on the linker to the catalytic domain could serve to regulate the movement of the catalytic domain, from the *apo* eIF2B conformation to the eIF2-bound conformation, allowing eIF2B-mediated guanine nucleotide exchange.

In these experiments the eIFs were endogenously-expressed and the cells harvested during the log phase of cell growth. Many PTMs are known to occur as a consequence of exposure to various stresses, including but not limited to hyperosmotic stress [238], DNA damage [239], heat shock [240], oxidative stress [241], and amino acid deprivation the latter of which is the case in pSer51 on eIF2α. Many of the PTMs identified in databases arise from studies under which cells were subjected to stress including DNA damage [242], α-factor arrest of the cell cycle [243], and amino acid deprivation [215]. Consequently, PTMs reported in the literature but not here could be a result of PTMs occurring in response to factors not present here, as in this study cells were harvested under ideal growth conditions with an excess of nutrients.

4.4.1.2 Cross-linking of Yeast eIF2B

The structure of eIF2B has remained elusive with multiple proposals for its subunit arrangement [209, 210]. Experimental evidence has produced conflicting results with respect to the correct arrangement, with one study on endogenous *S. cerevisiae* eIF2B proposing a $\gamma\epsilon$ core and α subunits at opposing ends of the complex while other studies on overexpressed mammalian eIF2B [209], *Chaetomium thermophilum* eIF2B [244] and *E. coli*-expressed human eIF2B [245] have indicated that eIF2Bα dimerises to form an α_2 subcomplex and eIF2B has a core $\alpha_2\beta_2\delta_2$ regulatory subcomplex. To attempt to resolve some of this uncertainty, the subunit organisation of eIF2B was interrogated using chemical cross-linking.

It has been noted that purification of eIF2B without SEC can be insufficient in obtaining a homogenous sample, free of contaminants and aggregation [246]. An eIF2B protein sample that was purified by SEC prior to analysis was therefore used. eIF2B was cross-linked with a 1:1 mixture of non-deuterated (d0) and deuterated (d4) BS3. After cross-linking and LC-MS/MS analysis, 75 different cross-links were identified, of which 20 were inter-subunit cross-links, (Table 4.1). The inter-subunit cross-links were visualised with xVis software [247] under the assumption that eIF2B is decameric [210] with two copies of each of the five subunits present (Figure 4.6). All subunits formed inter-subunit cross-links with at least one other subunit. One cross-link was particularly interesting; α(K160)–α(K160) as it can only be rationalised by an eIF2Bα_2 unit within the complex structure, indicating that the eIF2B complex cannot contain two eIF2Bα subunits at opposing ends as proposed previously [210]. In addition, cross-links between eIF2Bα and ϵ cannot be accommodated in another model previously proposed, in which eIF2Bα_2 forms a core and eIF2Bϵ is located at the periphery [209, 214]. These cross-linking results, together with reported native MS

Subunit 1	Subunit 2	Residue 1	Residue 2	Peptide 1	Peptide 2	Distance/Å
α	α	160	160	CVVTESRPSKQGNQLYTLLEQK	PSKQGNQLYTLLEQK	16.0
α	ε	111	356	SRNKIAEIGVDFIADDDIILVHGYSR	CTAIGSGTKIGEGTKIENSVIGR	37.4
α	ε	145	329	AVFSLLNHAANKFIR	EKDVVLAQSCK	12.1
β	δ	232	587	KLAQHNIETLVVPDSAVFALMSR	GSEGESKDASNEEDSNSK	26.4
γ	δ	332	365	NHDDDDDYKPSATSIQPTYFK	KDLCEK	33.6
γ	δ	332	376	NHDDDDDYKPSATSIQPTYFK	IGQFAKEK	28.7
δ	γ	311	116	DYQTPKGTTLSR	KQQFELIYK	34.1
δ	γ	311	140	DYQTPKGTTLSR	KINFIPSK	27.5
γ	ε	302	208	QASFKDPFTGNQQQQNPPTTDDDEDR	EKTSIQIDPELLDNVDEFVIR	24.0
γ	ε	376	176	SWQHSKPR	ASTYPKTR	26.0
ε	γ	208	251	EKTSIQIDPELLDNVDEFVIR	TKYLQIR	17.1
ε	γ	208	376	EKTSIQIDPELLDNVDEFVIR	SWQHSKPR	32.9
δ	ε	421	356	VLTELLLHNAISLKK	IGEGTKIENSVIGR	23.1
δ	ε	422	356	KNIKVIVVDSRPLFEGR	IGEGTKIENSVIGR	21.9
δ	ε	425	356	NIKVIVVDSRPLFEGR	IGEGTKIENSVIGR	19.3
δ	ε	587	154	GSEGESKDASNEEDSNSK	KMHLQDK	27.9
ε	δ	189	547	TIEPAAFVLDKSTSR	RGNKGALLNQFIK	25.5
ε	δ	356	33	IGEGTKIENSVIGR	KLTNK	20.3
ε	δ	356	425	IGEGTKIENSVIGR	NIKVIVVDSR	30.1
ε	δ	375	425	IKNSFIWDDCIIGNNSIIDHSLIASNATLGSNVR	NIKVIVVDSRPLFEGR	19.8

Table 4.1: Inter-subunit chemical cross-links identified in yeast eIF2B with BS3 cross-linker. The protein subunits, cross-linked residues, cross-linked peptide sequences, and measured Cα-Cα distances are given. An inter-subunit cross-link from α to α was identified due to overlap of the cross-linked peptide sequences.

data [209, 210] and X-ray crystallography of eIF2Bα dimers [213] suggest that while an α_2 unit is present in the assembled complex and may play a role in stabilising the decamer, it is most likely not located at the core of the complex, but is found in close proximity to eIF2Bϵ.

During the course of this study, a new structure from X-ray crystallography of the related yeast *Schizosaccharomyces pombe* eIF2B was published [248]. Interestingly, this structure places two α subunits in a dimeric unit at the bottom of the complex, with a $\beta_2\delta_2$ core agreeing well with studies indicating eIF2Bα dimerisation [209, 244, 245] and formation of the $\alpha_2\beta_2\delta_2$ regulatory subcomplex [244, 245]. A model of eIF2B from *S. cerevisiae* was therefore constructed by homology modelling of individual subunits using the Phyre2 web server [237] and alignment to the *S. pombe* crystal structure [248]. Placing the cross-links identified here onto this homology model (Figure 4.7) shows that the

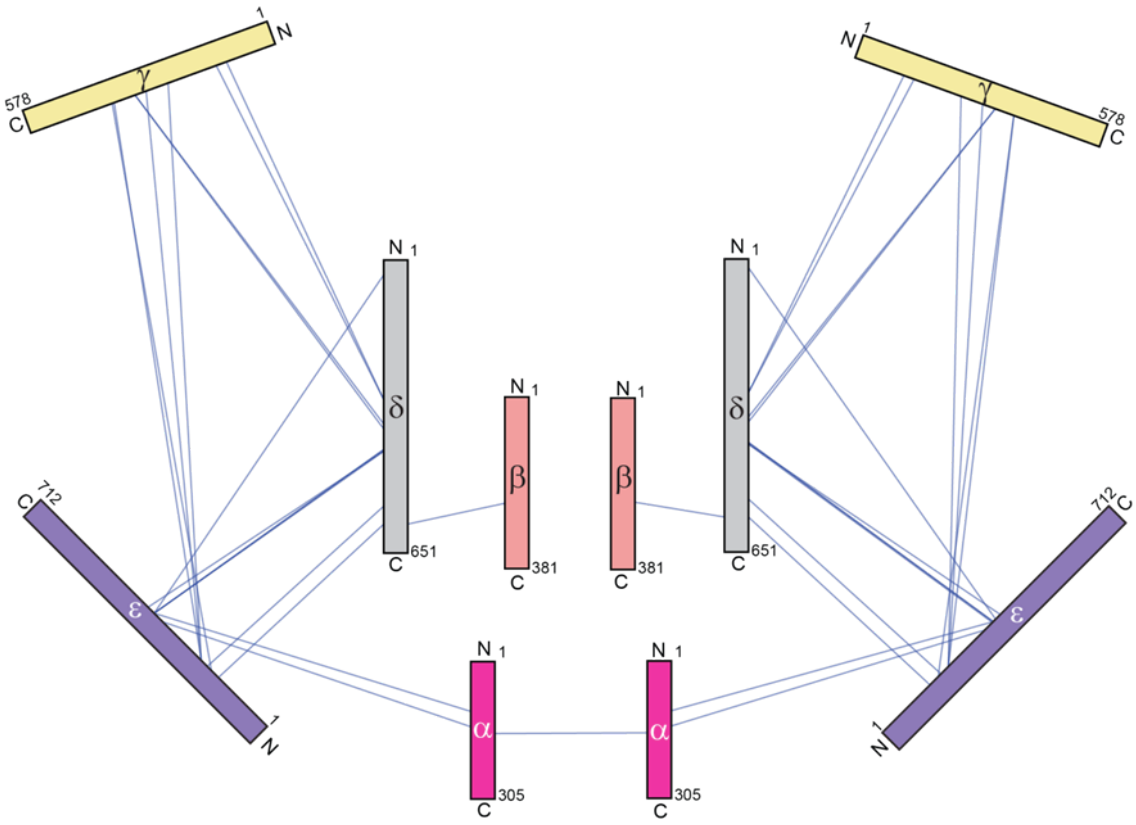


Figure 4.6: Identified eIF2B inter-subunit cross-links displayed as a network. Two copies of each of the five subunits are displayed, based on a decameric structure. All subunits cross-link to at least one other subunit. Subunit eIF2Bα contains an inter-subunit cross-link with another eIF2Bα subunit, indicating that an eIF2Bα_2 unit is present in the structure. Figure produced with xVis software [247].

cross-linking data agrees well with the crystal structure, with reasonable measured Cα-Cα distances (Table 4.1). In particular, the 'eIF2Bα_2' cross-links identified here are fully rationalised by the subunit arrangement of eIF2B from *S. pombe*.

4.4.2 Human eIF2B

4.4.2.1 Native Mass Spectrometry of eIF2B

Having explored the role of PTMs as a mechanism for regulation in endogenously-expressed eIF2B (*S. cerevisiae*) and placed these in the context of the structure, I next considered the non-natural regulation of eIF2B by the small

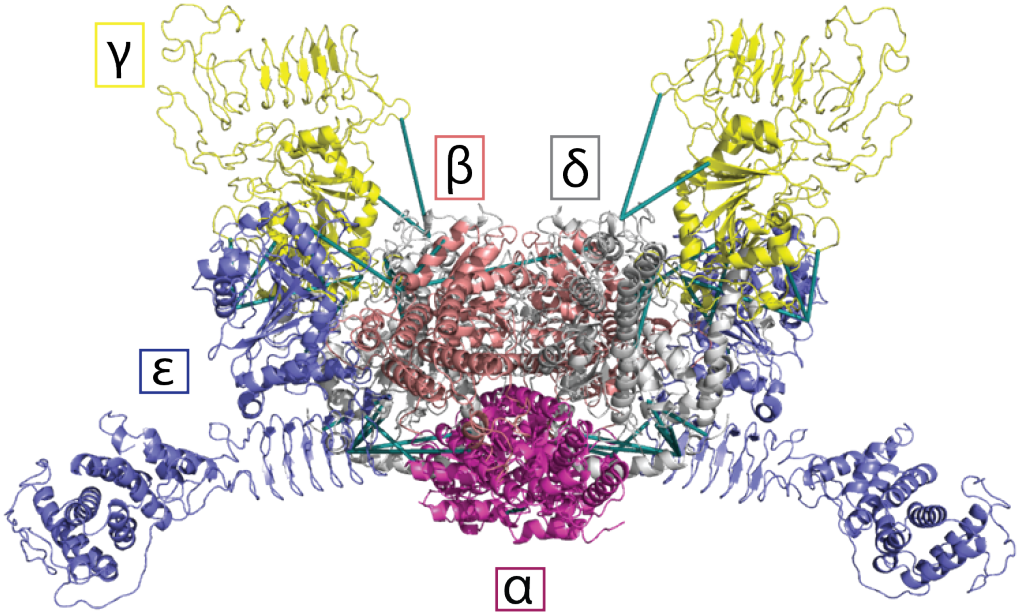


Figure 4.7: Identified inter-subunit cross-links displayed on a homology model of eIF2B from *S. cerevisiae*, with eIF2B subunits α (fuchsia), β (light pink), γ (yellow), δ (grey), and ϵ (purple). All cross-links can be rationalised with this model, although longer distances for some cross-links indicate either flexibility or deficiencies in the homology model. The model was generated by homology modelling of individual eIF2B subunits using the Phyre2 web server [237] and aligning with the crystal structure of eIF2B from *S. pombe* (PDB ID: 5B04 [248]).

molecule ISRIB. For this work, human eIF2B overexpressed in *E. coli* was utilised, to reduce the sample heterogeneity as found in the endogenously-expressed yeast protein and to investigate the human homologue, which is particularly interesting for drug-discovery. Purified human eIF2B samples were provided by Jordan Tsai and Professor Peter Walter (The University of California, San Francisco).

Firstly, to determine the stoichiometry of human eIF2B, the sample was analysed by native MS. In the mass spectrum (Figure 4.8A), a charge-state series is observed centred around 10,000–13,000 m/z, to which a mass of 525,670 ±89 Da is assigned, corresponding to the decameric eIF2B species; $\alpha_2\beta_2\gamma_2\delta_2\epsilon_2$ (Appendix Table B.2). Under these conditions, this complex dominates the spectrum. The broadness of the

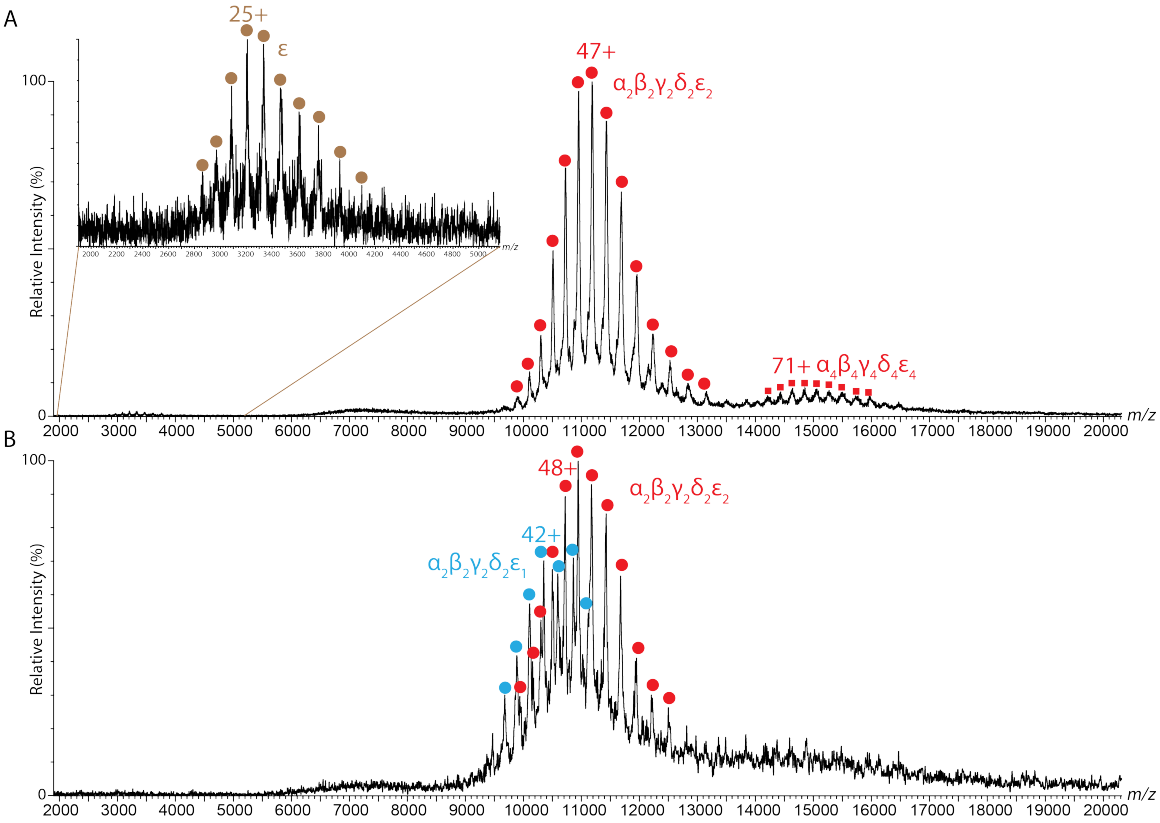


Figure 4.8: Native mass spectrum of recombinantly-expressed human eIF2B recorded on a QToF II Mass Spectrometer. (A) The main charge-state series observed is the decameric species $\alpha_2\beta_2\gamma_2\delta_2\epsilon_2$ (red circles, 525,670 ±89 Da). A small population of $\alpha_4\beta_4\gamma_4\delta_4\epsilon_4$ is observed (red squares, 1,053,714 ±127 Da) which is likely non-specific. Expansion of the low m/z region (inset) reveals dissociated subunit ϵ (brown circles, 80,213 ±11 Da). On close inspection of the main $\alpha_2\beta_2\gamma_2\delta_2\epsilon_2$ species, a second species is apparent. (B) When the collision voltage is increased from 40 V to 150 V, these peaks become more intense and resolved (blue circles, 445,951 ±99 Da), allowing assignment of the ϵ-dissociated complex $\alpha_2\beta_2\gamma_2\delta_2\epsilon_1$.

peaks is greater than expected and could be a result of heterogeneity in the eIF2B complex. On closer inspection, a smaller peak series of lower intensity was observed. To further resolve this species, the collision voltage was increased and the resulting mass spectrum (Figure 4.8B) displayed a second well-resolved charge-state series with increased intensity, in place of the previously-unresolved peaks, with a mass of 445,951 ±99 Da corresponding to the decameric complex which lost one ϵ copy; $\alpha_2\beta_2\gamma_2\delta_2\epsilon_1$.

To further probe the subunit architecture of the decameric eIF2B species, tandem MS was employed. The 46+ charge state of the eIF2B complex was selected and subjected to CID (Figure 4.9). The decamer complex was found to be reasonably stable under the conditions employed here, however two small populations of charge-stripped complexes were formed, one corresponding to the loss of subunit α

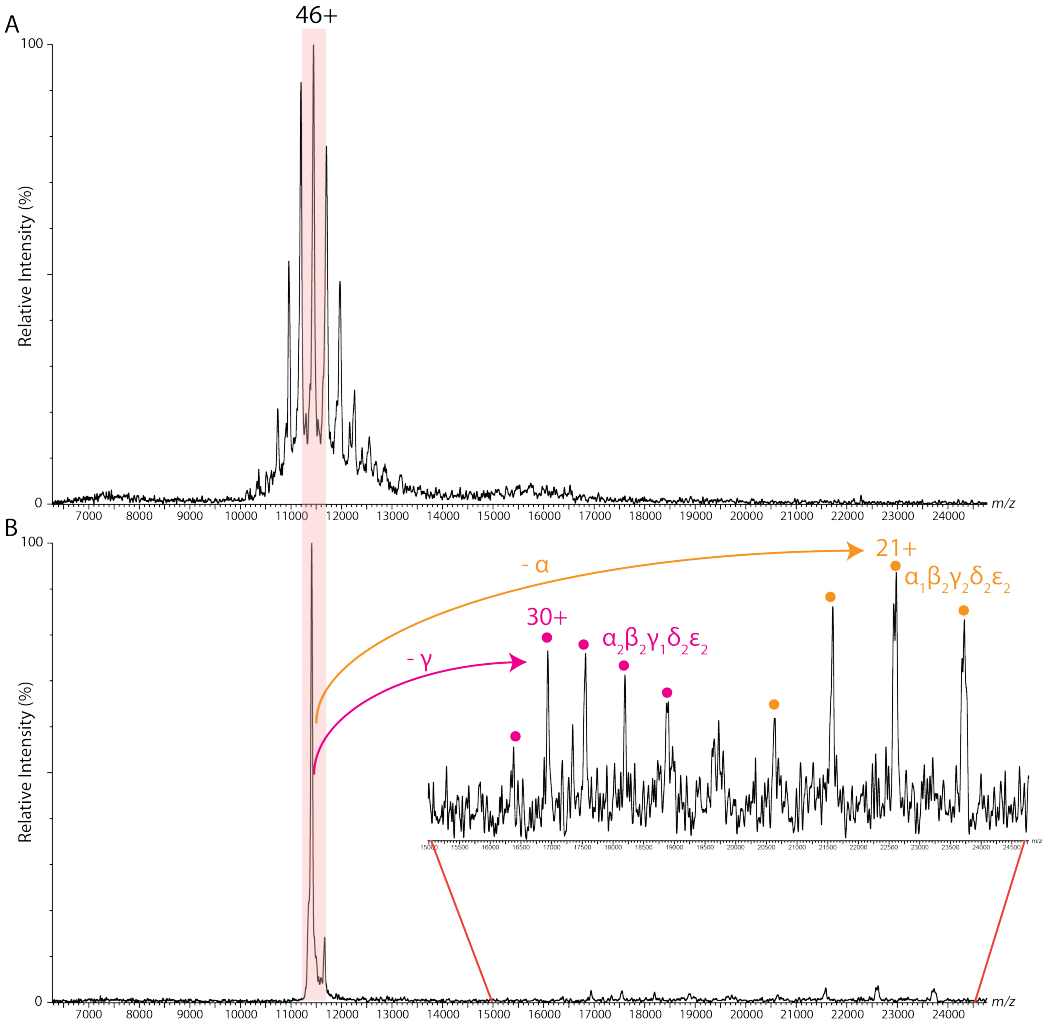


Figure 4.9: (A) Mass spectrum of the intact eIF2B decamer recorded on a QToF II Mass Spectrometer with a collision voltage of 40 V. The 46+ charge state was selected for tandem MS analysis (selection of 11,474 m/z) (B) and a collision voltage of 200 V applied, with two charge-stripped complexes observed in the high m/z region (expanded in inset). These complexes correspond to loss of subunits γ and α to form $\alpha_2\beta_2\gamma_1\delta_2\epsilon_2$ (pink circles, 475,165 $\pm$61 Da) and $\alpha_1\beta_2\gamma_2\delta_2\epsilon_2$ (orange circles, 491,767 $\pm$70 Da) complexes, respectively.

($\alpha_1\beta_2\gamma_2\delta_2\epsilon_2$) and the other to the loss of subunit γ ($\alpha_2\beta_2\gamma_1\delta_2\epsilon_2$). Combining these results suggests that subunits α, γ, and ϵ are more labile than subunits β and δ.

As the decameric eIF2B complex was transferred intact into the gas phase under the MS conditions employed, with little dissociation by CID, the complex was disrupted in-solution with 15% methanol, prior to MS analysis, to further investigate subunit interactions. After this disruption, a number of subcomplexes were observed in the mass spectrum (Figure 4.10). The predominant complex was still the intact $\alpha_2\beta_2\gamma_2\delta_2\epsilon_2$ complex, however significant populations of the individual subunits and subcomplexes α, α_2, $\gamma\epsilon$, and $\beta\gamma\delta\epsilon$ were observed.

4.4.2.2 eIF2Bα Dimerises

As the mass spectrum after in-solution disruption showed the presence of both monomeric and dimeric eIF2Bα, and considering previous reports of human eIF2Bα

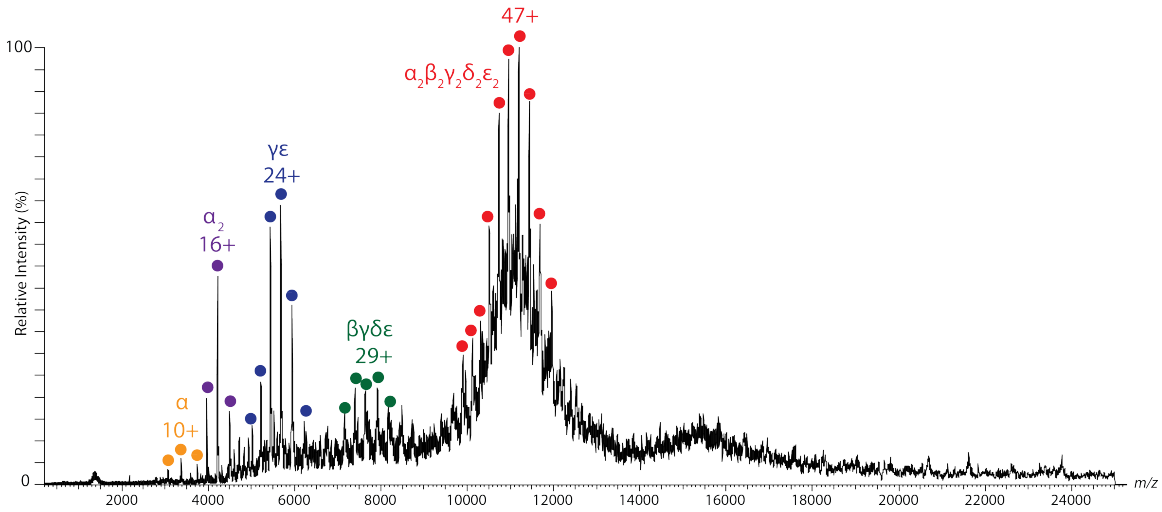


Figure 4.10: Native mass spectrum of recombinantly-expressed human eIF2B disrupted in-solution with 15 % methanol prior to MS analysis. The predominant complex is the intact $\alpha_2\beta_2\gamma_2\delta_2\epsilon_2$ complex (red circles, 526,091 ±81 Da). The subcomplexes α (orange circles, 33,741 ±2 Da), α_2 (purple circles, 67,466 ±9 Da), $\gamma\epsilon$ (dark blue circles, 130,534 ±29 Da), and $\beta\gamma\delta\epsilon$ (dark green circles, 229,249 ±80 Da) are observed.

dimerisation in isolation [249] and the α_2 unit in the structure of *S. pombe* eIF2B [248], the oligomeric state of isolated eIF2Bα was investigated using MS. Purified human eIF2Bα was provided by Jordan Tsai and Professor Peter Walter (The University of California, San Francisco). In the mass spectrum (Figure 4.11) a charge-state series corresponding to eIF2Bα_2 is observed, indicating that eIF2Bα is predominantly dimeric under the conditions employed here. Significant peak broadening was observed for subunit eIF2Bα, indicating a high degree of heterogeneity in the sample, arising from possible truncations and phosphogluconylation of the His-tag [250].

4.4.2.3 Native Mass Spectrometry of eIF2B and ISRIB

Having characterised the *apo* human eIF2B complex, the interaction of eIF2B with the small molecule ISRIB was next investigated. As velocity sedimentation has indicated that in the presence of ISRIB larger eIF2B complexes form than without

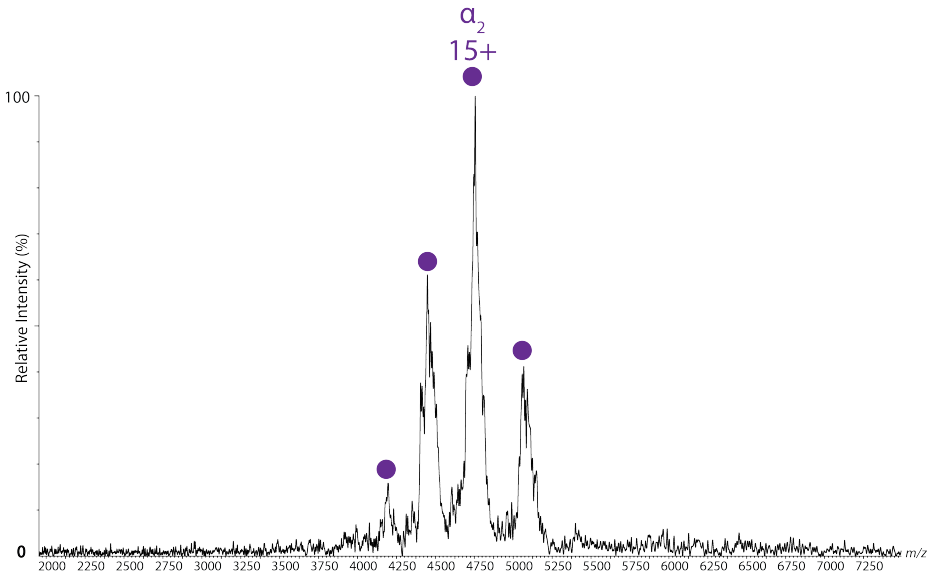


Figure 4.11: Native mass spectrum of recombinantly-expressed human eIF2Bα recorded on a QToF II Mass Spectrometer at a collision voltage of 25 V. Under these conditions, the oligomeric state is dimeric (71,386 ±37 Da).

ISRIB treatment [233], the effect of ISRIB incubation on eIF2B was investigated by native MS. Human eIF2B was incubated with ISRIB (solubilised in DMSO) or with DMSO as a control before the buffer was exchanged to 500 mM ammonium acetate prior to MS analysis. The resulting mass spectra (Figure 4.12) are highly similar to one another with the predominant complex both in the presence and absence of ISRIB being the intact decamer, i.e. eIF2B$\alpha_2\beta_2\gamma_2\delta_2\epsilon_2$.

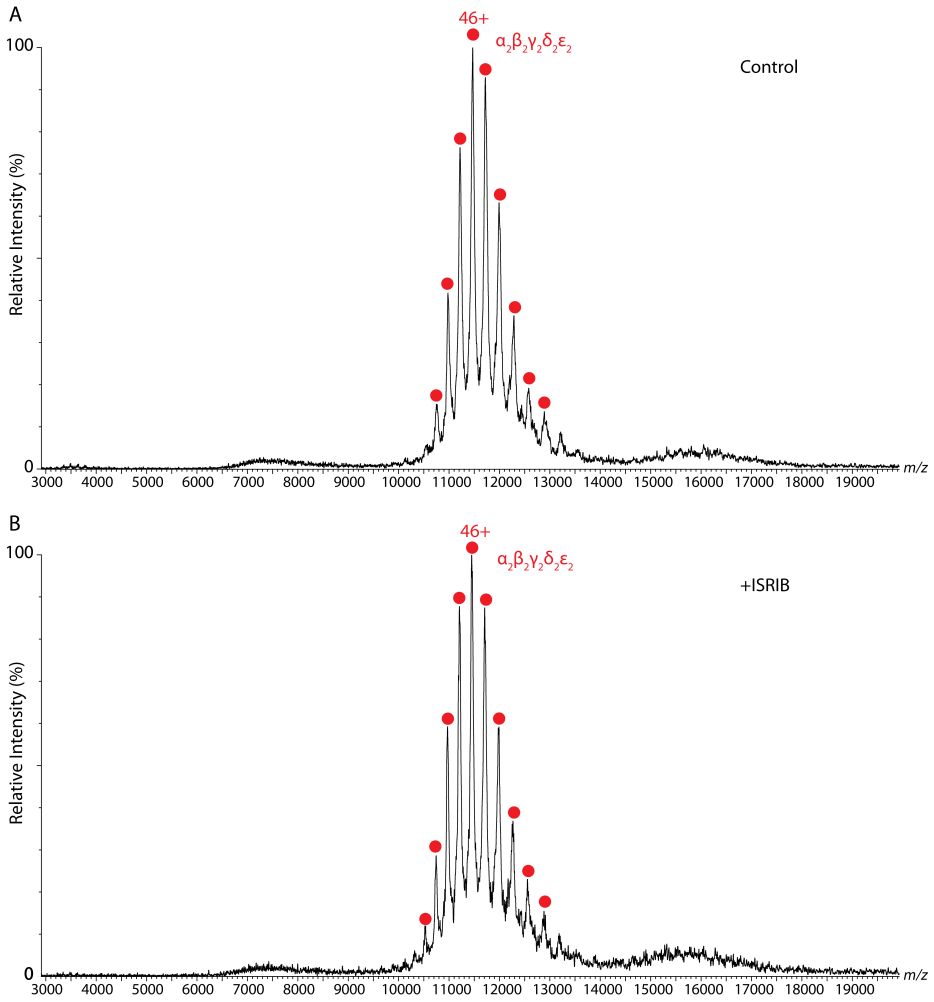


Figure 4.12: Native mass spectra of recombinantly-expressed human eIF2B incubated with (A) DMSO (Control) or (B) ISRIB solubilised in DMSO (+ISRIB), prior to buffer exchange, recorded on a QToF II Mass Spectrometer. The intact eIF2B decamer is observed in both spectra (519,075 ±129 Da and 518,360 ±74 Da).

As eIF2B was observed as a decamer under the conditions here, even without the potentially stabilising effect of ISRIB, the sample was incubated with ISRIB for an extended time period and analysed by native MS using higher-energy conditions designed to test the stability of the complexes formed. Without ISRIB present (Figure 4.13A) multiple homomeric subunits and subcomplexes were observed, including eIF2Bα, γ, α_2, $\gamma\epsilon$, $\beta\gamma\delta\epsilon$, and $\alpha_2\beta_2\gamma_2\delta_2\epsilon_1$. When incubated in the presence of ISRIB (Figure 4.13B) many of the same complexes formed. However, in the ISRIB-treated sample there was a modest decrease in the intensity, measured from the most intense peak, of dissociated subunit eIF2Bγ (from $\approx$ 15 to 8 %), α_2 (from $\approx$ 17.5 to 12 %) and the subcomplex $\gamma\epsilon$ (from $\approx$ 87.5 to 67.5 %), relative to the intensity of the decamer. Most notably, the population of eIF2B$\beta\gamma\delta\epsilon$, observed in the control sample with a well-resolved charge-state series, was not detected although a small population of this species may be present and unassignable due to peak overlap and an elevated baseline.

To further study the stability of the complex in response to ISRIB treatment, the subunit interfaces of the complex were disrupted in-solution. Aliquots of eIF2B were incubated with ISRIB/DMSO followed by exchange of the buffer to ammonium acetate. The samples were then acidified with acetic acid to disrupt the complex, prior to MS analysis. The resulting mass spectra (Figure 4.14) show that although some of the intact decamer remains, significant populations of subcomplexes are now observed. A second set of charge-state series is observed alongside the charge-state series for the two larger complexes (green and red stars). These are each approximately 1.8 kDa lighter than their 'associated' charge-state series (green and red circles). Their identity is not clear but may be due to degradation of a protein subunit under the acidic conditions used here. A decrease in the intensity of dissociated eIF2Bα_2 (from $\approx$ 50 to 20 % relative intensity, measured by the most

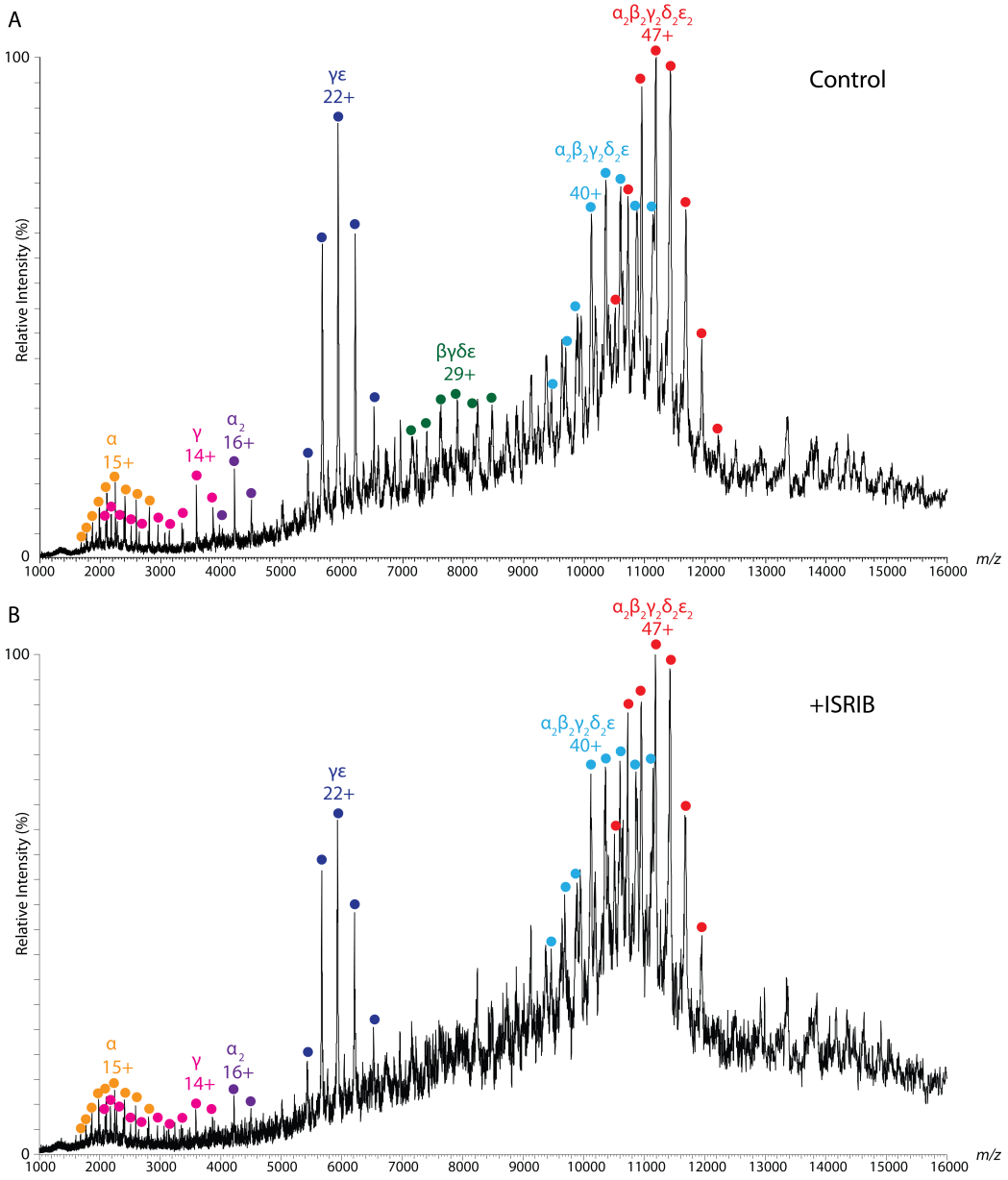


Figure 4.13: Native mass spectra of recombinantly-expressed human eIF2B incubated overnight with (A) DMSO (Control) or (B) ISRIB solubilised in DMSO (+ISRIB). The most intense complex is decameric $\alpha_2\beta_2\gamma_2\delta_2\epsilon_2$ (red circles, A: 525,646 ±59 Da, B: 525,683 ±71 Da). A complex which lost one copy of ϵ is observed (light blue circles, A: 445,036 ±44 Da, B: 445,005 ±24 Da) as well as the subcomplexes $\gamma\epsilon$ (dark blue circles, A: 130,461 ±18 Da, B: 130,433 ±15 Da), and α_2 (purple circles, A: 67,470 ±10 Da, B: 67,470 ±8 Da), and subunits γ (pink circles, A: 50,239 ±1 Da, B: 50,238 ±2 Da), and α (orange circles, A: 33,739 ±1 Da, B: 33,724 ±5 Da). In the control sample, a tetrameric $\beta\gamma\delta\epsilon$ complex (dark green, A: 229,145 ±25 Da) is observed. Following incubation with ISRIB, most of the complexes remain unchanged, however the $\beta\gamma\delta\epsilon$ complex is not observed and modest differences in intensity are observed for some complexes, including α, α_2, γ, and $\gamma\epsilon$.

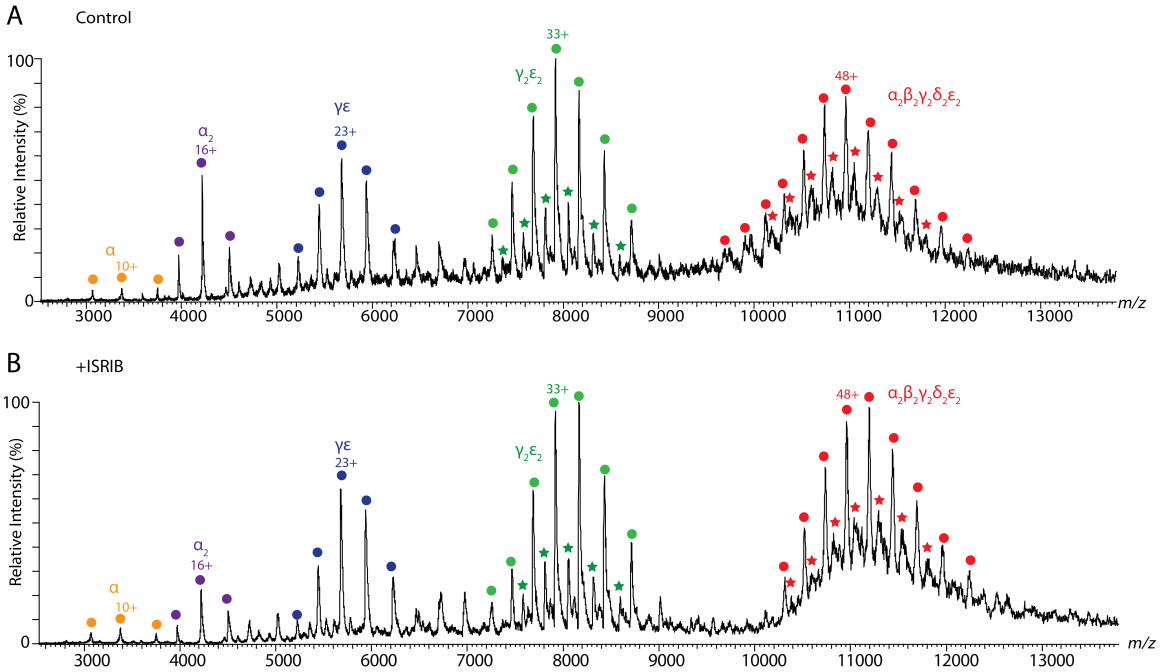


Figure 4.14: Native mass spectrum of recombinantly-expressed human eIF2B disrupted in-solution with acetic acid, with (A) DMSO (Control) or with (B) ISRIB solubilised in DMSO (+ISRIB). The predominant charge-state series observed corresponds to $\alpha_2\beta_2\gamma_2\delta_2\epsilon_2$ (red circles, A: 526,167 ±55 Da, B: 526,116 ±42 Da). Additional peak series are observed; $\gamma_2\epsilon_2$ (green circles, A: 261,256 ±15 Da, B: 261,332 ±18 Da), $\gamma\epsilon$ (blue circles, A: 130,556 ±8 Da, B: 130,587 ±10 Da), α_2 (purple circles, A: 67,491 ±10 Da, B: 67,502 ±14 Da), and α (orange circles, A: 33,737 ±1 Da, B: 33,755 ±14 Da). Additional species are assigned to degraded $\alpha_2\beta_2\gamma_2\delta_2\epsilon_2$ (red stars, A: 519,212 ±43 Da, B: 519,208 ±80 Da) and $\gamma_2\epsilon_2$ (green stars, A: 257,725 ±18 Da, B: 257,804 ±44 Da) complexes.

intense peak) is observed in the ISRIB-treated sample compared to the control, indicating that the interactions between eIF2Bα_2 and the other protein subunits are strengthened in the presence of ISRIB, a conclusion which is supported by the previous experiments (Figure 4.13). An eIF2B$\gamma_2\epsilon_2$ complex at high intensity is observed in both spectra, the same complex observed by Gordiyenko *et al.* [210].

4.5 Discussion

4.5.1 Post-Translational Modification of Yeast eIF2B

Translation initiation is a tightly-regulated process and the identified phosphorylation and acetylation sites on eIF2 and eIF2B from *S. cerevisiae* may play a regulatory role. In particular, eIF2Bϵ was found to be highly modified and locating the sites identified here and in the databases onto the generated homology model (Figure 4.15) reveals these modification sites to be enriched on the linker connecting the catalytic domain to the core of the complex. In addition, another protein region which shows a similar clustering of modification sites is the N-terminal region of eIF2β. It is notable that the linker region of eIF2Bϵ is involved in the interaction of eIF2Bϵ and eIF2. Furthermore, the N-terminal region of eIF2β has been reported to interact with eIF5 and the catalytic HEAT domain of eIF2Bϵ during guanine nucleotide exchange [201, 251]. The mechanism of eIF5 displacement from eIF2 by eIF2B is currently unknown, but the high level of modification at these critical interfaces suggests that PTMs may play a crucial role in regulating these competing interactions by stabilising and destabilising subunit interfaces.

A role has already been assigned to some PTMs, such as the Ser525 phosphorylation site on eIF2Bϵ, which is critical in regulation [252]. This is located on the flexible linker connected to the catalytic domain, on which several other PTMs were clustered. PTMs are rarely at 100% occupancy, and the clustering of sites seen on eIF2Bϵ and also on eIF2β may be on account of multiple modification sites acting as alternative sites for a similar role.

To explore the biological relevance of the PTMs identified here in the wider context of eukaryotic translation initiation, the conservation of these PTMs across different

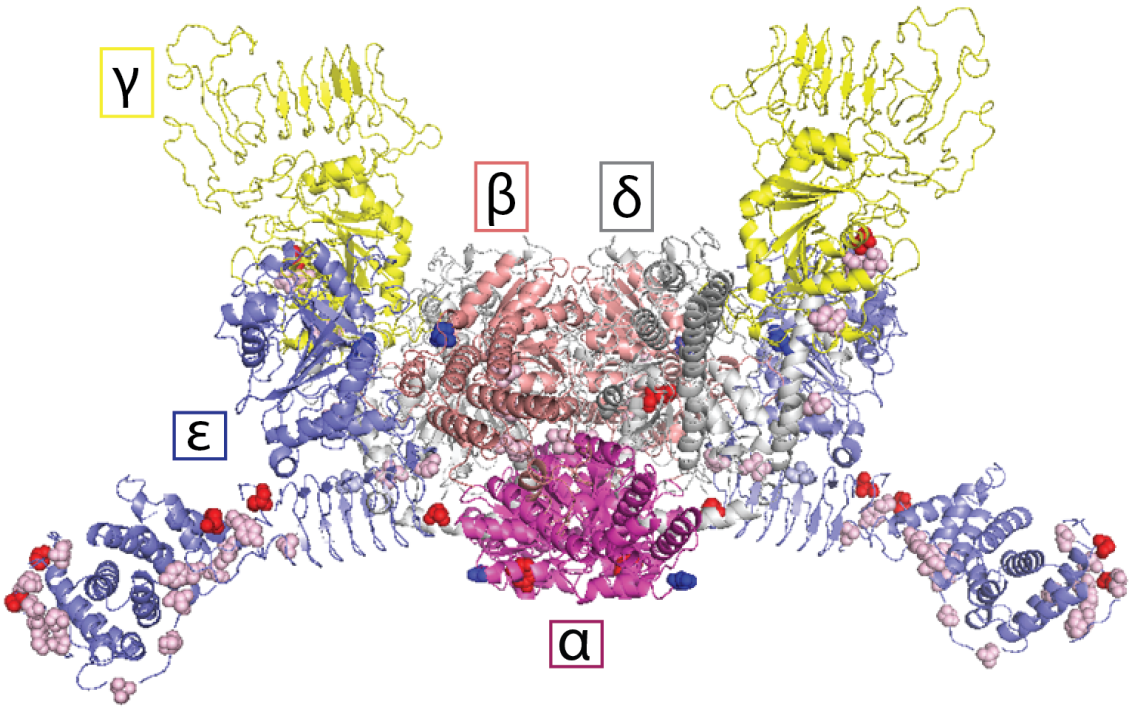


Figure 4.15: Post-translational modifications of eIF2B located on a homology model of eIF2B, with two copies each of eIF2Bα (fuchsia), β (light pink), γ (yellow), δ (grey), and ϵ (purple). Phosphorylation sites identified in this study (red) and the literature (light pink) as well as acetylation sites identified in this study (dark blue) and in the literature (light blue) are highlighted as spheres. The eIF2Bϵ subunit in particular is highly-modified, with phosphorylation sites concentrated on the flexible linker between the catalytic HEAT domain and the non-catalytic domain. The homology model was generated with the Phyre2 web server [237] and manually aligned to the structure of eIF2B from *S. pombe* (PDB ID: 5B04 [248]) using Pymol.

species was investigated using the PTM*func* database [170]. Of the 20 phosphorylation sites identified in this study, ten are conserved in *Candida albicans, Caenorhabditis elegans, Drosophila melanogaster, H. sapiens,* or *S. pombe* (Appendix Table B.1). The fact that they are highly conserved suggests a functional or structural role, which in some cases has already been ascribed [200, 252].

4.5.2 Chemical Cross-Linking of Yeast eIF2B

Cross-linking of the yeast eIF2B complex revealed multiple inter-subunit cross-links. The stoichiometry of eIF2B, with two copies of each subunit, presents a challenge in the analysis of cross-linking data. However, one of the cross-links was determined to arise from interactions of two copies of the same subunit deduced from overlap of cross-linked peptides. The cross-link α(K160)–α(K160) places two eIF2Bα subunits in close proximity. This interaction is particularly interesting as it cannot be accommodated in the model proposed by Gordiyenko *et al.* [210], but can be accommodated in the homology model based on the crystal structure solved by Kashiwagi *et al.* [248]. Surprisingly, few cross-links were observed between subunit α and any other subunit, with the exception of two α–ϵ cross-links. This could be the result of dissociation of subunit α, which the native MS results indicate to be labile.

4.5.3 Native Mass Spectrometry of Human eIF2B

Human eIF2B was observed as a heterodecamer by native MS, with a mass corresponding to an $\alpha_2\beta_2\gamma_2\delta_2\epsilon_2$ complex, consistent with the stoichiometry observed previously for human eIF2B by native MS [210] and for the yeast (*S. pombe*) eIF2B by X-ray crystallography [248]. Dissociation of subunits α, γ, and ϵ indicate that these are located at the periphery of the complex, corroborated by the yeast crystal structure [248], in which eIF2B$\beta_2\delta_2$ comprises the core surrounded by two distal $\gamma\epsilon$ units and an α_2 unit.

In-solution disruption of the complex with methanol generated a number of subcomplexes. Subunits γ and ϵ dissociated as the heterodimer $\gamma\epsilon$, with the individual subunits not observed. Inspection of the yeast crystal structure [248] shows that γ and ϵ are adjacent to one another and are located on the periphery of the complex, rationalising the observation of the $\gamma\epsilon$ complex after in-solution

disruption. Subunit α dissociates as both monomer and dimer and when eIF2Bα is purified alone and analysed by native MS, an eIF2Bα dimer is detected, as supported by a similar analysis of eIF2Bα by native MS and SDS-PAGE [209]. This dimeric form can be rationalised by the presence of a Rossmann fold domain [209] and this dimeric unit is observed in the yeast eIF2B crystal structure [248]. A subcomplex of $\beta\gamma\delta\epsilon$ is perhaps surprising given that it might be expected that the decameric $\alpha_2\beta_2\gamma_2\delta_2\epsilon_2$ complex would dissociate into two similar $\alpha_1\beta_1\gamma_1\delta_1\epsilon_1$ complexes. This, together with the discussed dimerisation of subunit α, indicates that interactions between the two α subunits are stronger than those connecting α to the eIF2B complex. As such, an octamer lacking subunit α ($\beta_2\gamma_2\delta_2\epsilon_2$) might be expected, however the tetrameric $\beta\gamma\delta\epsilon$ complex is observed in the mass spectrum. Most likely, subunit α stabilises eIF2B in the intact decamer, as opposed to the pentamer, a conclusion supported by data from Wortham *et al.* [209].

Other than $\beta\gamma\delta\epsilon$, no smaller subcomplexes containing β or δ are observed as one might expect on account of the large populations of $\gamma\epsilon$ and α observed. This is likely due to low ionisation efficiencies, and consequently low signal intensities, for these species. If eIF2Bβ and δ are assembled as an eIF2B$\beta_2\delta_2$ tetramer, the high pK_a of eIF2Bδ (9.45) could cause a low ionisation efficiency in combination with signal suppression from other abundant species.

4.5.4 The Interaction of ISRIB and Human eIF2B

It has been suggested that ISRIB may stabilise the eIF2B decamer via bridging of two pentameric $\alpha\beta\gamma\delta\epsilon$ units [233]. In this study, addition of ISRIB to the samples without extended incubation or complex destabilisation did not induce significant changes in complex formation. The stability of the decameric complex employed here may be a result of the protein concentrations required for native MS which are

above those utilised by Sidrauski *et al.* [233].

When eIF2B is incubated overnight and destabilised, a $\beta\gamma\delta\epsilon$ complex is observed. When incubated with ISRIB, this subcomplex is not detected in the mass spectrum. As subunit α is often observed in the mass spectrum as a dissociated dimer, it is likely that the $\beta\gamma\delta\epsilon$ complex is formed by dissociation of α_2 from the decameric $\alpha_2\beta_2\gamma_2\delta_2\epsilon_2$ complex to form two copies of $\beta\gamma\delta\epsilon$ subcomplexes. This implies therefore that ISRIB stabilises eIF2B in the intact decamer conformation. A decrease in dissocated α and α_2 in the ISRIB-treated sample further suggests that while subunit α stabilises the two tetramers into the intact decamer, ISRIB not only stabilises the interaction between the two tetramers but also stabilises the interface between α and the octameric $\beta_2\gamma_2\delta_2\epsilon_2$ complex by stabilisation of the $\beta_2\delta_2$ core (Figure 4.16).

By locating the mutations of eIF2Bδ sensitive to the response of eIF2B to ISRIB [234] on the structure of eIF2B [248], Kashiwagi *et al.* [253] were able to propose a preliminary binding mode of ISRIB in which ISRIB bridges the two β subunits. This, together with the native MS results obtained here, suggests that ISRIB stabilises the eIF2B decamer by bridging of the two tetrameric $\beta\gamma\delta\epsilon$ complexes, holding them in place, which is then further stabilised by α_2 (Figure 4.16).

Addition of acetic acid to the complex resulted in significant disruption, although a sizeable population of the decamer remained, albeit with a truncation indicating some form of acid-mediated degradation. More α_2 relative to decamer is seen in the control sample compared to the ISRIB-treated sample, indicating that ISRIB stabilises α_2 binding to the complex. This finding is supported by data reported by Sidrauski *et al.* [233] who noted that in velocity sedimentation analysis of

ISRIB-treated cells, a higher proportion of eIF2Bα migrated from lighter fractions (thought to be the pentameric $\alpha\beta\gamma\delta\epsilon$) to heavier fractions (thought to be decameric $\alpha_2\beta_2\gamma_2\delta_2\epsilon_2$) than in control cells. It was suggested by Sidrauski *et al.* that ISRIB binding stabilises α_2 binding to $\beta_2\gamma_2\delta_2\epsilon_2$, a suggestion which is supported by the native MS data reported here.

The observation of $\gamma_2\epsilon_2$ complexes was made by Gordiyenko *et al.* when eIF2B was disrupted with ACN [210]. As this complex is not detected without in-solution

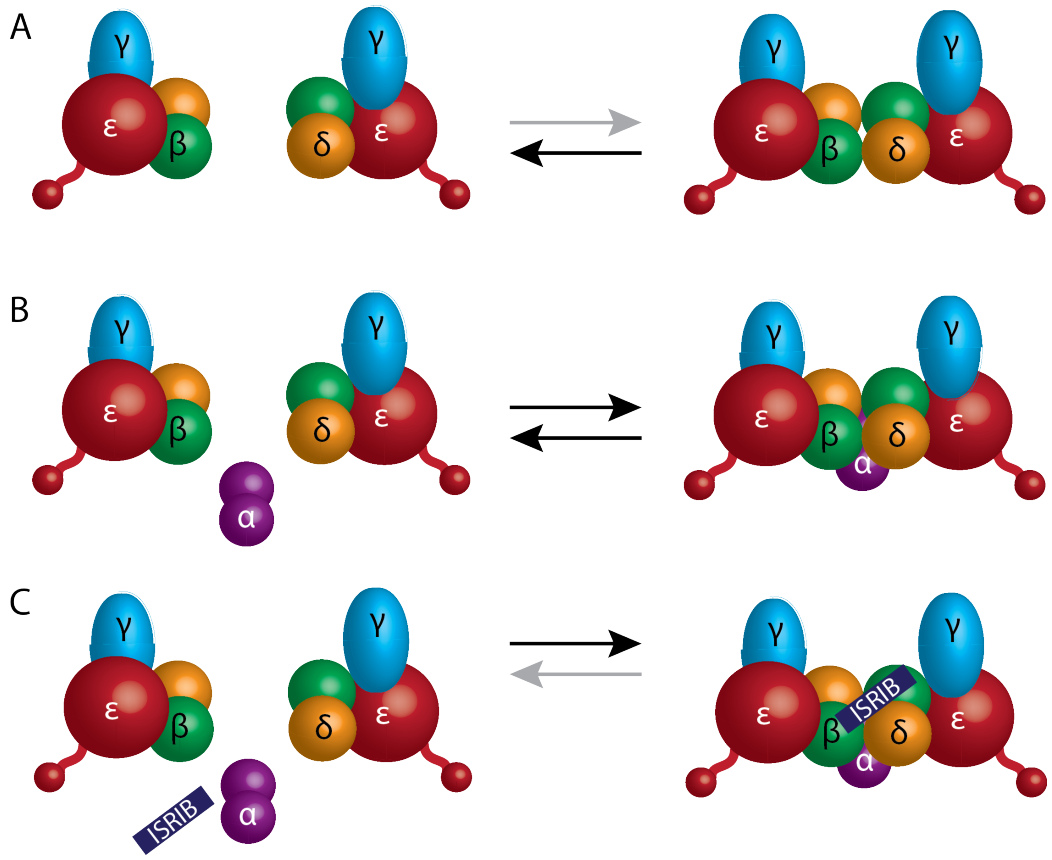


Figure 4.16: Proposed model for dissociation of decameric eIF2B into α_2 and two tetrameric $\beta\gamma\delta\epsilon$ complexes. (A) When α is not present to stabilise the full complex, the equilibrium between intact complex and tetramers shifts towards the tetrameric $\beta\gamma\delta\epsilon$ complexes. (B) When all five subunits are present, subunit α stabilises the decamer. (C) The small molecule ISRIB bridges the two β subunits, holding the two halves of the complex together, stabilising α association which further stabilises the decamer.

disruption both here and by Gordiyenko *et al.*, it is possible that this complex is not present under typical *in vivo* conditions and is the result of disruption to the native subunit interfaces, allowing non-native contacts to form between subunits γ and ϵ. In EM analysis however, investigation of yeast eIF2B complexed to gold nanoparticles on eIF2Bγ uncovered a conformational change in which the two γ subunits move together and make contact [246], suggesting that a conformational change may be responsible for the observed $\gamma_2\epsilon_2$ complex. The particles utilised for this reported EM study were purified with the full complement of subunits as eIF2B$\alpha_2\beta_2\gamma_2\delta_2\epsilon_2$, yet the particles displaying closer γ–γ interactions were those lacking eIF2Bα. It was therefore proposed that this conformational change may occur only when α is absent, and the $\gamma_2\epsilon_2$ tetramer observed by EM may have arisen due to disruption of the α_2 binding interface.

As eIF2Bα is known to dissociate from eIF2B *in vivo* [212], it is likely that this structural change also occurs *in vivo* with the γ–γ interaction possibly occluding or creating new reaction interfaces. If ISRIB not only stabilises the decameric complex but also stabilises α_2 binding, then ISRIB may reduce the incidence of such a conformational change, which could increase the proportion of eIF2B available for guanine nucleotide exchange on eIF2. It is notable that there is a small decrease in the relative intensity of $\gamma_2\epsilon_2$ tetramer compared to the decamer in the ISRIB-treated sample compared to the control suggesting that this may be the case however the change observed here was modest. Further studies on the effect of both eIF2Bα and ISRIB on the conformation of eIF2B, for example using EM with gold nanoparticle labelling, would help clarify if such a conformational change is influenced in this way. The work presented here draws further attention to a potential ISRIB-induced conformational change in eIF2B which is important for further research, especially as it would likely modify the eIF2B reaction interfaces significantly.

Considering these results in the context of *in vivo* studies, such as the effect of ISRIB on cognition after TBI, in which ISRIB returns cognition after TBI back to normal non-injured levels [254], suggests that this effect is the result of stabilisation of the eIF2B decamer. With this mechanism, many questions still remain with some of the most pertinent surrounding the issues of what causes eIF2B destabilisation after TBI and, in relation to treatment with ISRIB, how do the positive ISRIB-mediated effects remain long after the half-life of ISRIB, when one would expect ISRIB to have degraded [254].

Taking the first of these issues with regards to TBI-induced eIF2B instability, several possibilities can be proposed. The location of ISRIB binding is in a well-formed hydrophobic pocket observed by EM [255, 256], with this pocket observed in ISRIB-bound and *apo* eIF2B structures. It is possible that the ISRIB binding site serves as the binding site for other ligands, possibly metabolites, which could serve either stabilising or destabilising roles in healthy, or TBI subjects, respectively. In the case of a stabilising ligand this could be present only prior to TBI whereas in the case of a destabilising ligand, this could be present only after TBI. In both cases, ISRIB treatment may only prove therapeutic in TBI subjects, with either binding in the absence of a stabilising ligand or displacement of a destabilising ligand promoting eIF2B stability. It is therefore interesting that ISRIB did not appear to enhance cognition in the normal, healthy subjects [254]. Another mechanism of eIF2B instability in TBI subjects could be disruption of the interface between eIF2Bα and eIF2B$\beta_2\gamma_2\delta_2\epsilon_2$. Dissociation of α may not only destabilise eIF2B into two tetramers but also prevent eIF2 binding, as it has been proposed that the eIF2 binding site involves subunits eIF2Bsα, β, and δ [248, 253]. The influence of ISRIB on the interaction of eIF2B with eIF2 is consequently an interesting area for future investigation.

4.6 Conclusions

In this chapter the PTM status of endogenously-expressed eIF2 and eIF2B from *S. cerevisiae* was surveyed. Analysis revealed highly-modified subunits, with clustering of sites on the flexible linker connecting the catalytic and non-catalytic subdomains of eIF2Bϵ and also in a flexible region of eIF2Bγ and the N-terminus of eIF2β. These clusters of modifications occur at critical interfaces where eIF2 interacts with eIF2B and eIF5 and are therefore promising sites for further study. To investigate this, native MS could be used to analyse complexes formed by incubation of eIF2 with eIF2B and eIF5 both in the presence and absence of phosphatase enzymes. Furthermore, native MS could also be used to analyse complexes formed in phosphomimic variants of eIF2 and eIF2B at the conserved phosphorylation sites identified on flexible protein regions in this chapter (e.g. mutation of S40E on eIF2β).

Continuing with endogenously-expressed eIF2B from *S. cerevisiae*, chemical cross-linking defined inter-subunit interactions and revealed the close proximity of two copies of eIF2Bα, indicating an eIF2Bα_2 subcomplex within the complex. The cross-links are accommodated in a homology model based on the structure of the related yeast (*S. pombe*) eIF2B determined by X-ray crystallography and published since this cross-linking analysis was performed. Locating the previously-determined PTMs onto this structure reveals that the PTMs are located along protein-protein interfaces and therefore may play an important function in regulating interactions of the eIF2B enzyme. To investigate this, eIF2B could be treated with phosphatase and deacetylase enzymes followed by analysis by native MS or comparative cross-linking in order to assess changes in eIF2B complex stability and conformation.

Moving onto human eIF2B, native MS confirmed a decameric stoichiometry, with CID and in-solution disruption providing evidence that supports a similar subunit arrangement in human eIF2B as reported in the yeast *S. pombe*. Addition of ISRIB appears to stabilise the eIF2B decamer with less subunit dissociation observed, particularly into the $\beta\gamma\delta\epsilon$ and α_2 subcomplexes. Less dissociated α_2 is observed in the presence of ISRIB, suggesting that ISRIB not only stabilises two $\beta\gamma\delta\epsilon$ tetramers into $\beta_2\gamma_2\delta_2\epsilon_2$, but also stabilises α_2 binding to form the decameric $\alpha_2\beta_2\gamma_2\delta_2\epsilon_2$ complex, with the binding of α_2 itself stabilising two tetramers to the full decameric assembly $\alpha_2\beta_2\gamma_2\delta_2\epsilon_2$. In-solution disruption of eIF2B with acetic acid resulted in the formation of $\gamma_2\epsilon_2$ tetramers which may be explained by reports of an eIF2B conformational change in the absence of eIF2Bα, possibly changing the availability of eIF2 binding interfaces. Thus, there is a positive feedback mechanism on ISRIB binding; stabilisation of the eIF2B decamer and of α_2 binding which itself stabilises the decamer and prevents conformational change. These effects together would propagate a more significant impact of ISRIB binding and go some way to explaining its reported therapeutic effects in treating TBI in *in vivo* studies. The effect of ISRIB on such a conformational change could be explored by use of EM and gold nanoparticle labelling of eIF2Bγ in the presence and absence of ISRIB, as used previously on eIF2B [246].

ISRIB is a highly-promising candidate for pharmacological treatment for cognition defects arising from TBI and Alzheimer's disease. One area to explore is the influence of subunit α in promoting eIF2B GEF activity on both phosphorylated and non-phosphorylated eIF2. To this end, native MS could be used to compare the effects of ISRIB on eIF2B expressed with and without subunit α. Furthermore, the subsequent formation of eIF2-eIF2B complexes when incubated with phosphorylated and non-phosphorylated eIF2 could also be investigated with native

MS. Importantly, the effect of ISRIB on post-translationally-modified eIF2B should be explored in future studies as naturally-modified human eIF2B may show differences in reactivity and complex stability in response to PTMs such as phosphorylation, especially on account of the highly-conserved nature of the many PTMs identified on eIF2 and eIF2B from *S. cerevisiae* in this study. Native MS studies on the effects of ISRIB in human eIF2B expressed in cell culture from human cells such as HEK293 cells would contribute towards understanding the effects of ISRIB on post-translationally-modified human eIF2B.

4.7 Methods

4.7.1 Protein Purification

Purified His-tagged eIF2 from *S. cerevisiae* was provided by Dr Yuliya Gordiyenko and Jocky Kung (The University of Oxford, UK), purified as follows. His-tagged eIF2 from *S. cerevisiae* strain GP3511 (MATα leu2-3, -112 ura-52 inol gcn2Δ pep4::LEU2 sui2Δ HIS4-lacZ pAV1089 [SUI2 SUI3 6×HisGCD22 URA3]) was purified by Ni- and heparin-affinity chromatography as described [208]. Briefly, cells were grown in yeast extract peptone dextrose (YPD) media and lysed with a microfluidiser (M-110PS, Microfluidics) at 23,000 psi. The lysate was centrifuged and the supernatant incubated with Ni resin (GE Healthcare), with agitation on ice for 1 hour. Ni-bound eIF2 was eluted from the resin with a stepwise gradient of 15–100 mM EDTA. Eluted fractions were analysed by SDS-PAGE and eIF2-containing fractions pooled. The pooled fractions were loaded onto a heparin column and eIF2 eluted with a gradient up to 1M KCl over 20 column volumes. Eluted fractions were analysed by SDS-PAGE and eIF2-containing fractions pooled. Pooled fractions were dialysed overnight against storage buffer (38 mM HEPES (pH 7.4), 3.25 mM $MgCl_2$, 5 mM β-Mercaptoethanol (β-ME), and 135 mM KCl).

Purified eIF2B from *S. cerevisiae* was provided by Professor Graham Pavitt (The University of Manchester, UK), purified as follows. FLAG-tagged eIF2B from *S. cerevisiae* strain GP4109 (MATα leu2-3 leu2-112 ura3-52 ino1 gcd6Δ gcn2Δ::hisG ura3-52::HIS4-lacZ pAV1428[GCD6 GCD1-FLAG2-His6 URA3 2 μm] pAV1494[GCN3 GCD2 GCD7 LEU2 2 μm]) was purified by FLAG-affinity as described [257, 258]. Briefly, cells were lysed as described for eIF2, above. FLAG-tagged eIF2B complex was immobilised on anti-FLAG resin in a high-salt buffer (500 mM KCl). FLAG-tagged eIF2B was washed with 500 mM ammonium

acetate and eIF2B eluted with FLAG peptide (Sigma-Aldrich). The eluted fraction was further purified by SEC with a Superose 6 10/300 GL column (GE Healthcare) and concentrated with an Amicon Ultra concentrator (100 kDa MWCO) before flash-freezing and storage at -80 ^{o}C.

Purified human eIF2B expressed in *E. coli* was provided by Jordan Tsai and Professor Peter Walter (The University of California, San Francisco), purified as follows. Plasmids encoding the five subunits of human eIF2B (Appendix Figure B.1), including an N-terminal 6×His-tag and TEV cleavage site on subunit eIF2Bβ were transformed into One Shot BL21 Star (DE3) chemically-competent *E. coli* cells (Invitrogen). Cells were grown in LB supplemented with ampicillin, kanamycin, and chloramphenicol at 37 ^{o}C on an orbital shaker until an OD_{600} of 0.6 was reached. Cells were induced with 0.8 mM IPTG (Gold Biotechnology), then grown for 16 hours at 16 ^{o}C. Cells were harvested and lysed with the EmulsiFlex-C3 (Avestin) in lysis buffer (20 mM HEPES-KOH (pH 7.5), 250 mM KCl, 1 mM tris(2-carboxyethyl)phosphine (TCEP), 5 mM $MgCl_2$, 15 mM imidazole, and complete protease inhibitor cocktail (Roche)). The lysate was centrifuged at 30,000 × g for 20 minutes at 4 ^{o}C, before loading of the supernantant onto a 5 mL HisTrap HP column, equilibrated in binding buffer (20 mM HEPES-KOH (pH 7.5), 200 mM KCl, 1 mM TCEP, 5 mM $MgCl_2$, 15 mM imidazole). eIF2B was eluted with a linear gradient of binding buffer supplemented with 15 mM to 300 mM imidazole across 15 column volumes. The eIF2B-containing fraction was loaded onto a 20 mL MonoQ HR16/10 column (GE Healthcare), equilibrated in Buffer A (20 mM HEPES-KOH (pH 7.5), 200 mM KCl, 1 mM TCEP, 5 mM $MgCl_2$). eIF2B was eluted with a linear gradient of Buffer A supplemented with 200 mM to 500 mM KCl across 10 column volumes. The eIF2B-containing fraction was concentrated with an Amicon Ultra-15 concentrator (100 kDa MWCO) and loaded

onto a Superdex 200 10/300 GL column (GE Healthcare) equilibrated with Buffer A and eluted eIF2B fractions were stored at -80 ^{o}C prior to MS analysis.

Purified human eIF2Bα expressed in *E. coli* was provided by Jordan Tsai and Professor Peter Walter (The University of California, San Francisco), purified as follows. A plasmid encoding human eIF2Bα with an N-terminal 6×His-tag and TEV cleavage site was transformed into BL21 (DE3) *E. coli* cells and grown in LB supplemented with ampicillin at 37 ^{o}C on an orbital shaker, until an OD_{600} of 0.8 was reached. Cells were induced with 0.8 mM IPTG and grown for 16 hours at 20 ^{o}C. Cells were harvested and lysed in lysis buffer (20 mM HEPES-KOH (pH 7.5), 250 mM KCl, 1 mM TCEP, 5 mM $MgCl_2$, 20 mM imidazole, and protease inhibitor cocktail). The lysate was centrifuged at 30,000 × g for 20 minutes at 4 ^{o}C. The supernatant was loaded onto a 5 mL HisTrap HP column, equilibrated in a binding buffer (20 mM HEPES-KOH (pH 7.5), 30 mM KCl, 1 mM TCEP, 5 mM $MgCl_2$, 20 mM imidazole) and eluted with linear gradient of binding buffer supplemented with 20 mM to 300 mM imidazole across 15 column volumes. The eluted eIF2Bα-containing fractions were loaded on a MonoQ HR 16/10, equilibrated in buffer (20 mM HEPES-KOH (pH 7.5), 30 mM KCl, 1 mM TCEP, 5 mM $MgCl_2$). The collected flow-through contained eIF2Bα and was loaded on a MonoS HR 10/10 (GE Healthcare), equilibrated in buffer (20 mM HEPES-KOH (pH 7.5), 30 mM KCl, 1 mM TCEP, 5 mM $MgCl_2$). The collected flow-through again contained eIF2Bα and this was incubated with TEV protease (50 μg per litre culture) for 16 hours at 4 ^{o}C. The TEV protease reaction was loaded on a 5 mL HisTrap HP column, with the flow-through containing cleaved eIF2Bα. The protein was concentrated with an Amicon Ultra-15 concentrator (30 kDa MWCO) and loaded onto a Superdex 75 10/300 GL (GE Healthcare) column equilibrated in buffer (20 mM HEPES-KOH (pH 7.5), 200 mM KCl, 1 mM TCEP, 5 mM $MgCl_2$,

5 % glycerol). Eluted protein was stored at -80 ^{o}C prior to MS analysis.

4.7.2 Identification of Post-Translational Modifications

Purified eIF2 and eIF2B from *S. cerevisiae* were separated by SDS-PAGE and stained with Coomasie using InstantBlue (Invitrogen). Gel bands were excised and digested in-gel with trypsin according to Shevchenko *et al.* [182]. Phosphopeptides were enriched for phosphosite identification analysis with TiO_2 affinity columns according to Thingholm *et al.* [92].

For phosphosite identification, tryptic peptides were separated by nano-flow RPLC with a DionexUltiMate 3000 RSLC nano system (Thermo Fisher Scientific). Peptides were loaded onto a trap column (HPLC column Acclaim PepMap 100, C18, 100 μm inner diameter particle size 5 μm; Thermo Fisher Scientific) and separated on an analytical capillary column (50 cm, HPLC column Acclaim PepMap 100, C18, 75 μm inner diameter particle size 3 μm; Thermo Fisher Scientific), with a gradient of 5–80 % (v/v) mobile phase B over 74 minutes (mobile phase A: 0.1 % (v/v) FA; mobile phase B: 80 % (v/v) ACN/0.1 % (v/v) FA) with a flow rate of 300 nL min^{-1}. The LC was coupled to an LTQ Orbitrap XL Hybrid Mass Spectrometer (Thermo Fisher Scientific). MS instrument parameters were set: spray voltage 1.8 kV; capillary temperature 180 ^{o}C; normalised collision energy 35 % at an activation q = 0.25; and an activation time of 30 ms. The LTQ Orbitrap XL was operated in data-dependent mode. Survey full-scan spectra were acquired in the Orbitrap (300–2,000 m/z) with a resolution of 30,000 at 400 m/z and an AGC target at 1×10^{6}. For MS/MS, The top five most intense ions were selected for CID fragmentation in the linear ion trap at an AGC target of 30,000. Dynamic exclusion of 30 s for previously-selected ions was applied. Ions with an unknown charge state or a charge state of 1+ were excluded. The Orbitrap mass analyser was internally

calibrated using the lock mass option (445.120025 *m/z* [183]).

For acetylation site identification, tryptic peptide samples that were not enriched for phosphopeptides were separated by nano-flow RPLC (EASY nLC 1000, Thermo Fisher Scientific) with a mobile phase A, 0.1 % (v/v) FA/ 5 % (v/v) DMSO and a mobile phase B, 100 % (v/v) ACN/ 0.1 % (v/v) FA/ 5 % (v/v) DMSO). Peptides were loaded onto a trap column (5 mm, PepMap RSLC, C18, 300 μm inner diameter particle size 5μm; Thermo Fisher Scientific) and separated on an analytical capillary column (50 cm, PepMap RSLC, EASY-spray column, C18, 75 μm inner diameter particle size 3 μm; Thermo Fisher Scientific), with a gradient of 7-30 % (v/v) mobile phase B over 30 minutes at a flow rate of 200 nL min^{-1}. The column temperature was 40 ^{o}C. The LC was coupled to a Q Exactive Orbitrap Mass Spectrometer (Thermo Fisher Scientific). MS instrument parameters were set as follows: spray voltage 2.1 kV and capillary temperature 320 ^{o}C. The Q Exactive Orbitrap was operated in data-dependent mode. Survey full-scan MS spectra were acquired in the Orbitrap (*m/z* 350–1,500) with a resolution of 70,000 and an AGC target at 3×10^6. The top ten most intense ions were selected for HCD fragmentation at an AGC target of 50,000.

Data were converted from raw to mgf format using pXtract conversion software (http://www.pfindstudio.com/software/pXtract/index.html [259]). Data were searched against NCBI non-redundant database, with the *S. cerevisiae* taxonomy filter applied, using Mascot v2.4.1 search engine (Matrix Science). Search parameters were as follows: precursor mass accuracy 10 ppm (LTQ XL data) or 7 ppm (Q Exactive data); fragment ion mass accuracy 0.6 Da (LTQ XL data) or 0.01 Da (Q Exactive data); enzyme trypsin; missed cleavage sites up to 2; variable modifications of oxidation on methionine; fixed modification of

carbamidomethylation of cysteine. For PTM identification, variable modifications of phosphorylation on serine, threonine, and tyrosine; or acetylation on lysine or the protein N-terminus were allowed. Identified PTMs were compared to PTMs from three databases; the UniProtKB database (http://www.uniprot.org [260]), the PHOSIDA database (http://www.phosida.com [261]), and the PTM*func* database (http://ptmfunc.com [170]).

4.7.3 Homology Modelling

Homology models were generated using the Phyre2 web server [237] in intensive mode. Homology models for eIF2α, β, and γ (template PDB IDs: 1Q8K [262], 2DCU [263], 5K0Y [264], respectively) were aligned to the structure of archaeal initiation factor 2 from *S. solfataricus* (PDB ID: 3CW2 [204]) using Pymol software. Homology models of eIF2Bα, β, γ, δ, and ϵ (template PDB IDs: 5B04 and 1PAQ [248]) were aligned with the structure of eIF2B from *S. pombe* (PDB ID: 5B04 [248, 265]).

4.7.4 Chemical Cross-linking Mass Spectrometry

Purified yeast eIF2B was cross-linked with BS3-d0/d4 at 25 ^{o}C for one hour with 400 rpm shaking in a Thermomixer (Eppendorf). The cross-linking reaction was quenched and divided. Half was separated by SDS-PAGE and digested in-gel with trypsin. The other half was subjected to ethanol precipitation before in-solution digestion with trypsin. The resulting peptides were analysed by LC-MS/MS as described previously. Data analysis was performed with pLink software [259] against a user-defined database of eIF2B subunits α, β, γ, δ, and ϵ with the following search parameters: enzyme trypsin with up to three missed cleavage sites; cross-linkers BS3-d0 and BS3-d4; peptide tolerance 10 ppm. All spectra were validated manually by visual inspection. Cross-links were visualised with xVis software (https://xvis.genzentrum.lmu.de [247]).

4.7.5 Native Mass Spectrometry

Native tandem mass spectra of human eIF2B, isolated human eIF2Bα, and direct addition of ISRIB to human eIF2B (Figures 4.8, 4.9, 4.11, and 4.12) were acquired on a QToF II Mass Spectrometer (Micromass). Instrument parameters were optimised for each spectrum with typical values as follows: capillary voltage 1.6 kV; cone voltage 20 V; extractor 5–15 V; collision voltage 25–200 V; backing pressure 4–5 mbar; MCP 1800 V; collision cell gas 0.3 MPa. Spectra were processed with MassLynx v4.1 and calibrated externally with cesium iodide (10 mg mL^{-1}) Native mass spectra of in-solution disruption of eIF2B with methanol and acetic acid, and eIF2B after extended incubation with ISRIB (Figures 4.10, 4.14, & 4.13) were acquired on a Q Exactive Plus Mass Spectrometer with EMR (Thermo Fisher Scientific) and modifications for UHMR as described [44]. Instrument parameters were optimised for each spectrum with typical values as follows: capillary temperature 30–100 ^{o}C; capillary spray voltage 1.2 kV; desolvation voltage -100 V; HCD voltage 250 V; C-trap entrance lens 2; inject flatapole voltage 8 V; interflatapole voltage 4 V; bent flatapole voltage 4 V; trapping gas parameter 4–8; noise threshold 3. Protein was analysed at 1 μM unless stated otherwise. Solution conditions for native MS of eIF2B were as follows. Human eIF2B in 500 mM ammonium acetate (pH 7.5) buffer was disrupted with 15 % (v/v) methanol in-solution prior to MS. Human eIF2B was incubated in binding buffer (20 mM HEPES-KOH (pH 7.5), 200 mM KCl, 1 mM TCEP, 5 mM $MgCl_2$) with equimolar ISRIB (in DMSO, final DMSO concentration 1 % (v/v)) on ice for 3 hours, or overnight in the case of the extended incubation time. Sample buffer was exchanged to 500 mM ammonium acetate (pH 7.5) with Bio-spin gel filtration columns (Bio-Rad) twice, prior to MS analysis. For acetic acid disruption, human eIF2B in 500 mM ammonium acetate solution was incubated with 0.01 % (v/v) acetic acid for 30 minutes prior to MS analysis.

Chapter 5

Regulation of the Rotary ATP Synthases

5.1 Introduction

Adenosine triphosphate (ATP) is an energy storage molecule in many organisms. Its hydrolysis to adenosine diphosphate (ADP) and inorganic phosphate (P_i) liberates the free energy stored in its high-energy phosphate bond [267]. Cells are able to couple the energetically-favourable ATP hydrolysis reaction to accelerate other reactions. The synthesis of ATP however requires energy. Consequently, cells possess enzymes that are able to convert the energy from photosynthesis and respiration to generate a proton motive force (PMF). This PMF can then be used for ATP synthesis by the ATP synthase enzyme [268].

Parts of this chapter have been published in: Carla Schmidt, Victoria Beilsten-Edmands, Shabaz Mohammed, and Carol V. Robinson, *Scientific Reports*, 2017 [266]. Gel bands of purified V-type ATPase samples from *S. cerevisiae* were provided by Stephanie Bueler, Thamiya Vasanthakumar, and Professor John Rubinstein (The University of Toronto, Canada). Professor Carla Schmidt (The Martin Luther University of Halle-Wittenberg) purified cATP synthase from *S. oleracea* and performed cATP synthase comparative cross-linking experiments. Purified mitochondrial F-type ATP synthase from *B. taurus* was provided by Professor Christoph Gerle (The University of Osaka, Japan). Lipidomics experiments were performed by Dr Di Wu and Dr Dror Chorev (The University of Oxford, UK). All data analysis presented is my own.

5.2 ATP Synthases

The rotary ATP synthases are conserved in the membranes of bacteria, chloroplasts, and mitochondria. They harness the PMF in order to synthesise ATP but also work in the reverse direction to acidify organelles, powered by ATP hydrolysis [268]. The rotary ATP synthase family has three members; the F-type, the V-type, and the A-type ATP synthases; the first two types will be investigated in this chapter. The F-type ATP synthases generally work to synthesise ATP, while the V-type ATPases are proton pumps powered by ATP hydrolysis. Despite these different roles, the structure and functional mechanisms of the rotary ATP synthases are similar [269].

Their structure can be divided into two parts; the Oligomycin-binding Fraction (F_o/V_o) and Fraction 1 (F_1/V_1) (Figure 5.1). The F_o/V_o consists of the membrane-embedded rotor and proton pump and F_1/V_1 consists of the soluble hexameric head, housing the catalytic sites, which is attached to the F_o/V_o via central and peripheral stalks [270]. Although the structures deviate in part from one another, many of the individual subunits are homologous, but for historical reasons are named inconsistently [269]. Despite these differences, the overall mechanisms of ATP synthesis/hydrolysis and proton translation are thought to be conserved.

In the presence of sufficient PMF, proton translocation induces rotation of the rotor ring. The rotor ring is comprised of hydrophobic c subunits, with a stoichiometry dependent on species, and is located adjacent to the a subunit which together form the proton pathway. This pathway is thought to take the form of two half-channels [271, 272], the exit pores of which have recently been visualised in yeast F-type ATP synthase by cryo-EM [273]. Protons enter the first half-channel on the lower side of subunit a (Figure 5.2), on the p-side (protochemically-positive side) of the membrane, and bind to a negatively-charged carboxylate on a conserved

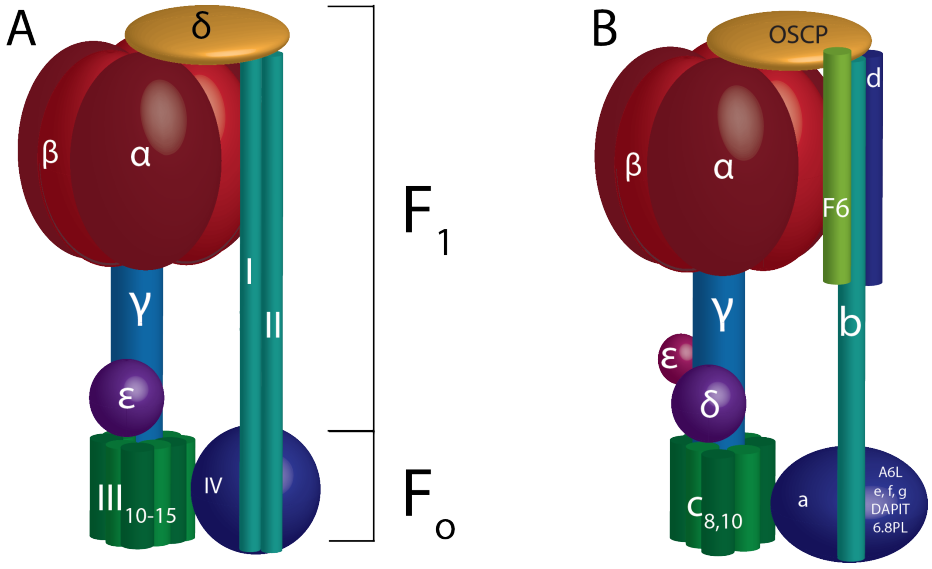


Figure 5.1: Cartoon representation of the structure of F-type ATP synthase in (A) bacteria and chloroplasts and in (B) mitochondria. The soluble F_1 catalytic domain contains an $\alpha_3\beta_3$ head (dark red and red) which is the site of ATP synthesis or hydrolysis. The central stalk comprises the γ subunit (blue) which connects F_1 to the membrane-embedded F_o domain. In addition, a second subunit termed ϵ (purple) in bacteria and chloroplasts, and termed δ (purple) in mitochondria, also connects the central stalk to the F_o domain. In mitochondria an additional subunit, ϵ (magenta) binds the central stalk and contacts the F_o domain. The proton pathway is located in the F_o domain between the *c*-ring (dark green) and subunit *a* (dark blue). Additional membrane-embedded supernumerary subunits are found in the mitochondrial enzyme, including *e*, *f*, *g*, *A6L*, *DAPIT*, and *6.8PL* subunits. In bacteria and chloroplasts the peripheral stalk is comprised of two subunits; *b* and *b'* (teal) and are coupled to F_1 by subunit δ (orange). In mitochondria, the peripheral stalk is comprised of three subunits; *b*, *d*, and F_6 (teal, dark blue, and light green respectively), while coupling of F_1 to the peripheral stalk is performed by oligomycin sensitivity-conferring protein (*OSCP*) (orange).

acidic glutamate or aspartate residue at the C-terminus of the *c* subunit, which is isolated from the surrounding lipid membrane. The neutralised carboxylic acid residue moves via Brownian motion towards the more hydrophobic environment of the lipid bilayer. This in turn exposes another negatively-charged carboxylate to the half-channel for proton binding. Successive proton-binding and *c*-ring rotation occurs until the proton reaches the upper half-channel on subunit *a*, which is exposed to the upper side of the membrane. At this point, the carboxylic acid

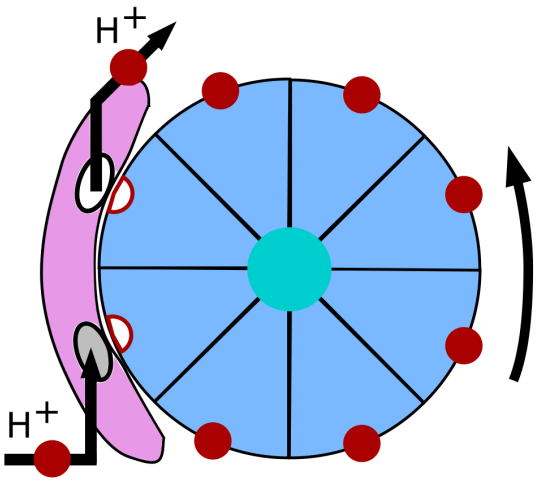


Figure 5.2: Schematic of the proton pathway during ATP synthesis. Subunit a (pink) and the c-ring (blue) together form the proton pathway via two half-channels. Protons (red circles) enter at the base of ATP synthase, on the p-side of the membrane, and bind to negatively-charged C-terminal aspartate or glutamate residues on the c-ring. The newly-neutralised residue rotates into the hydrophobic lipid membrane, revealing another negatively-charged residue for proton binding. When a neutralised aspartic or glutamic acid residue arrives at the upper subunit a half-channel, the residue is deprotonated and the proton released. This rotation of the c-ring and resulting rotation of the central stalk drives ATP synthesis.

residue is exposed to the local environment on the n-side (protochemically-negative side) of the membrane and deprotonated, thus leaving an acidic carboxylate available for the next round of proton binding. When the PMF is sufficient, the c-ring rotation induced by this proton binding and release cycle drives ATP synthesis. Alternatively, under different conditions such as in the absence of such a proton gradient, or in the case of V-type ATPases, the rotor reverses direction and ATP hydrolysis drives the translocation of protons [268].

As the rotor turns, the attached central stalk rotates with it. The central stalk consists of subunit γ which protrudes through the centre of the soluble hexameric head and extends downwards to form the base of the central stalk, contacting the membrane-embedded F_o/V_o domain. The soluble hexameric head contains α and β

subunits (termed B and A in the V-type ATPases) in an $\alpha_3\beta_3$ stoichiometry, with the catalytic nucleotide-binding sites located on the β subunits near the α–β interfaces. The γ subunit is asymmetrical and as it rotates, this asymmetry induces conformational changes in the $\alpha_3\beta_3$ head [268].

Described as the binding change mechanism [274], the conformational changes induced by 360^o rotation of subunit γ inside the $\alpha_3\beta_3$ head take each catalytic site through three distinct conformational states (Figure 5.3) and therefore induce ATP synthesis. Rotation of γ takes place in steps of 120^o, divided into sub-steps of 40^o and 80^o [275]. The conformational states of the β-catalytic sites are termed 'open', 'loose', and 'tight' and refer to the nucleotide-binding affinity of the site in each conformation [270, 274]. The catalytic ATP synthesis cycle starts with a nucleotide-free catalytic site in the 'open' conformation binding ADP and P_i.

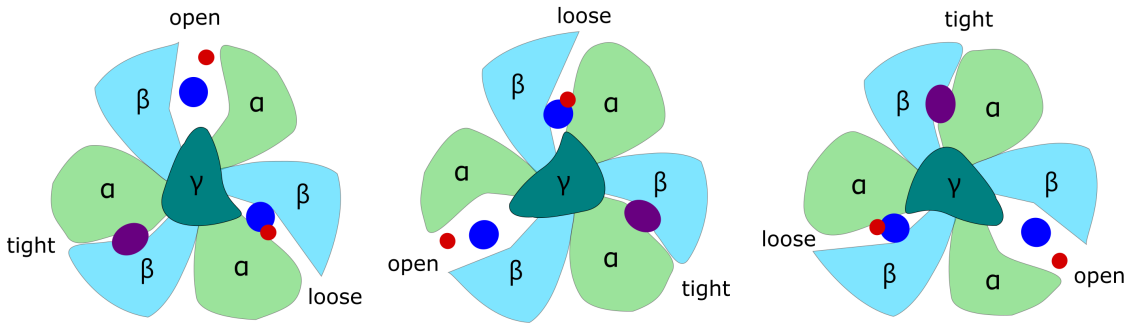


Figure 5.3: The binding change mechanism of ATP synthesis. Nucleotide binding sites are located at the interfaces of α and β subunits, alternating between α non-catalytic and β-catalytic sites. At the β-catalytic sites, each site is in one of three conformations; loose, open, or tight. Rotation of the central γ stalk by 360^o induces conformational changes taking each site sequentially through the three conformations. In the open conformation, ADP (blue circles) and P_i (red circles) bind to the catalytic site. Rotation of γ by 120^o induces a change to the loose conformation in which ADP and P_i are pushed closer together. Further γ rotation by 120^o induces a change to the tight conformation with ADP and P_i reacting to form ATP (purple circles). Further γ rotation induces a change to the open conformation, releasing newly-synthesised ATP and allowing ADP and P_i to bind for another round of ATP synthesis.

Rotation of the central γ stalk by 120^o induces a conformational change to the 'loose' conformation, with ADP and P_i binding to the site. ADP and P_i are converted to ATP as γ rotates by 120^o transferring the site to the 'tight' conformation, with the newly-synthesised ATP molecule bound with high affinity. Finally, the 'open' conformation is restored as γ completes a 360^o rotation, and the newly-synthesised ATP is released. As each of the three β subunits cycles through the three conformations during the 360^o rotation of γ, three ATP molecules are synthesised [268].

Although the general structure and mechanisms for proton translocation and ATP synthesis/hydrolysis are conserved between the different rotary ATPases, there are numerous structural differences which have implications on their functional and regulatory mechanisms.

5.2.1 The Structure of the F-type ATP Synthases

The F-type ATP synthase is perhaps the simplest member of the rotary ATPase family, in terms of structure (Figure 5.1). Like other ATP synthases, the F_1 domain is comprised of two alternating subunits; α and β with a stoichiometry of $\alpha_3\beta_3$. The F_1 domain, known as the catalytic domain, binds nucleotides on either α or β subunits. The nucleotide-binding sites on the α subunits are non-catalytic and are not known to have a direct role in ATP synthesis or hydrolysis [276]. On the contrary, the nucleotide-binding sites on the β subunits are catalytic and are responsible for ATP synthesis or hydrolysis, depending on the direction of the motor (see above).

Binding of the central γ stalk to F_o is aided by another subunit called ϵ in bacteria and chloroplasts, while the mitochondrial homologue is called δ and is further aided

by a second subunit ϵ in its attachment to the rotor ring of F_o [269]. The rotor ring of F_o is comprised of multiple copies of a hydrophobic subunit *c* (termed subunit *III* in chloroplasts), with a stoichiometry that varies by species. The membrane-bound *c*-ring is located adjacent to subunit *a* (subunit *IV* in chloroplasts). The proton pathway (discussed above) is located between these two subunits [269].

The F_o domain is connected to the F_1 domain via a second peripheral stalk, the structure of which varies between mitochondrial enzymes and those in bacteria and chloroplasts (Figure 5.1). In bacteria and chloroplasts, subunits *b* and *b'* (*II* and *I* in chloroplasts) extend from the membrane to the top of the F_1 head, with subunit δ coupling the peripheral stalk to the $\alpha_3\beta_3$ head [269]. In mitochondria, subunits *b*, *d*, and F_6 comprise the main peripheral stalk with the mitochondrial δ homologue oligomycin sensitivity-conferring protein (*OSCP*) ensuring strong coupling between the head and peripheral stalk. Overall, the peripheral stalk takes the role of a static stator, whose role it is to ensure that the $\alpha_3\beta_3$ head does not rotate with subunit γ.

In green plants, F-type ATP synthase is located on the chloroplast thylakoid membranes and utilises the PMF established from oxidative phosphorylation to synthesise ATP [269]. The stoichiometry of the ring is III_{14} and contains a stable lipid 'plug', predominantly of sulfolipids, which binds stoichiometrically to the *III* subunits [99]. The lipid plug is thought to reduce the central orifice of the rotor ring to better accommodate subunit γ. Numerous phosphorylation sites have been identified in the chloroplastic F-type ATP synthase from *Spinacia oleracea* (cATPase) [99, 277] and phosphorylation has been shown to have an important role in stabilising cATP synthase subunit interfaces and nucleotide binding [99].

Mammalian mitochondrial F-type ATP synthase, isolated from bovine heart, has

been well-studied with a combination of crystal structures [279] and cryo-EM density maps [278] giving a good picture of the subunit organisation and structures of many of the core subunits (Figure 5.4), including different conformational states of the $\alpha_3\beta_3$ head. Mammalian mitochondrial F_oF_1 contains additional subunits in the membrane-embedded region, near to subunit *a* and are termed the supernumerary subunits. The roles of the supernumerary subunits, which include subunits *e*, *f*, *g*, diabetes-associated protein in insulin-sensitive tissues (*DAPIT*), and 6.8 kDa proteolipid (*6.8PL*) are not fully understood with little known about their structure or location within the F-type ATP synthase [280, 281], although it is

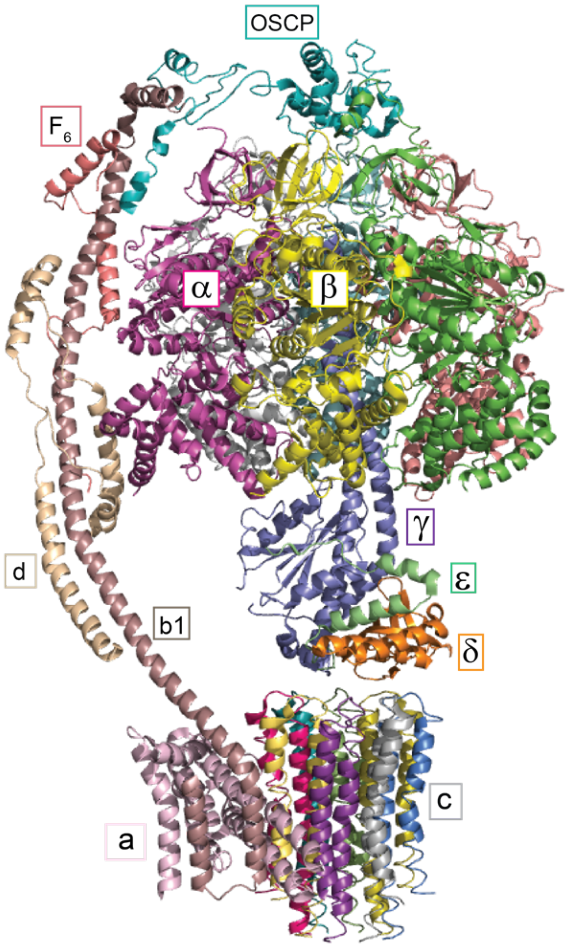


Figure 5.4: Partial structure of the bovine ATP synthase, excluding the supernumerary membrane subunits (PDB ID: 5ARA [278]).

known that their association with the protein complex is dependent on the presence of phospholipids during protein purification [282]. Subunits *e*, *f*, and *g* are thought to contribute to the membrane-bending kink imparted by F_oF_1, crucial for mitochondrial cristae formation [283–285]. Cryo-EM studies have identified an unassigned feature on the F_o domain, termed the intermembrane space domain (IMD) [286], however the scarcity of structural data on the F_o domain has impeded conclusive assignment.

5.2.2 The Structure of the V-type ATPases

Unlike their ATP-synthesising counterparts, V-type ATPases primarily serve as proton pumps to acidify organelles, and in some cases assist in protein transport [287]. Significant amounts of energy are expended in maintaining appropriate pH balance, as effective control of pH is an essential part of homeostasis, with the malfunctioning of pH-stabilising machinery such as V-type ATPases leading to multiple diseases, for instance renal tubular acidosis [288] and osteopetrosis [289]. Their general structure is similar to that of F-type ATP synthases however, they are more complex and contain more protein subunits (Figure 5.5). The V_1 domain contains eight subunits *A*-*H* [290]. The catalytic hexameric head is conserved, comprising alternating *A* and *B* subunits (β and α in F-type ATP synthases) in an A_3B_3 stoichiometry, with the catalytic nucleotide-binding sites located on the three *A* subunits. The central stalk connects the catalytic head to the V_o domain and consists of the asymmetric subunits *D* and *F* (γ and δ in mitochondrial F-type ATP synthase). While the F-type ATP synthases contain a single peripheral stalk, the V-type ATPases have three stalks, each composed of subunit *E* and *G* (three copies each per enzyme), along with subunits *H* and *C* which have regulatory roles [287].

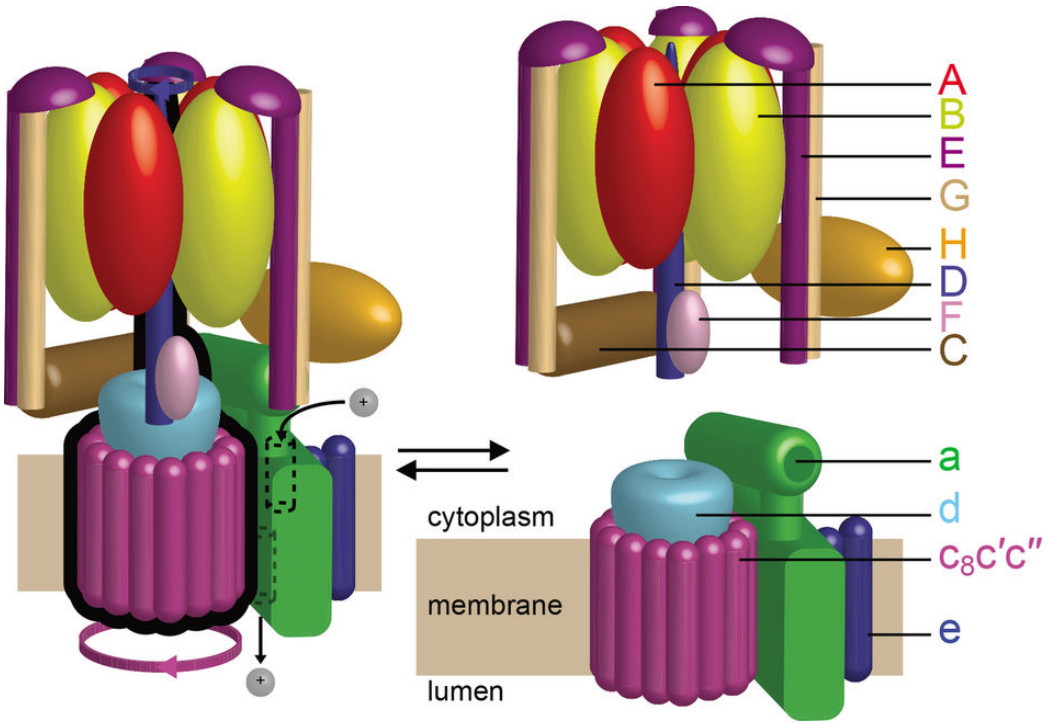


Figure 5.5: The structure of V-type ATPases. Left: The soluble F_1 catalytic domain contains subunits *A* (red) and *B* (yellow) in an A_3B_3 stoichiometry. The central stalk connects F_1 to F_o and is comprised of subunit *D* (blue), *F* (light pink) and *d* (cyan). The *c*-ring is comprised of three subunits; *c*, *c'* and *c"* (magenta). The F_o contains in addition further membrane subunits *a* (green), and *e* (navy). The V-type ATPase contains three peripheral stalks, made of subunits *C*, *E*, *G*, and *H* (brown, purple, beige, and orange, respectively). Right: Autoinhibition occurs when V_1 and V_o dissociate under conditions of nutrient starvation, thus halting ATP hydrolysis. Figure from Mazhab-Jafari *et al.* [14].

The V_o domain contains six subunits; *a*, *d*, *e*, *c*, *c'*, and *c"*. Subunit *d* couples the central stalk to the *c*-ring, but unlike in the F-type ATP synthases where the rotor ring is homo-oligomeric, the V-type *c*-ring is hetero-oligomeric and contains three different proteolipid subunits; *c*, *c'*, and *c"*. The stoichiometry of the *c*-ring has been an area of uncertainty, however high-resolution cryo-EM data of the yeast V_o has shown the *c*-ring to be arranged with a stoichiometry of $c_8c'_1c''_1$ [291]. Each ring subunit contains the conserved acidic residue required for proton translocation, with the adjacent membrane-embedded subunit *a* forming the rest of the proton pathway, with its two half-channels each open to opposite sides of the membrane.

While most of the subunits of V_oV_1 in yeast exist as a single isoform, subunit *a* is encoded by two genes (VPH1 and STV1) and therefore exists as two isoforms (Vph1p and Stv1p) [287]. The two isoforms are important in targeting of the ATP synthase to different organelles, with Vph1p-containing complexes being found primarily on vacuolar membranes whereas Stv1p-containing enzymes are predominantly found on the membranes of the Golgi apparatus [292, 293]. The N-terminal region encodes the signalling for this differential targeting, with higher levels of sequence identity found in the C-terminal region. Further differences in the two isoforms are unclear, although discrepancies in the coupling efficiency and extent of dissociation have been reported and proposed to explain the different levels of acidification in vacuoles and the Golgi apparatus [293, 294].

Additional small membrane subunits have been identified, such as the hydrophobic subunits *e* and *f*, however their functions are unclear [14]. Some of the best V_oV_1 structures were obtained by cryo-EM from the yeast *S. cerevisiae* enzyme [14, 291, 295]. These analyses have revealed invaluable information on the structure and organisation of the protein subunits, particularly in the V_o domain, which has been challenging to study due to its hydrophobic nature.

5.2.3 The Regulation of ATP Synthases

Multiple mechanisms of regulation have been identified across the various types and species of ATP synthases to regulate their activity in response to environmental factors such as low PMF [296], glucose starvation [297], and darkness [298]. Here, I will focus on three inhibitory mechanisms which are pertinent to the work in this chapter; the autoinhibitory dissociation of V_oV_1 in yeast, the conformational change of ϵ in F_oF_1 in chloroplasts, and the binding of IF_1 to F_oF_1 in mitochondria.

5.2.3.1 V_oV_1 Autoinhibition

The V_oV_1 complex exhibits a mode of autoinhibition via reversible dissociation of V_oV_1 into V_o and V_1 (Figure 5.5B) in response to various stimuli including glucose starvation in yeast [297]. Uncoupled V_o does not pump protons as V_1 is not present to power the *c*-ring rotation, and the V_1 does not hydrolyse ATP due to silencing by subunit *H* [299]. Thereby ATP hydrolysis is stopped and wastage of ATP is prevented when glucose is scarce. Reassembly of V_oV_1 requires interaction with the protein complex RAVE (regulator of the ATPase of vacuolar and endosomal membranes) [300] which not only assists in reassembly of starvation-dissociated V_oV_1 but also in normal V_oV_1 complex assembly [299]. Conversely, dissociation does not require interaction with RAVE or synthesis of any other new protein [297]. How this dissociation is induced is to date unknown, although various mechanisms have been proposed including modulation of protein-protein interactions between the V_o and V_1 domains.

5.2.3.2 A Conformational Change of Subunit ϵ

In bacteria and chloroplasts, subunit ϵ regulates ATP synthesis in response to the cellular concentration of ATP. The α-helical C-terminus of subunit ϵ binds to ATP and adopts a folded conformation. When ATP is absent however, the helices adopt an extended conformation, interact with the $\alpha_3\beta_3$ head and thus prevent ATP hydrolysis [301–303]. This results in a sixteen-fold decrease in ATP hydrolysis [304] in spinach chloroplasts. It was further found that replacement of basic C-terminal Arg residues in chloroplast ϵ significantly decreases the inhibitory effect [305] as did replacement of negatively-charged glutamate in the DELSEED motif of subunit β from *E. coli*, indicating that electrostatic interactions are important in maintaining the inhibitory effect [306].

5.2.3.3 The IF_1 Inhibitor

Although there is no evidence that ϵ-mediated inhibition occurs in mitochondrial F_oF_1, parallels have been drawn with the IF_1 inhibitor protein [308] characterised in yeast [309], plants [310], and mammals [311, 312]. In bovine, IF_1 is a basic 84 amino acid protein [312] that inhibits F_oF_1 in mitochondria. The sequence is highly conserved, especially at residues 14-47 which form the F_1 binding site. The activity of IF_1 is determined by its pH-dependent oligomerisation state. At a pH < 6.7, IF_1 forms an active inhibitory dimer [313] able to bind to up to two F_oF_1 complexes [314]. At pH > 6.7, dimerisation to tetrameric IF_1 occludes the F_oF_1 binding site resulting in an inactive non-inhibitory form. The pH-sensitivity of oligomerisation is thought to depend on the charge state of the His residue 49 [313]. Instead of the dimer:tetramer equilibrium as observed in bovine IF_1 (Figure 5.6), characterisation of IF_1 from *S. cerevisiae* has revealed a monomer:dimer equilibrium in which monomeric IF_1 forms the active protein [315].

A crystal structure of bovine IF_1 in complex with the ATP synthase F_1 domain was

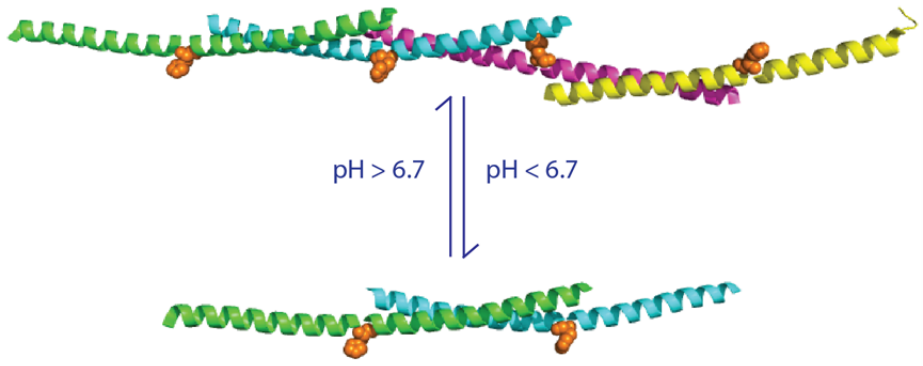


Figure 5.6: The oligomerisation state of IF_1 is dependent on pH. The crystal structure of tetrameric IF_1, determined at 2.2 Å is shown (PDB ID: 1GMJ [307]). At pH > 6.7, an inactive tetramer is formed (top), while at pH < 6.7, an active dimeric inhibitor forms. Histidine residue 49, thought to be responsible for the pH-dependent oligomerisation, is highlighted in orange.

solved by Cabezón *et al.* in 2003 [316]. In this structure, crystals were preformed in ATP and grown with both AMP-PNP and ADP present. IF_1 contacts the central γ stalk and resides at the interface between α and β subunits (Figure 5.7). IF_1 penetrates into the $\alpha_3\beta_3$ core and also contacts inner residues of α and β. By comparison to a reference structure without IF_1 present, Cabezón *et al.* found that ATP or AMP-PNP resides in a β-catalytic site in the IF_1-bound complex whereas in the absence of IF_1 this site is occupied by ADP. This was attributed to the inhibition of hydrolysis by IF_1 via prevention of the necessary conformational changes of subunits α and β.

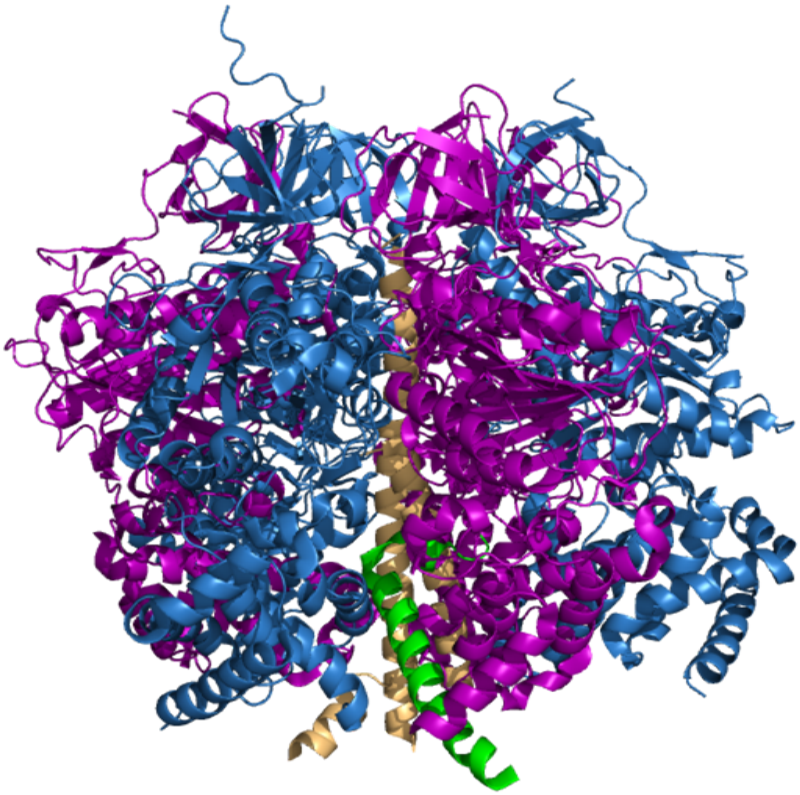


Figure 5.7: The structure of bovine IF_1 in complex with F_1, determined by X-ray diffraction at 2.8 Å resolution (PDB ID: 1OHH [316]). Three copies of subunit α (blue), three copies of subunit β (purple), subunit γ (gold) and IF_1 (green) are shown. IF_1 contacts γ subunit and neighbouring α and β subunits.

5.3 Research Aims

In spite of the significant progress in elucidating the structures of ATP synthases in the past few years, many uncertainties remain to be understood about these molecular machines. By studying both V-type and F-type ATP synthases from three different species I aim to exploit MS-based methods to expand our knowledge of various aspects of ATP synthases.

The first ATP synthase investigated is the V-type ATPase from the yeast *S. cerevisiae*. While the level of structural and functional detail known about the yeast V-type ATPase has increased significantly over the past few years, many questions remain to be answered. I aim to address two of these in this chapter. The structural insights gained from a recent study of the vacuolar V_o from *S. cerevisiae* combining cryo-EM with proteomics analysis [14], revealed the presence and location of subunit *e* and the newly-identified subunit *f*, both in contact with subunit *a* (Vph1p) (Figure 5.8). An isoform of subunit *a* exists (Stv1p), which is found in V_oV_1 in the Golgi apparatus of *S. cerevisiae*, and Stv1p has only 49 % sequence identity with Vph1p. It is unknown if subunits *e* and *f* are present in the Golgi-form of V_oV_1 and if they are the same isoforms as those found in the vacuolar enzyme. Therefore I will first address the presence and form of these supernumerary subunits in the Golgi-form of V_oV_1.

A second open question regarding the yeast V_oV_1 concerns its autoinhibitory dissociation to V_o and V_1 and the underlying mechanism. Most likely phosphorylation plays a role as it often does in regulation of various protein complexes and furthermore considering the location of subunits within the V_oV_1 complex, subunit *a* is a likely target. I will explore the possible role for phosphorylation in acting as a molecular switch for autoinhibition.

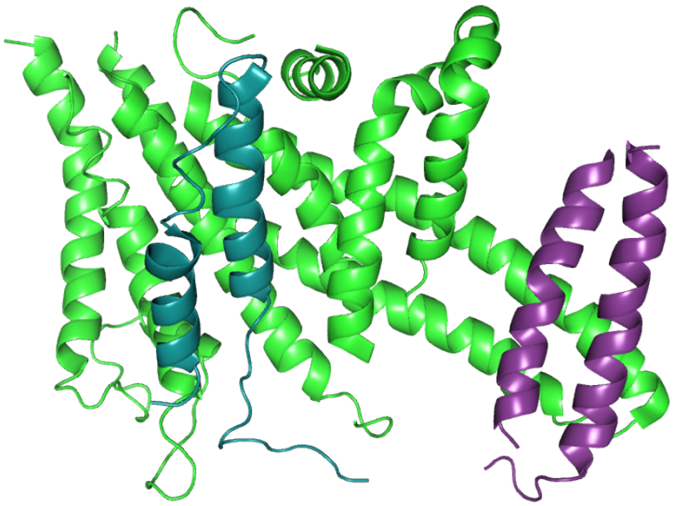


Figure 5.8: Subunit organisation of subunits *a* (Vph1p), *e*, and *f* from cryo-EM analysis of the V_o domain (PDB ID: 5TJ5 [14]). Subunits *e* (cyan) and *f* (purple) contact subunit *a* (green).

Next, and continuing on the topic of PTM regulation of ATP synthases, I will focus on F_oF_1 found in the chloroplasts of *S. oleracea*. PTMs have been shown previously [99] to have an important structural role, yet our understanding is currently incomplete. In addition, while phosphorylation has been studied in some detail, acetylation has not in spite of its increasingly-recognised role in protein regulation [317], both alone [318] and in conjunction with phosphorylation [169]. The role of PTMs, particularly acetylation together with phosphorylation, will be explored and the details of structural changes examined, by use of chemical cross-linking.

Finally, having considered the role of PTMs in eukaryotic V- and F-type ATP synthases, I will study the mammalian F-type ATP synthase from the mitochondria of *Bos taurus*. As PTMs often play a regulatory role, I will first assess the PTM status of F_oF_1 through a combination of denaturing MS and proteomics. In order to understand the role of PTMs in ATP synthase regulation, it is necessary to expand

our knowledge on its structure and subunit organisation, with information on the membrane subunits being the most elusive. I will analyse F_oF_1 with cross-linking and native MS, with the aim of uncovering the placement and organisation of these membrane-situated subunits. I will then consider other ways of regulating F_oF_1, primarily looking at the known inhibitor IF_1.

5.4 The Yeast V-type ATPase

5.4.1 Results and Discussion

5.4.1.1 The Identification of the V-type ATPase Subunits *e* and *f*

Given the recent identification of subunits e and f in association with the vacuolar V-type ATPase from *S. cerevisiae* [14], the presence or absence of subunits e and f in the Golgi-form of V_oV_1 was investigated. Excised lower molecular weight gel bands from SDS-PAGE of purified Golgi-form V_oV_1 were provided by Thamiya Vasanthakumar and Professor John Rubinstein (The University of Toronto, Canada). Due to the highly hydrophobic nature of these subunits which often results in low sequence coverage with tryptic peptides [319], the gel bands were digested in-gel with trypsin, chymotrypsin or a combination of both enzymes. After LC-MS/MS analysis and database search, subunits e and f were both identified in the Golgi-form of V_oV_1 (Table 5.1 and Figure 5.9). Although 100 % sequence coverage was not achieved for the two proteins (83 % for subunit e and 69 % for subunit f), they appear to be the same form as those found in the vacuolar form of the complex, notwithstanding any variation that may occur in the unsequenced protein regions. This is surprising on account of the low sequence identity of the two subunit a isoforms (54 % identity). Inspection of the cryo-EM structure of subunits a, e, and f in the vacuolar enzyme (Figure 5.9B, PDB ID: 5TJ5 [14]) shows that

Protein Description	Accession	Protein Mass [Da]	Hit no.	Protein score	# Spectra	# Peptides	Seq. coverage [%]
			A1 Trypsin				
VPH2	P32341	25271	1	499	19	12	39
Subunit F	P39111	13453	2	456	15	11	56
Subunit e	**Q3E7B6**	**8375**	**3**	**216**	**37**	**3**	**23**
STV1	P37296	101595	4	181	11	10	9
Subunit f	**P0C5R9**	**9355**	**6**	**148**	**8**	**3**	**20**
Subunit d	P32366	39765	7	142	3	3	12
			A2 Trypsin				
STV1	P37296	101595	1	295	18	14	15
Subunit e	**Q3E7B6**	**8375**	**4**	**91**	**7**	**3**	**23**
V0 assembly	P53262	29659	7	46	1	1	2
			A3 Trypsin				
Subunit d	P32366	39765	2	108	2	2	6
STV1	P37296	101595	4	81	7	6	6
			B1 Chymotrypsin				
VPH2	P32341	25271	1	269	16	12	**39**
Subunit f	**P0C5R9**	**9355**	**2**	**204**	**23**	**10**	**69**
Subunit c'	P32842	17026	3	195	14	9	54
Subunit F	P39111	13453	4	164	22	19	94
V0 assembly	P53262	29659	5	140	22	3	12
Subunit e	**Q3E7B6**	**8375**	**6**	**102**	**16**	**10**	**79**
STV1	P37296	101595	7	44	3	3	3
			B2 Chymotrypsin				
V0 assembly	P53262	29659	1	239	24	4	**16**
STV1	P37296	101595	3	96	15	10	**9**
Subunit e	**Q3E7B6**	**8375**	**4**	**88**	**13**	**9**	**83**
Subunit F	P39111	13453	5	85	13	11	72
			B3 Chymotrypsin				
STV1	P37296	101595	2	88	10	5	**5**
V0 assembly	P53262	29659	4	46	5	1	5
			C1 Trypsin and Chymotrypsin				
Subunit F	P39111	13453	1	345	18	14	**73**
Subunit c'	P32842	17026	2	268	25	15	58
Subunit f	**P0C5R9**	**9355**	**3**	**188**	**11**	**7**	**57**
VPH2	P32341	25271	4	134	9	9	**24**
STV1	P37296	101595	5	124	9	9	8
Subunit e	**Q3E7B6**	**8375**	**6**	**99**	**10**	**8**	**50**
V0 assembly	P53262	29659	7	80	5	5	9
Subunit c	P25515	16340	9	66	3	3	18
			C2 Trypsin and Chymotrypsin				
Subunit F	P39111	13453	1	366	20	15	**77**
Subunit e	**Q3E7B6**	**8375**	**2**	**117**	**13**	**10**	**64**
V0 assembly	P53262	29659	3	99	9	6	15
STV1	P37296	101595	4	98	9	9	6
Subunit d	P32366	39765	5	65	1	1	2
Subunit c	P25515	16340	10	36	1	1	6
			C3 Trypsin and Chymotrypsin				
STV1	P37296	101595	1	211	16	12	77
Subunit d	P32366	39765	3	64	3	3	**15**
V0 assembly	P53262	29659	4	48	4	4	**6**
Subunit c'	P32842	17026	7	44	2	2	**9**
Subunit e	**Q3E7B6**	**8375**	**12**	**34**	**2**	**2**	**27**

Table 5.1: Proteomics analysis of the Golgi-form of V_oV_1. All ATPase protein identifications including the protein subunit, accession code, protein mass, Mascot hit number, protein score, number of identified spectra, number of identified peptides, and sequence coverage are given. Identifications of subunits *e* and *f* are shown in blue and red, respectively.

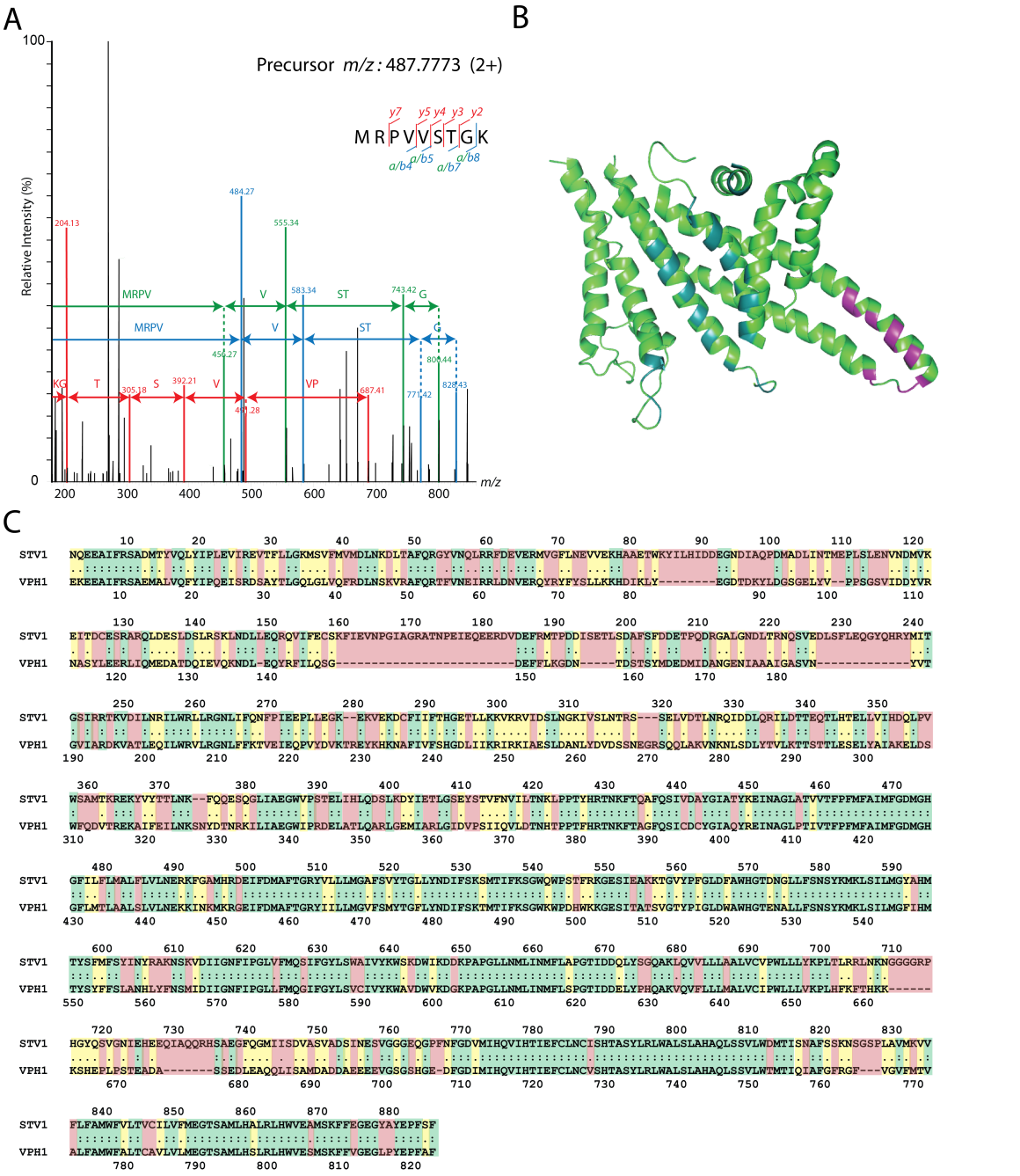


Figure 5.9: (A) Example MS/MS spectrum for protein identification. The peptide MRPVVSTGK from subunit *f* is identified with a, b, and y ions (green, blue, and red, respectively). (B) Interfaces between subunits *a*, *e*, and *f* as identified in the cryo-EM structure (PDB ID: 5TJ5 [14]). Subunit *a* (green) is shown with residues highlighted which contact subunit *e* (cyan) and subunit *f* (purple). (C) Comparison of amino acid sequences of subunit *a*; vacuolar (Vph1p) and Golgi (Stv1p). Amino acid residues between the two isoforms that are identical (green), similar (yellow), and disimilar or absent (red) are indicated.

subunit *e* contacts residues towards the centre of the subunit *a* (Vph1p) protein chain. A sequence alignment of Vph1p and Stv1p shows that most of the residues comprising this interaction interface are conserved between Vph1p and Stv1p (Figure 5.9C). Subunit *f* contacts the C-terminal region of subunit *a* (Vph1p), which while sharing some sequence homology with Stv1p has substantial differences, in some cases near the interaction interface between subunits *a* and *f*, making the conservation of the interaction between subunit *f* and Stv1p surprising.

5.4.1.2 The Role of Phosphorylation in V-type ATP Synthase Autoinhibition

The switch controlling disassembly and reassembly of the V_o/V_1 complex is thus far unknown. As phosphorylation has been shown to affect protein-protein interactions in multiple cases, it was thought to be responsible for the disassembly/reassembly of V_oV_1 in this case. Furthermore, a phosphorylation site on subunit *a* has been previously identified in a cytosolic loop between helices 6 and 7 (Ser677). This phosphorylation site is conserved in the Stv1p isoform (Ser735) and its location would place it such that phosphorylation/dephosphorylation may control the disassembly of the enzyme. To investigate this, excised gel bands corresponding to subunit *a* (Vph1p) from SDS-PAGE of samples of purified intact V_oV_1 and glucose starvation-induced dissociated V_o were provided by Stephanie Bueler and Professor John Rubinstein (The University of Toronto, Canada). Protein was digested in-gel with trypsin and/or chymotrypsin. Half of the samples was directly analysed by LC-MS/MS while the other half was treated with TiO_2 to enrich phosphopeptides, before LC-MS/MS analysis. Data were database searched with Mascot search engine for phosphorylation sites (Table 5.2).

Five phosphorylation sites (Ser46, Thr161, Ser162, Ser676, Ser677) were identified

Sample	Enriched	Enzyme	Protein score	Phosphosite	Score
Vo	No	Trypsin	75393		
	Yes	Trypsin	44038	S46	23
	No	Chymotrypsin	25936	S677	81
				S162	30
	Yes	Chymotrypsin	12070	T161	39
				S162	54
	No	Trypsin and Chymotrypsin	26058	S162	24
				S676	164
	Yes	Trypsin and Chymotrypsin	9948	S46	29
				S677	66
VoV1	No	Trypsin	62778		
	Yes	Trypsin	16089	S46	13
	No	Chymotrypsin	21888	S162	32
				S676	73
	Yes	Chymotrypsin	7035	S162	44
	No	Trypsin and Chymotrypsin	23442	S676	136
				S677	124
	Yes	Trypsin and Chymotrypsin	7381	S162	26
				S677	74

Table 5.2: Proteomics of gel bands. Gel bands were excised for V_o or V_oV_1 and either digested with trypsin, chymotrypsin or both enzymes. After digestion, some samples were enriched for phosphopeptides with TiO_2 as indicated. Data were searched with Mascot search engine and the Mascot scores are indicated.

in the V_o samples and four (Ser46, Ser162, Ser676, Ser677) in the V_oV_1 samples. One phosphosite was only identified in the V_o sample; *p*-Thr161. *p*-Thr161 however had low localisation scores and therefore it is ambiguous if these spectra correspond to *p*-T161 or *p*-S162, which was identified in the V_oV_1 sample. Importantly, *p*-Ser677 was observed in both V_oV_1 and V_o samples, indicating that this is likely not the switch controlling autoinhibition.

This analysis therefore did not support the hypothesis of a single phosphorylation site on subunit *a* acting as a switch for V_oV_1 dissociation. It is possible that phosphorylation does play a role in regulating this autoinhibitory mechanism, but the phosphorylation site may not have been identified here on account of the difficult nature of phosphosite identification coupled with the hydrophobicity of

subunit *a*, which makes bottom-up proteomics approaches challenging. It is also possible that another subunit is the control point for autoinhibition. Inspection of the structure and subunit organisation of V_oV_1 shows that numerous subunits are located at the dissociation interface, including subunit *H*, located next to subunit *a*. Further proteomics studies of the other subunits of V_oV_1 therefore may yield useful insights into this as yet unresolved question.

5.5 The Spinach Chloroplast F-type ATP Synthase

Having considered the role of PTMs in the regulation of the V-type ATPase from the yeast *S. cerevisiae*, the cATP synthase located on the thylakoid membrane of chloroplasts in *S. oleracea* was next investigated. Phosphorylation has been shown to play a role in regulating and stabilising protein-protein interfaces in cATP synthase [99], however our current understanding is far from complete. In this section I will use quantitative cross-linking to investigate the effects of two important PTMs; phosphorylation and acetylation which have both been identified on cATP synthase but have not been fully characterised [266, 320]. I will therefore investigate the effect of these PTMs on the stability and subunit conformation of the cATP synthase.

5.5.1 Results and Discussion

To investigate the effect of deacetylation on the cATP synthase, a comparative chemical cross-linking strategy was employed along with the deacetylase enzyme SIRT3, a mitochondrial enzyme which deacetylates cATP synthase in *H. sapiens*, *B. taurus*, and *A. thaliana* [321–323]. Professor Carla Schmidt (The Martin Luther University of Halle-Wittenberg) purified cATP synthase samples and performed the

comparative cross-linking experimental procedure up to and including data acquisition. cATP synthase was incubated with SIRT3 and cross-linked with BS3-d4 while an aliquot of untreated cATP synthase was cross-linked with BS3-d0. After cross-linking, the two aliquots were mixed 1:1, digested with trypsin, and analysed by LC-MS/MS. After data analysis, 54 unique cross-links could be quantified (Figure 5.10 and Appendix Table C.1), of which 25 were inter-protein cross-links. Most of these cross-links showed no significant difference in intensity after SIRT3 incubation. As expected, many of the cross-linked peptides were found to be acetylated. As acetylated lysine residues are not able to undergo BS3 cross-linking, an increase in the signal intensity of cross-linked peptides would be expected after deacetylation. On the contrary, a significant decrease in intensity was observed for two cross-links located on subunit ϵ (Figure 5.11). A cross-link on subunit ϵ between the N-terminus and adjacent β-domain ϵ(N-term)–ϵ(K20) showed intensity decreased by a factor of seven while a cross-link in the hinge region of the anti-parallel double-α-helix decreased by two orders of magnitude ϵ(K105)–ϵ(K112).

The reduction in two cross-links on the ϵ subunit in response to deacetylation can be rationalised by one of two scenarios (i) in-solution dissociation and unfolding of subunit ϵ or (ii) a conformational change of subunit ϵ, known to be an important switch between active and inactive states during the light/dark cycle. To distinguish between these two scenarios, reported native MS data [266] can be considered. Here, after deacetylation, an increase in in-solution dissociation of ϵ was observed, along with an increase in gas-phase dissociation of subunits ϵ and δ. The gas-phase dissociation product arising from ϵ dissociation increases much more so than that arising from δ dissociation. If the first scenario of in-solution unfolding and subunit dissociation is occuring, then an increase in dissociated ϵ would be expected by native MS, without CID. However, if a conformational change and elongation of the

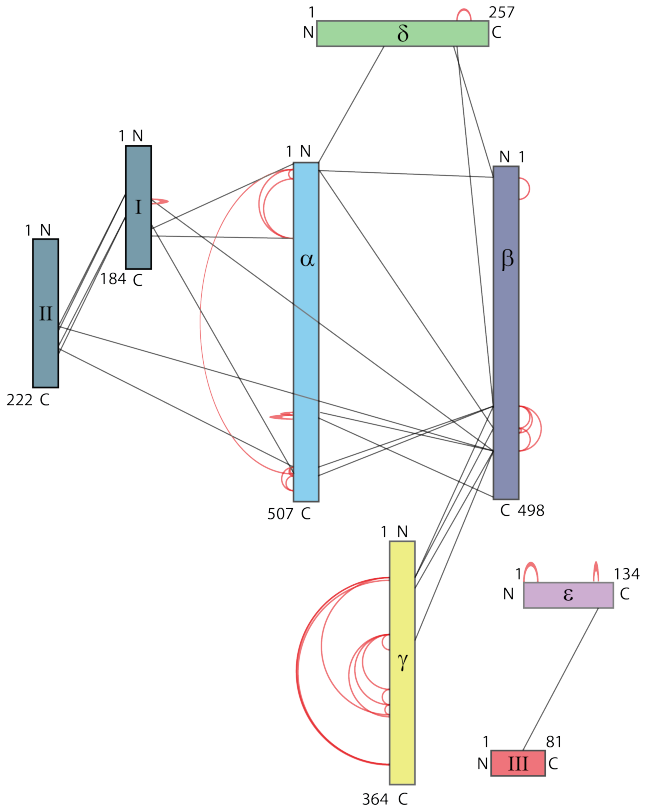


Figure 5.10: Comparative cross-linking of the cATP synthase quantified after treatment with SIRT3 deacetylase enzyme. Protein subunits are indicated by bars representing the amino acid sequence, with cross-links depicted as lines in red and grey for intra-protein and inter-protein cross-links, respectively. Figure from Schmidt *et al.* [266].

hinge-region is occurring, then an increase in gas-phase dissociation of ϵ would be expected as it is known that extended subunits dissociate more readily with CID [324, 325]. The reported MS data show an increase in both in-solution and gas-phase dissociation of ϵ. However, the gas-phase dissociation product arising from ϵ dissociation increases relative to δ dissociated product. This indicates elongation of ϵ without in-solution dissociation, consistent with inactivation of cATP synthase in response to deacetylation.

Previous studies have shown that the basic arginine residues at the C-terminus of ϵ are important in maintaining the inhibitory effect of extended ϵ [305], most likely

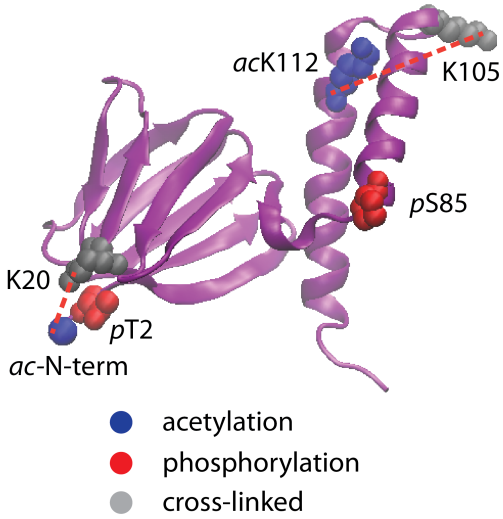


Figure 5.11: Comparative cross-links of the ϵ subunit of the cATP synthase, treated with SIRT3 deacetylase enzyme. Known phosphorylation (red) and acetylation sites (blue) [266], cross-linked lysine residues (grey) and cross-links (dashed lines) and highlighted on a homology model of subunit ϵ [99]. Figure from Schmidt *et al.* [266].

via electrostatic interactions with the DELSEED motif of subunit β [308]. As acetylation neutralises an otherwise basic residue (lysine/arginine), one would expect that deacetylation would promote the ϵ-β electrostatic interactions and thus the inhibitory effect of ϵ. It is this scenario which is observed by comparative cross-linking, with a reduction in the cross-links corresponding to the 'compact' form of ϵ after deacetylation. Therefore, these cross-links indicate that deacetylation promotes the inhibitory role of subunit ϵ in its extended conformation, through strengthening of electrostatic interactions. It has been proposed that the availability of nucleotide determines the conformation of subunit ϵ, with the extended conformation favoured at low ATP availability. It is plausible therefore, that deacetylase enzymes are activated when ATP decreases and thus deacetylation contributes to the ϵ-mediated inhibition of F_1 to prevent wasteful hydrolysis of ATP.

As phosphorylation has been shown previously to decrease nucleotide occupancy

and affect protein-protein interactions in cATP synthase [99], the simultaneous effects of both deacetylation and dephosphorylation were investigated using comparative cross-linking. Professor Carla Schmidt (The Martin Luther University of Halle-Wittenberg) performed comparative cross-linking experiments on cATP synthase, as before. An aliquot of cATP synthase, incubated with SIRT3 and a phosphatase enzyme, was cross-linked with BS3-d4 while untreated cATP synthase was cross-linked with BS3-d0. The two aliquots were combined 1:1, digested with trypsin and analysed by LC-MS/MS. The presence of multiple modifications on peptides presented difficulties in quantification. However, after data processing, 19 cross-links could be quantified (Figure 5.12 and Appendix Table C.2), of which eight were inter-protein cross-links. Unlike for deacetylated cATP synthase, the intensity of most cross-links changed significantly upon simultaneous deacetylation and dephosphorylation with most cross-linking intensities decreasing. This can be accounted for by a decrease in the stability of protein-protein interactions in response to removal of both PTMs.

The simultaneous removal of acetylation and phosphorylation appears to have a greater effect than the sum of separately deacetylating and dephosphorylating [99]. Taken together with the cross-linking of deacetylation only, these results imply that deacetylation induces local changes, for example on subunit ϵ but phosphorylation is important for global stability with dephosphorylation destabilising various protein-protein interfaces.

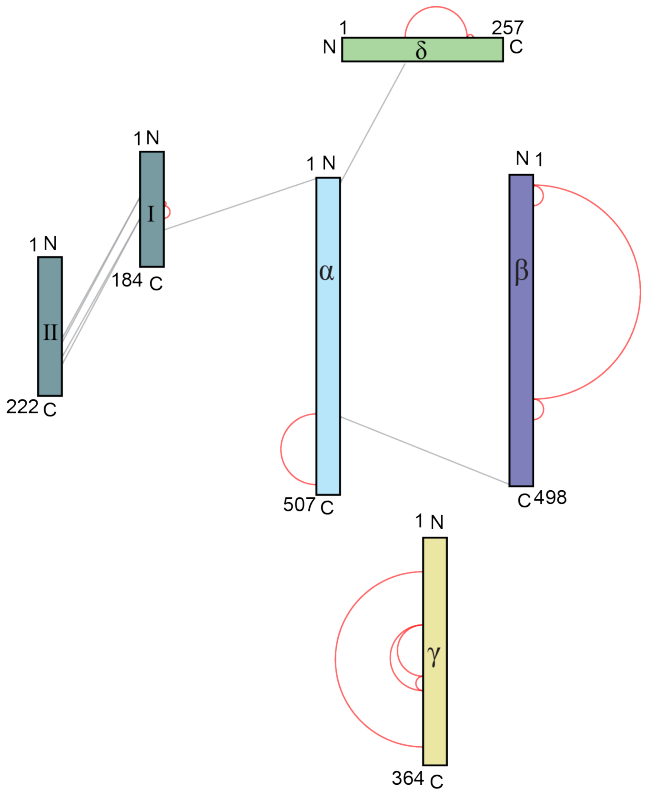


Figure 5.12: Comparative cross-linking of the cATP synthase quantified after treatment with both deacetylase and phosphatase enzymes. Protein subunits are indicated by bars representing the amino acid sequence, with lines indicating intra-protein (red) and inter-protein (grey) cross-links. Fewer cross-links can be quantified than when cATP synthase is treated with SIRT3 deacetylase enzyme alone. Figure produced with xVis software [247].

5.6 Bovine Mitochondrial F-type ATP Synthase

Having investigated the influence of PTMs on ATP synthases in yeast and plants, I will now consider the role of PTMs and other regulatory mechanisms in a mammalian ATP synthase from the mitochondria of *B. taurus* heart. Purified mitochondrial ATP synthase from *B. taurus* heart was provided by Professor Christoph Gerle (The University of Osaka, Japan). Numerous PTMs have been described in bovine ATP synthase with a high level of interest deriving from their implication in heart disease [297], however the identification and characterisation of these PTMs is far from complete. Furthermore, details of the interaction of the

endogenous inhibitor IF_1 with ATP synthase remain unclear, including an apparent discrepancy in stoichiometry between different species and uncertainty in the effect of IF_1 on nucleotide binding [268]. Finally, the location and interaction of many of the small, membrane subunits of the ATP synthase have remained elusive, despite previous work motivated by their connection with diseases such as diabetes [280]. In this section I will use MS-based methods to explore these multiple areas of uncertainty in order to provide a more complete picture of the regulatory mechanisms of ATP synthase employed in mammals.

5.6.1 Results

5.6.1.1 Proteomics Analysis

To verify that the intact bovine ATP synthase was purified with a full complement of protein subunits, proteomics analysis was performed. As various PTMs have been described previously in mitochondrial ATP synthases (reviewed in [326]), the presence of PTMs in the mitochondrial ATP synthase from *B. taurus* was also investigated. The ATP synthase was digested with trypsin and analysed by LC-MS/MS. To aid in the identification of phosphorylation sites, which have been previously shown to modify ATP synthases, phosphopeptides were also enriched by TiO_2-affinity purification prior to LC-MS/MS analysis. Data were searched with Mascot search engine and spectra verified by visual inspection. As the protein sequences listed in this database did not take into account the expected cleavage of N-terminal transit (signal) peptides, data were also searched with MaxQuant software against the expected mature protein sequences.

After this analysis, all 17 expected protein subunits, in addition to the inhibitor IF_1, were identified. While most subunits were identified with high sequence coverage, some of the hydrophobic subunits had lower sequence coverage, notably

subunits *a*, *c*, and *6.8PL*. In order to identify PTMs, data were searched using Mascot search engine and MaxQuant software allowing for variable modifications of phosphorylation, methylation, trimethylation, and acetylation. These modifications have been shown previously to occur in bovine mitochondria and play important roles in regulation [326]. After database searching and manual validation of spectra, 23 phosphorylation, 8 methylation, and 46 acetylation sites (of which subunits *f* and *g* were acetylated on their N-termini) were identified (Tables 5.3 and 5.4). In total, 77 modification sites were found, which were distributed across 14 of the protein subunits.

In order to place the identified PTMs in the context of the structure, modified residues were located on a structure of the bovine ATP synthase determined by cryo-EM (Figure 5.13, PDB ID: 5ARA [278]). Although most PTMs could be placed on the structure, the lack of structural information for some regions of the protein complex, including many of the membrane subunits, prevented all modifications from being located. Acetylation, the most prevalent of the PTMs identified here, is distributed across the protein surface, particularly on the soluble $\alpha_3\beta_3$ head. On the other hand, the phosphorylation and methylation sites are preferentially located at protein-protein interfaces.

5.6.1.2 Denaturing Mass Spectrometry

The identification of a large number of PTMs across multiple protein subunits indicates that the subunit masses may differ from the protein sequence mass and in some cases may differ significantly, potentially hindering assignment of any native mass spectra. Furthermore, PTM sites are rarely at 100 % occupancy and thus accurate subunit masses can give an indication of the overall PTM status of the

Protein	Coverage (%)		Modification	Residue	Peptide sequence	Direct		Enriched	
	M	MQ				M	MQ	M	MQ
α	93	91	Ph (T)	48	TGT*AEVSSILEER			82	150
			Ph (S)	52	TGTAEVS*SILEER			75	101
			Ph (S)	53	TGTAEVSS*ILEER			104	127
			Ph (T/S)	64/65	ILGADT*S*VDLEETGR				50
			Ph (S)	65	ILGADTS*VDLEETGR			41	
			Ac (K)	103	NVQAEEMVEFSSGLK*GMSLNLEPDNVGVVVFGNDKLIK		82		
			Ac (K)	167	VVDALGNAIDGKGPIGSK*AR	95	146		
			Ac (K)	175	VGLK*APGIIPR	53	62		
			Ph (T)	236	RFNDGT*DEK			31	76
			Ac (K)	239	FNDGTDEK*K				77
			Ac (K)	240	FNDGTDEKK*			52	
			Ac (K)	305	DNGK*HALIIYDDLSK	57	68		
			Ac (K)	427	AMK*QVAGTMKLELAQYR		88		
			Ac (K)	434	QVAGTMK*LELAQYREVAAFAQFGSDLDAATQQLLSR	53			
			Me (R)	527	IR*TDGKISEESDAK		58		
			Ac (K)	531	TDGK*ISEESDAKLK	51	107	95	
β	95	96	Ac (K)	124	GQK*VLDSGAPIRIPVGPETLGR	38	98		
			Ac (K)	159	IMNVIGEPIDERGPIK*TK		105		
			Ac (K)	159/161	IMNVIGEPIDERGPIK*TK*	66			
			Ac (K)	212	IGLFGGAGVGK*TVLIMELINNVAK	44	46		
			Ac (K)	351	K*GSITSVQAIYVPADDLTDPAPATTFAHLDATTVLSR	76	144		
			Ph (S)	415	IMDPNIVGS*EHYDVAR				86
			Ac (K)	480	FLSQPFQVAEVFTGHLGK*LVPLKETIK	54			
			Ac (K)	522	ADK*LAEEHS		71		
γ	88	85	Ac (K)	79	VYGVGSLALYEK*ADIKTPEDK	87	81		
			Ph (T)	84	ADIKT*PEDK			40	46
			Ac (K)	89/90	K*K*HLIIGVSSDR	32	58		
			Me (R)	100	HLIIGVSSDR*GLCGAIHSSVAK		67		
			Ac (K)	129	EVK*IIGVGDKIR	47	105		
			Ph (T)	144	T*HSDQFLVTFK			48	
			Ph (S)	146	THS*DQFLVTFK			27	40
			Ac (K)	154	THSDQFLVTFK*EVGR	68	132		
			Ac (K)	201	TEEK*PIFSLDTISSAESMSIYDDIDADVLR		55		
			Me (K)	262	MTAMDNASK*NASEMIDKLTLTFNR		85		
			Ph (S)	265	NAS*EMIDKLTLTFNR		49		
			Ac (K)	270	NASEMIDK*LTLTFNR		60		
δ	88	76	Ac (K)	136	ANLEK*AQSELLGAADEATRAEIQIR	60	228		
ε	94	94	Me (K)	21	VAYWRQAGLSYIRYSQICAK*		52		
			Ph (T)	29	DALKT*EFK			35	

Table 5.3: PTMs of the core F_o subunits identified by proteomics analysis and database search using Mascot search engine (M) and MaxQuant software (MQ). The protein subunit, sequence coverage, modification (Acetylation (Ac), Methylation (Me), Phosphorylation (Ph)) and modified residue, peptide sequence, and peptide scores from direct analysis and TiO_2 enrichment are given. If a modification could not be unambiguously assigned to one residue, both possibilities are listed; i.e. 64/65.

Protein	Coverage (%)		Modification	Residue	Peptide sequence	Direct		Enriched	
	M	MQ				M	MQ	M	MQ
a	36	12							
b1	77	85	Ac (K)	53	PVPPLPEHGGK*VR		87		
			Ph (S)	119	YGAS*VGEFADK			36	
			Ac (K)	144	QASIK*QIQDAIDMEK		85		
			Ac (K)	154	QIQDAIDMEK*SQQALVQKR	99	117		
			Ph (S)	155	S*QQALVQK			36	87
			Ac (K)	221	QKEQEHMINWVEK*R/EQEHMINWVEK*R	45	86		
			Ph (S)	226	VVQS*ISAQQEK			51	138
			Ph (S)	228	VVQSIS*AQQEK			29	
c	0	9							
d	97	98	Ac (K)	48	LATLPEK*PPAIDWAYYK	63	64		
			Ac (K)	58	LATLPEKPPAIDWAYYK*ANVAK	46	62		
			Ac (K)	63	ANVAK*AGLVDDFEKK	71	94		
			Ac (K)	73	K*FNALKVPIPEDK	47	63		
			Ac (K)	78	FNALK*VPIPEDK	48	89		
			Ac (K)	85	VPIPEDK*YTAQVDAEEKEDVK	68	135		
			Ph (Y)	86	Y*TAQVDAEEK				79
			Ac (K)	95	VPIPEDKYTAQVDAEEK*EDVK	71	97		
			Ac (K)	117	TRIQEYEK*ELEK		49		
			Me (K)	121	TRIQEYEKELEK*		49		
			Ac (K)	144	NIIPFDQMTIEDLNEVFPETK*LDK		49		
			Ac (K)	147	NIIPFDQMTIEDLNEVFPETKLDK*K		123		
			Me (K)	148	NIIPFDQMTIEDLNEVFPETKLDKK*		71		
			Ac (K)	149	K*YPYWPHRPIETL	40	88		
e	81	87	Ph (S)	8	VPPVQVS*PLIK			34	43
			Ph (Y)	30	Y*NYLKPR			41	118
			Ph (Y)	32	YNY*LKPR			40	133
f	90	66	Ac (N-term)	2	A*SVVPLKEKK	34		33	
			Ac (K)	16	LLEVK(1)LGELPSWILMR		50		
g	87	70	Ac (N-term)	2	A*EFVRNLAEKAPALVNAAVTYSKPR	77			
			Ac (K)	11	NLAEK*APALVNAAVTYSKPR	87	87		
			Ac (K)	54	VELVPPTPAEIPTAIQSLK*K		86		
A6L	75	59							
F6	84	88	Ph (S)	108	FEVVEKPQS*			54	119
OSCP	92	95	Me (K)	26	FAK*LVRPPVQIYGIEGR		69		
			Ac (K)	60	QNKLEQVEK*ELLR	69	167		
			Ac (K)	100	EK*FSPLTSNLINLLAENGR	61	147		
			Me (R)	136	LTNTPAVISAFSTMMSVHR*GEVPCTVTTASALDEATLTELK		55		
			Ac (K)	158	GEVPCTVTTASALDEATLTELK*TVLK	52	63		
DAPIT	82	45	Ph (T)	25	YFNSYTLT*GR			64	93
6.8 PL	0	13	Ph (S)	52	ASS*AAPAHGHH			23	90
IF1	53	74							

Table 5.4: PTMs of the F_1 domain and peripheral stalk subunits identified by proteomics analysis and database searching using Mascot (M) and MaxQuant software (MQ). The protein subunit, sequence coverage, modification (Acetylation (Ac), Methylation (Me), Phosphorylation (Ph)) and modified residue, peptide sequence, and peptide scores from direct analysis and TiO_2 enrichment are given.

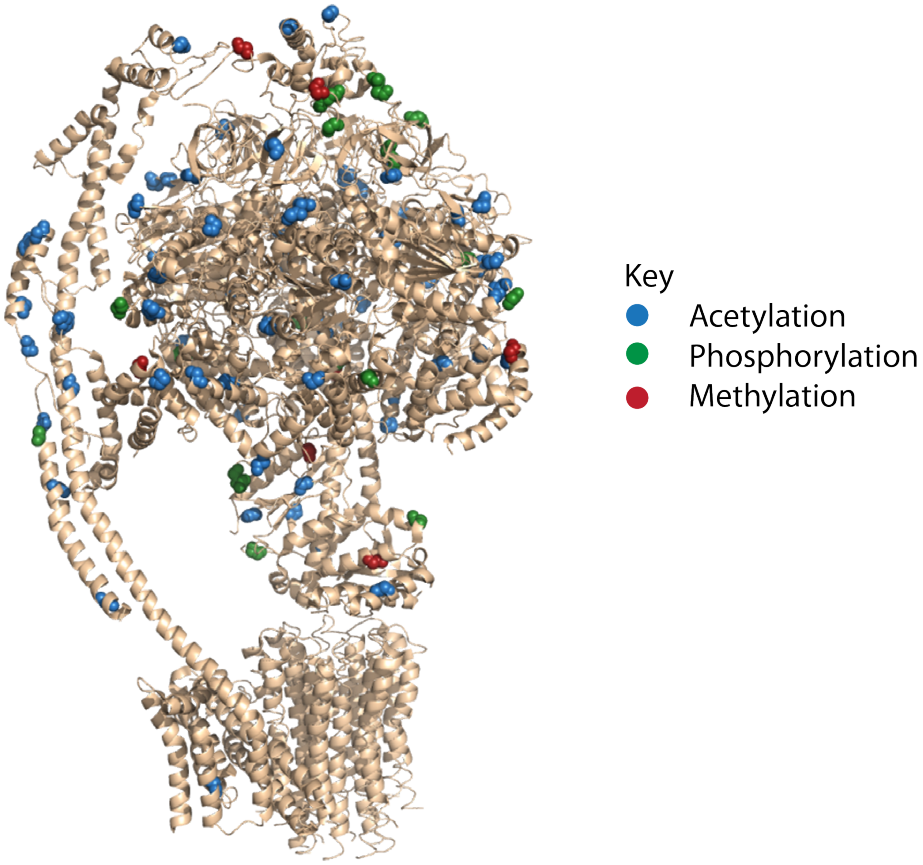


Figure 5.13: Post-translational modifications identified by proteomics are highlighted on the cryo-EM structure of F_oF_1 (PDB ID: 5ARA) [278]. Acetylation (blue), phosphorylation (green) and methylation (red) sites were identified.

protein subunits. Therefore, denaturing MS was used to determine accurate subunit masses. Proteins were separated by LC under denaturing conditions and analysed by MS (Figure 5.14). In a second experiment, F_oF_1 was denatured on a C4 column and analysed by MS (Figure 5.15). By combining both analyses, accurate masses for all but one (*b1*) of the subunits were determined (Table 5.5).

The inhibitory subunit IF_1 was identified by LC-MS, eluting early in the gradient (Figure 5.14A). Limited N-terminal proteolysis was observed with masses assigned to the intact subunit (9,580 Da) and also to species successively lighter by 57 Da, 87 Da, 129 Da, and 87 Da corresponding to the loss of amino acid residues

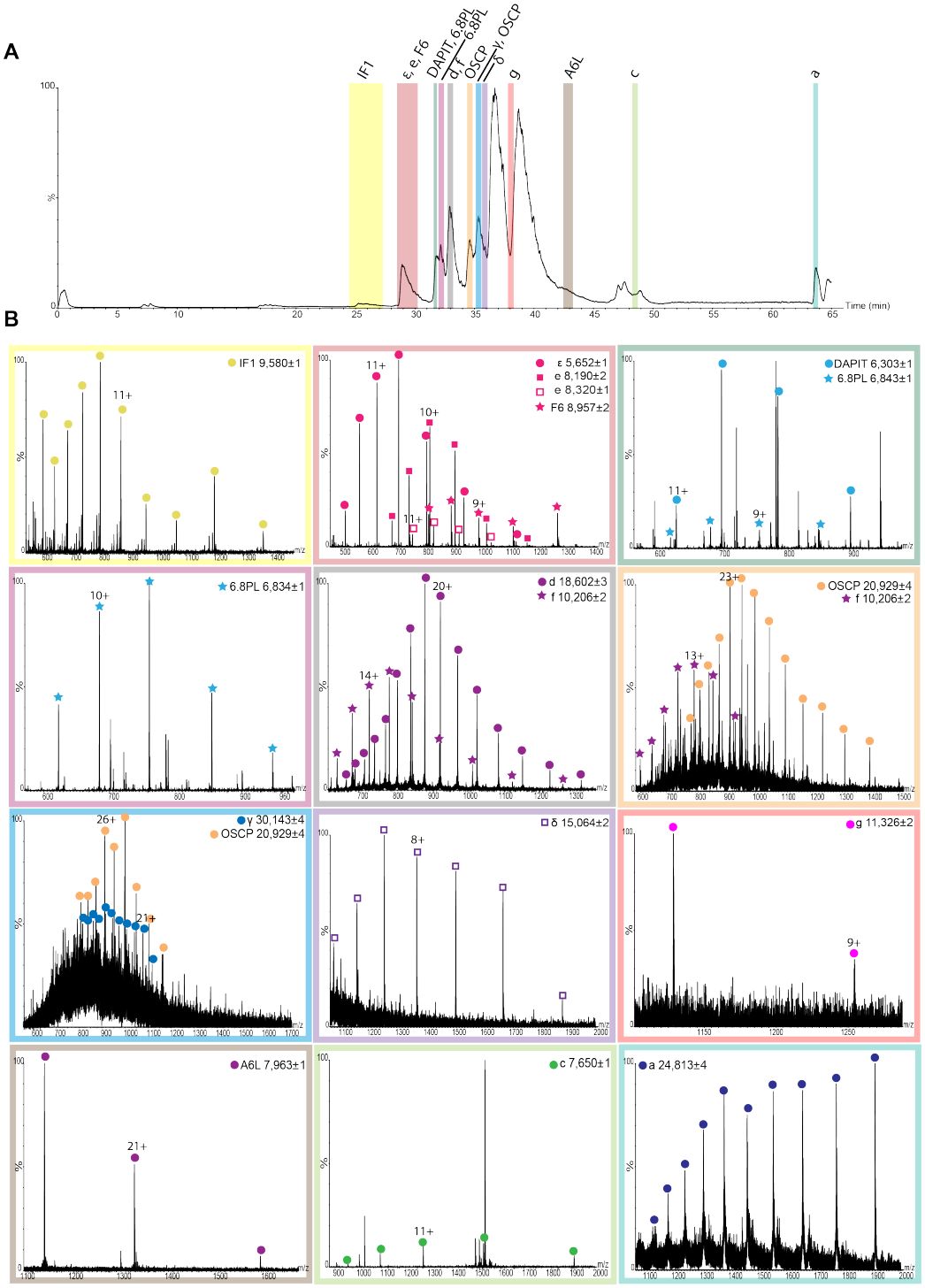


Figure 5.14: Accurate subunit mass determination using LC-MS. (A) The chromatogram with the elution of subunits identified. (B) Mass spectra of subunits allows determination of accurate masses for subunits IF_1, γ, δ, ϵ, *OSCP*, F_6, *A6L*, *a*, *c*, *d*, *e*, *f*, *g*, *6.8PL*, and *DAPIT*.

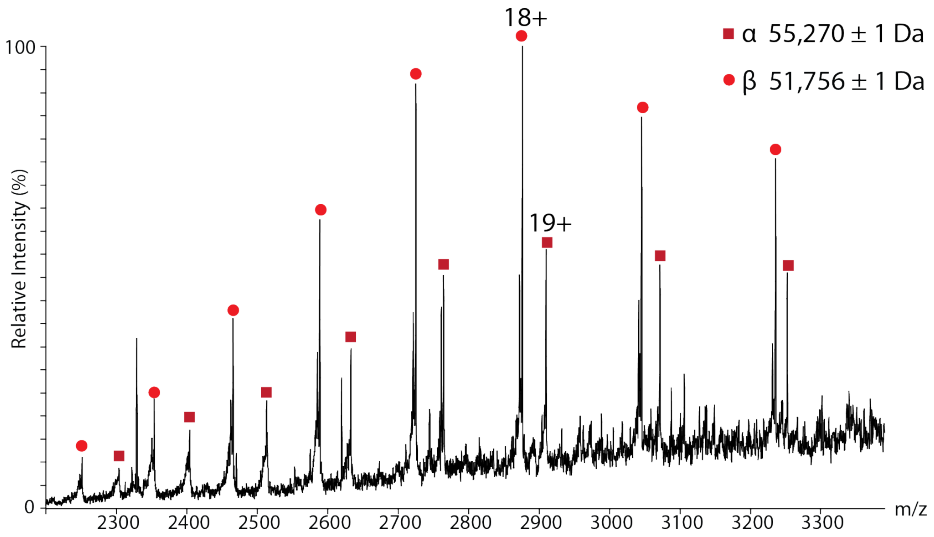


Figure 5.15: (A) Mass spectrum of denatured subunits α and β, for accurate mass determination.

Protein	Sequence Mass (Da)	Mature Subunit Mass (Da)	Measured Mass (Da)	Mass Deviation
α	59,720	55,229	55,270	Transit peptide
β	56,284	51,737	51,756	Transit peptide
γ	33,072	30,141	30,143	Transit peptide
δ	17,612	15,065	15,063	Transit peptide
ε	5,783	5,783	5,650	NME
a	24,788	24,788	24,813	N-term Form
b1	28,822	24,669	-	-
c	14,223	7,608	7,650	Transit peptide; Trimethyl
d	18,692	18,692	18,602	NME; N-term Acet
e	8,321	8,321	8,203	NME
f	10,297	10,297	10,206	NME; N-term Acet
g	11,417	11,417	11,326	NME; N-term Acet
A6L	7,973	7,973	7,963	-
F6	12,532	8,958	8,957	Transit peptide
OSCP	23,320	20,930	20,929	Transit peptide
DAPIT	6, 435	6,435	6,306	NME
6.8 kDa PL	6,834	6,834	6,834	-
IF1	12,301	9,576	9,580	Transit peptide

Table 5.5: Subunit masses determined by denaturing MS. The protein subunits, theoretical sequence masses, theoretical mature sequence masses after transit peptide cleavage, experimentally-measured mass, and proposed explanation of mass deviation are given.

Gly, Ser, Glu, and Ser, respectively and matching the N-terminal sequence of IF_1.

Although many of the subunit masses were close to the sequence mass, some deviated from this. In most cases, loss of a known N-terminal transit peptide sequence (subunits α, β, γ, δ, F_6, *OSCP*, and IF_1) or N-terminal methionine excision (NME) (ϵ, *e*, and *DAPIT*) accounted for these discrepancies. For subunit *e*, two masses were detected (8,190 and 8,320 Da), differing by $\approx$ 130 Da, which is attributed to NME, with the methionine-retaining protein being the minor form. The efficiency of NME can be influenced by the adjacent amino acid residue, in this case Val, which has been shown previously to inhibit the excision reaction [327] and is likely the reason for both forms being detected here. The mass deviations of the remaining subunits *a*, *c*, *d*, *f*, and *g* can be attributed to the previously observed modifications of NME coupled with N-terminal acetylation (subunits *d*, *e*, and *f*), N-terminal formylation (subunit *a*), and loss of an N-terminal transit peptide coupled with trimethylation (subunit *c*) [328].

5.6.1.3 Native Mass Spectrometry

Having characterised the PTM status of the subunits and determined accurate subunit masses, the non-covalent interactions present in the bovine F_oF_1 complex were analysed by native MS. The F_oF_1 buffer was exchanged to ammonium acetate supplemented with the mild detergent lauryl maltose neopentyl glycol (LMNG), in order to prevent complex dissociation, and analysed by native MS under conditions designed to promote the transmission of high mass complexes. In the resulting mass spectrum (Figure 5.16), the most intense charge series is observed around 9,000–11,000 *m/z* and the mass assigned to a well-resolved 405,352 ±102 Da complex (Appendix Table C.3). In addition, a complex of 821,444 ±129 Da is observed with peak broadness indicating a high degree of heterogeneity, possibly

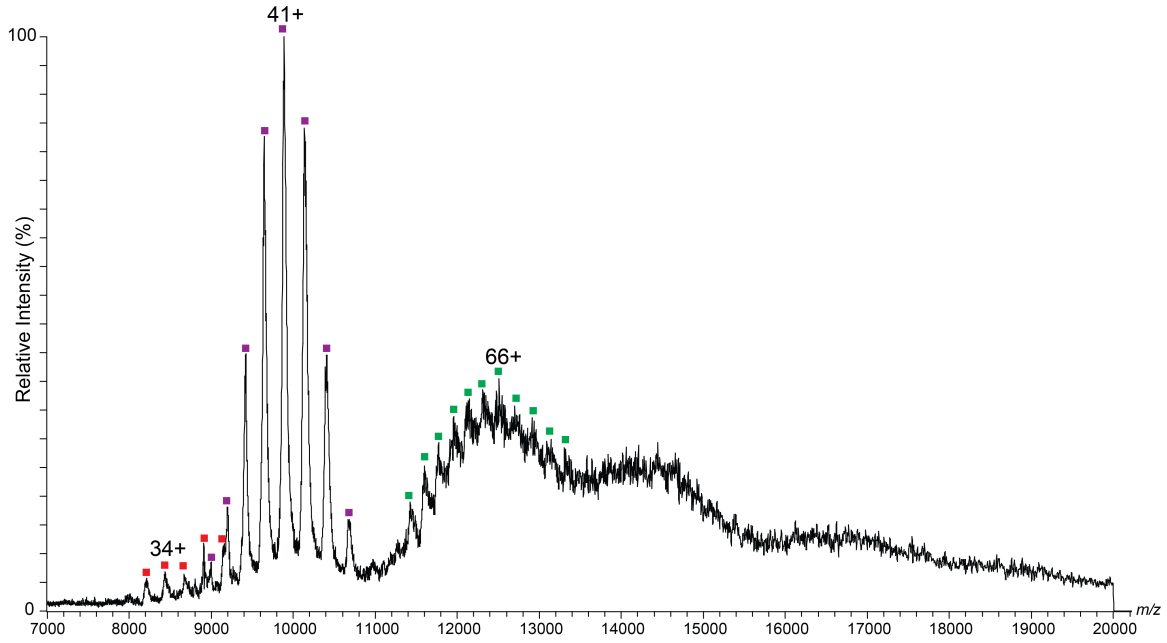


Figure 5.16: Native mass spectrum of mitochondrial F-type ATP synthase from *B. taurus*. A 287,105 ±31 Da complex (red squares), 405,352 ±102 Da complex (purple squares), and 821,444 ±129 Da complex (green squares) are observed. Broad humps around 14,000 m/z and 17,000 m/z cannot be assigned but may be dimers and other higher-order oligomers.

due to PTMs and adducts such as salts, nucleotides, lipids, and detergent. This mass (≈ 821 kDa) is between the mass of monomeric and dimeric F_oF_1 (theoretical masses ≈ 585 kDa and 1.17 MDa, excluding contributions from bound nucleotides, lipids, and detergent) and may be a disrupted dimer complex. Some poorly-resolved peaks are observed around 14,000 m/z and thought to be a dimer, however conclusive assignment cannot be made due to the significant peak broadening and noise observed, indicating the presence of multiple species with similar m/z values.

In order to better resolve the charge state series in the mass spectrum, the 821 kDa complex was selected with a wide m/z window and low collision energy applied to prevent complex dissociation. In the resulting mass spectrum (Figure 5.17) there was an increase in the population of the 821 kDa species detected, with less peak

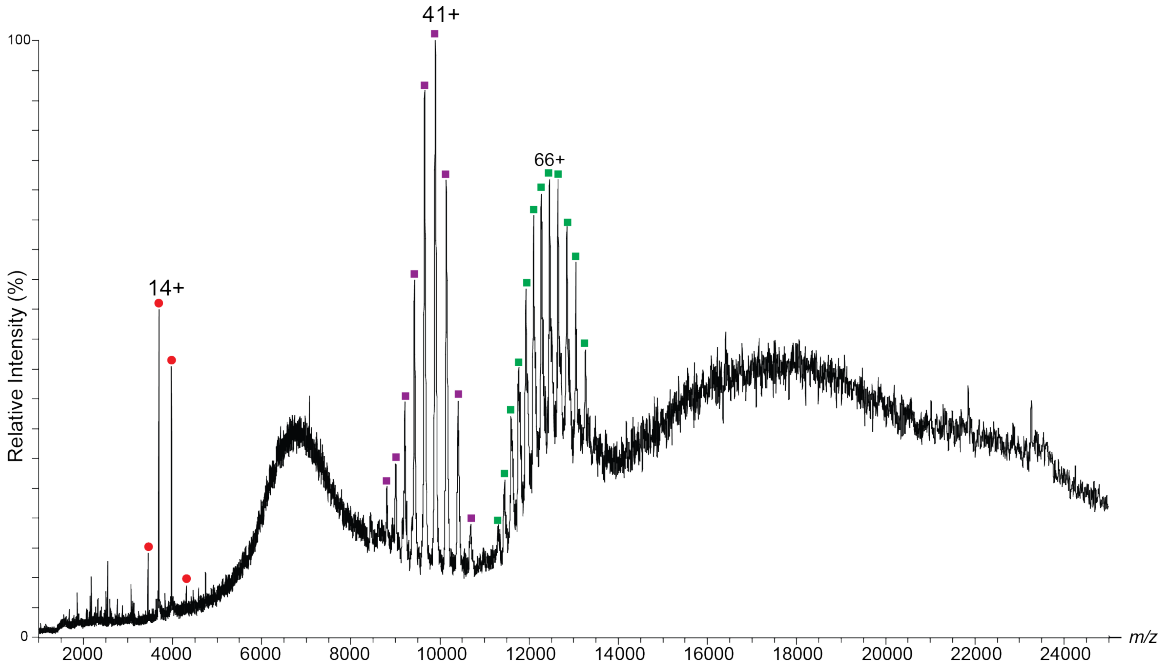


Figure 5.17: Native mass spectrum of F_oF_1. A complex of 822,006 ±74 Da, detected previously, is observed with increased intensity and resolution (green squares). In addition, species of mass 405,547 ±82 Da (purple squares) and 51,704 ±1 Da (red circles, subunit β) are observed.

broadening observed allowing a more accurate mass to be assigned (822,006 ±74 Da). The mass of this complex is between that expected for monomeric and dimeric F_oF_1, indicating that the complex identified here is an intermediate subcomplex, such as a dimer lacking one F_1, a complex previously observed by cryo-EM [329]. This charge state series is therefore tentatively assigned as such. Monomeric intact F_oF_1 is not however observed here, suggesting that the dimer interface is stronger than that between the F_o and F_1 domains in the monomer, under the conditions employed here. Although the mass selection together with lower collision energy afforded greater resolution of the 822 kDa charge state series, the overall baseline is increased with a broad hump now present around 17,000 *m/z*.

It was reasoned that if PTMs, including the phosphorylation sites identified earlier, are contributing to the peak broadness in the mass spectra then dephosphorylation would theoretically improve resolution as a result of reducing sample heterogeneity. Therefore, the F_oF_1 protein was partially dephosphorylated by incubating with a phosphatase enzyme prior to MS analysis. The resulting spectrum showed an increased resolution (Figure 5.18), with some complex dissociation observed, presumably due to a stabilising role that the removed phosphorylation was serving.

Close inspection of the 405 kDa complex revealed a second charge series corresponding to a 395 kDa complex, partially overlapping with the 405 kDa series indicating that the presence of this 395 kDa species may have contributed towards

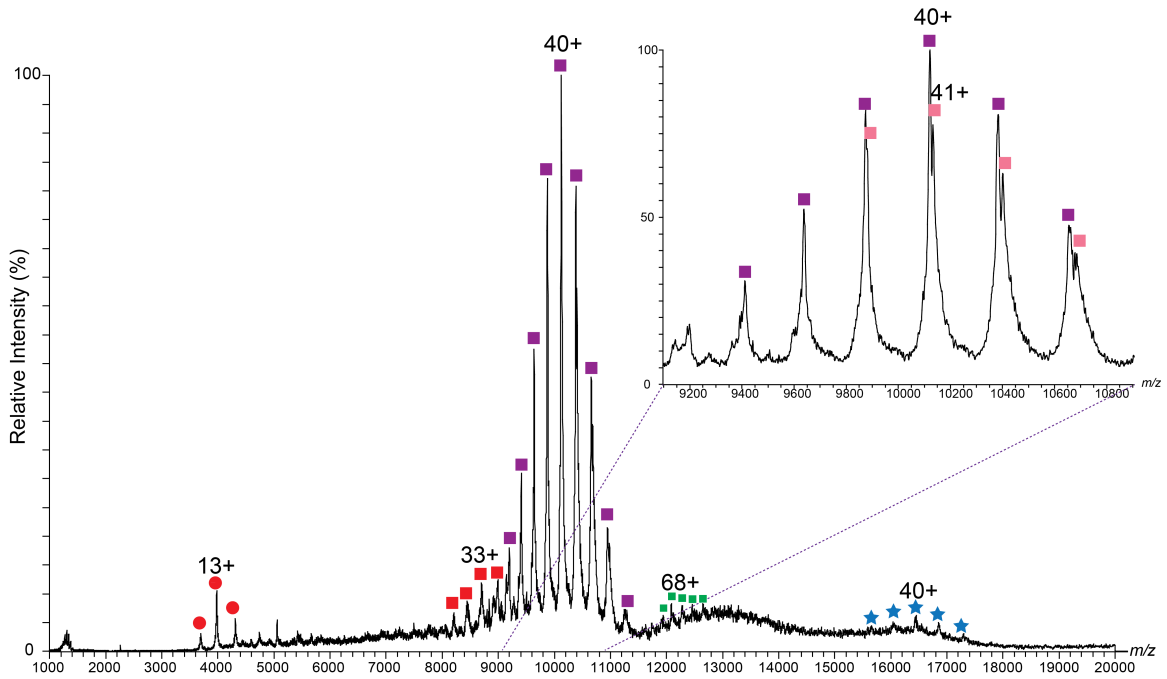


Figure 5.18: Native mass spectrum of F_oF_1 after phosphatase treatment. Close inspection of the 404,783 ±32 Da complex (purple squares) observed previously reveals the presence of a second species of 395,107 ±31 Da (inset, pink squares). In addition species of 51,700 ±1 Da (red circles, subunit β), 287,092 ±30 Da (red squares), 822,081 ±42 Da (green squares), and 657,328 ±93 Da (blue stars) are observed.

the peak broadness observed in Figure 5.16. The mass determined for the complex is now smaller by ≈ 700 Da (404.8 kDa vs 405.5 kDa), indicating removal of phosphorylation sites. In addition to these species, the 822 kDa complex identified previously was observed as a more clearly resolved series, albeit less intense than previously, around 12,000–14,000 *m/z*. A charge-stripped complex was observed around 15,500–17,500 *m/z* assigned to loss of subunits α and β. Finally, smaller complexes and subunits were observed in the lower *m/z* region of 1,000–6,000 *m/z*, including subunit β (51,700 ±1 Da, red circles) and a highly-charged subunit β at 1,500 *m/z*, indicating β loss by CID. The mass difference of the 405.5 and 395 kDa species is calculated to be 9,6766 Da which is very close to the mass of the IF_1 inhibitor determined by denaturing MS, the presence of which was also identified by proteomics analysis.

In order to assign the 405.5/395 kDa species to specific subunit compositions, the non-phosphatase treated F_oF_1 was analysed by MS and a peak selected for CID (m/z = 9642, charge state 42+ of 405.5 kDa complex). The resulting mass spectrum contains multiple collision products (Figure 5.19). In the region of the charge-stripped complexes (around m/z = 10,000) subcomplexes originating from both the 405.5 kDa and 395 kDa complexes are observed. Three sets of charge-stripped subcomplexes are observed (Figure 5.19 main panel and Figure 5.19B), and by deducting their mass from 405.5/395 kDa and comparing to the masses determined from denaturing MS, the CID products can be assigned. Charge-stripped complexes were therefore attributed to the loss of; 1) Subunit δ, 2) Subunit ϵ, and 3) *OSCP*. Highly-charged monomers corresponding to the mass of δ, ϵ, and *OSCP* were observed around 2,000 *m/z* (Figure 5.19A). In addition, a subunit with mass corresponding to the *b1* subunit, the only subunit not identified by denaturing MS analysis, was observed (24,691 ±1 Da, Appendix Figure C.1).

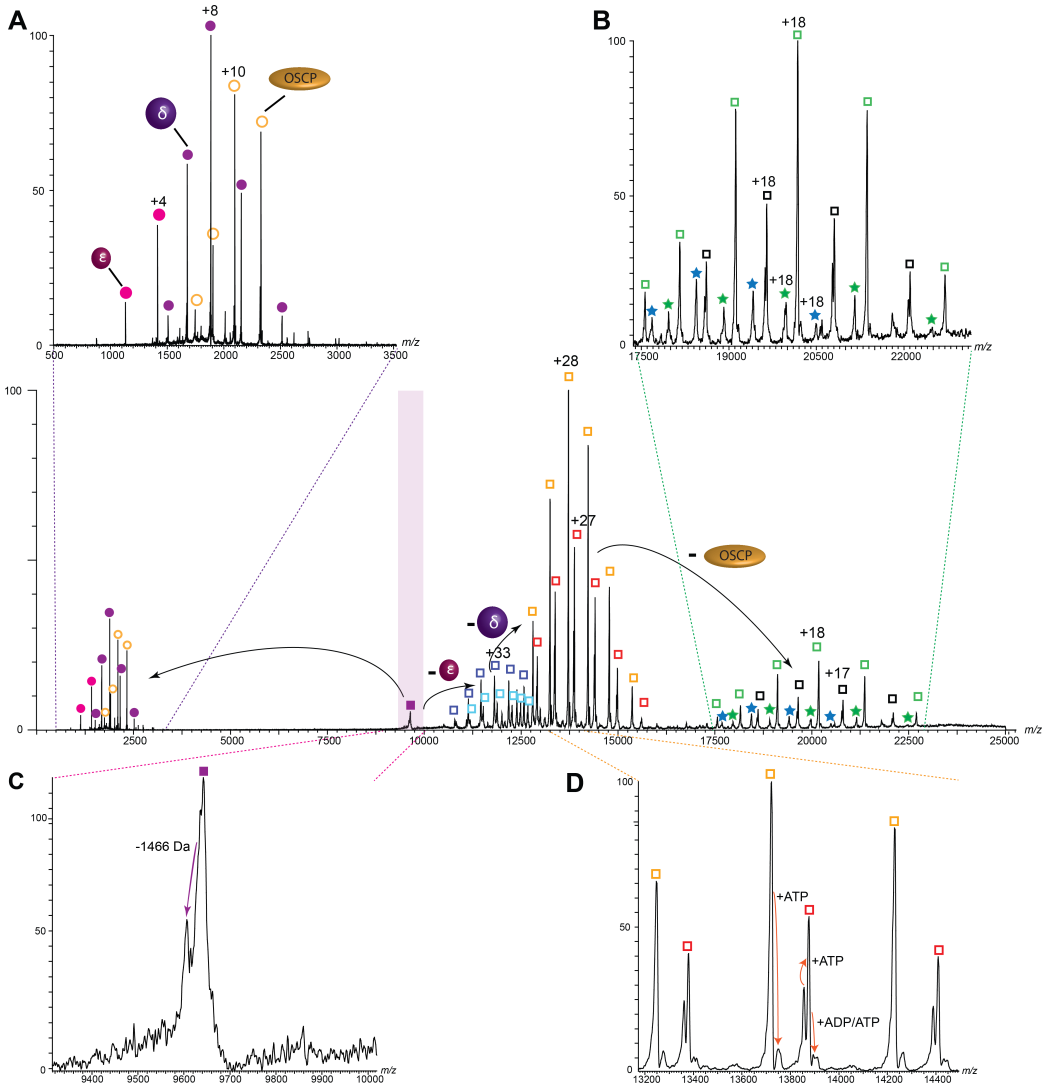


Figure 5.19: Native CID mass spectrum of 405.5 kDa and 395 kDa complexes. The 405.5/395 kDa complex was selected for CID ($m/z = 9642$ Da, purple). (A) Expansion of the low m/z region shows the highly-charged CID products; subunit ϵ (pink circles, 5,656 ±1 Da), subunit δ (purple circles, 15,080 ±1 Da), and subunit *OSCP* (open orange circles, 20,947 ±1 Da). In the 10,500-15,500 m/z region, charge-stripped complexes are formed by loss of ϵ with/without IF_1 (dark/light blue squares, 389,783 ±7 Da/380,235 ±13 Da) and loss of ϵ and δ with/without IF_1 (orange/red squares, 384,139 ±12 Da/374,586 ±7 Da). (B) Expansion of the high m/z region shows loss of ϵ, δ, and *OSCP* with/without IF_1 (green/black squares, 363,202 ±13 Da/353,647 ±7 Da) and loss of γ and *OSCP* with/without IF_1 (blue/green stars, 368,824 ±46 Da/359,453 ±42 Da). (C) Close inspection of the precursor peak reveals loss of a small ≈ 1466 Da molecule, close to the mass of cardiolipin. (D) Comparison of charge-stripped complexes reveals additional peak series arising from nucleotide binding. When IF_1 is bound (orange squares) the major peak is ATP_1-bound and a small ATP_2-bound population is observed. When IF_1 is not bound (red squares), a smaller proportion is ATP-bound with populations of ATP_0, ATP_1, and ATP_2-bound species observed.

Closer inspection of the charge-stripped complexes reveals an additional peak series attributed to ATP binding (Figure 5.19D). The ratio of ATP-bound and ATP-unbound peaks is not uniform between species originating from the 405.5 kDa (IF_1-bound) and 395 kDa (IF_1-unbound) precursors. For the IF_1-unbound species, three populations are detected corresponding to the *apo* form and binding of one or two nucleotides, with ≈30 % in the *apo* form. Conversely, the complexes originating from the 405.5 kDa precursor have a greater proportion nucleotide-bound, with the predominant population being bound to one nucleotide and a smaller population bound to two nucleotides.

Closer inspection of the mass spectrum also reveals a second peak on the CID precursor peak (Figure 5.19C). Calculating the mass difference of the two peaks results in an adduct mass of 1,466 Da. This mass indicates a lipid binding as lipids are known to be important in F_oF_1 function. Therefore, F_oF_1 was subjected to a lipidomics analysis, using LC-MS/MS. Dr Di Wu and Dr Dror Chorev (The University of Oxford, UK) performed the experimental lipidomics procedure. The resulting mass spectrum (Figure 5.20) shows the presence of multiple lipid species with two species being particularly prevalent; at 766.54 Da and 1447.95 Da. A search of the LIPID MAPS database [330] assigned the first of these (766.54 Da) to phosphatidylethanolamine (PE 38:4) and the second (1447.95 Da) to cardiolipin (CDL 72:8), with MS/MS fragmentation spectra indicating these to be PE(20:4/18.0) and CDL(1'-[18:2/18:2],3'-[18:2/18:2]) (Figure 5.21). The cardiolipin species fits well with the mass observed bound to the CID precursor ion (Figure 5.19), indicating that cardiolipin is bound.

Summarising the results above allows the subunit composition of the 405.5 kDa and 395 kDa complexes to be determined: 1) All 17 F_oF_1 subunits and the inhibitor

protein IF_1 were identified by proteomics, 2) Intact subunit masses were determined by denaturing MS for the majority of the subunits, 3) Subunits δ, ϵ, and *OSCP* were ejected from the 405.5 and 395 kDa complexes by CID, 4) The mass difference between the 405.5 kDa and 395 kDa complexes fits that of IF_1, 5) A mass of 1466 Da is bound to the 405.5/395 kDa peak, and 6) The presence of cardiolipin is indicated by lipidomics.

To assign the subunit composition of the complexes detected by native MS, potential subunit compositions were calculated using the iterative search algorithm SUMMIT [331], using the restraints that at least one copy of IF_1, δ, ϵ, *OSCP*, and cardiolipin were present in the 405.5 kDa complex. The 395 kDa complex was therefore assigned as $\alpha_3\beta_3\gamma_1\delta_1\epsilon_1\mathrm{OSCP}_1\mathrm{CDL}_1\mathrm{ADP}_1\mathrm{ATP}_1$ and the 405.5 kDa complex

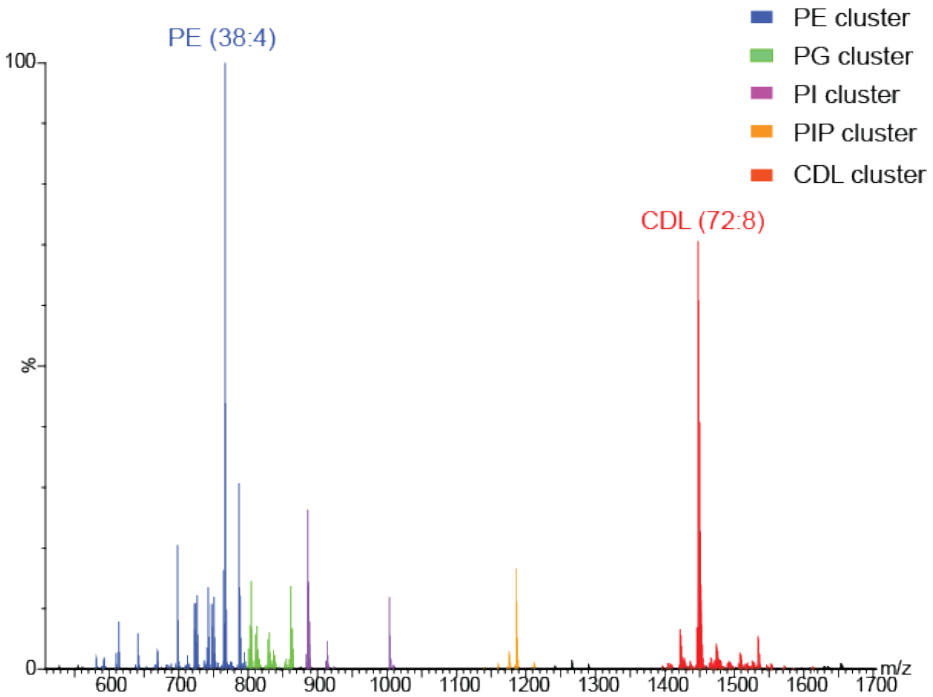


Figure 5.20: Mass spectrum from lipidomics experiment. The two main lipid species are assigned to phosphatidylethanolamine (PE 38:4, blue) and cardiolipin (CDL 72:8, red).

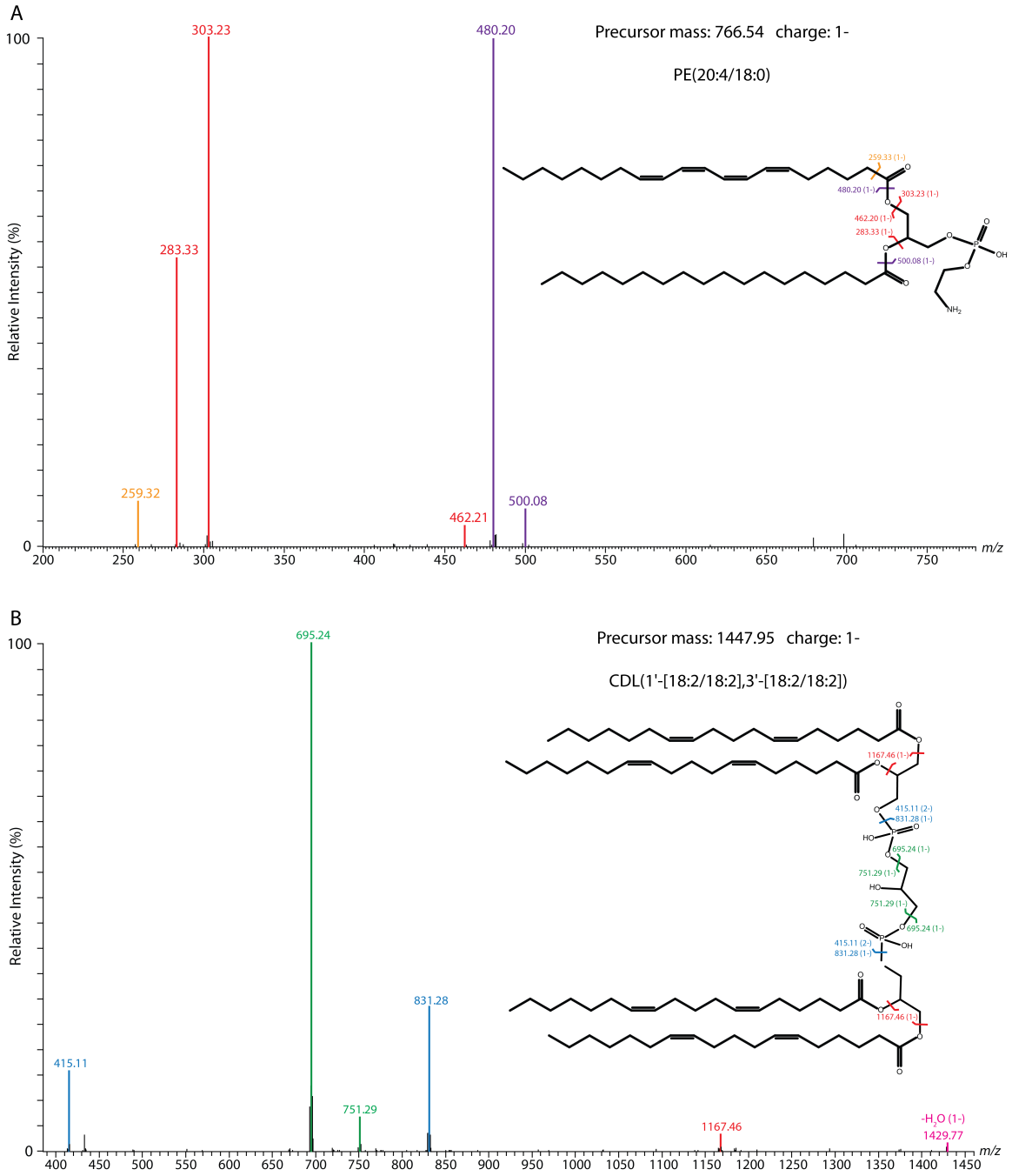


Figure 5.21: MS/MS spectra of lipid species. (A) Fragmentation of 766.54 *m/z* precursor ion is consistent with the phosphoethanolamine species PE(20:4/18:0). (B) Fragmentation of 1447.95 *m/z* precursor ion is consistent with the cardiolipin species CDL(1'-[18:2/18:2],3'-[18:2/18:2]).

assigned as the addition of monomeric IF_1 to the 395 kDa complex. The binding of monomeric IF_1 is surprising as the active bovine IF_1 is thought to bind as a dimer to F_1, in contrast to the yeast IF_1 which binds as a monomer.

5.6.1.4 Cross-Linking Reveals Subunit Contacts of the Intermembrane Space Domain and the IF_1 Inhibitor

The next step was to identify the location of IF_1 binding to F_oF_1 and to determine if monomeric or dimeric IF_1 binds. Furthermore, having placed the identified PTMs on the available structure of F_oF_1, it was clear that some of the PTMs were located on subunits without available structures or with unknown placement within the F_oF_1, including those which are candidates to be constituents of the intermembrane space domain (*e*, *f*, *g*, *A6L*, *DAPIT*, *6.8PL*). In order to gain insights into the location of these subunits and IF_1 and their organisation within the complex, F_oF_1 was chemically cross-linked. Two cross-linking reagents were used; BS3 and DSS in order to probe both soluble and more hydrophobic parts of the complex. After cross-linking, the complex was separated by SDS-PAGE and digested. Trypsin and chymotrypsin enzymes were both employed to digest cross-linked proteins as tryspin is highly specific but cleaves at lysine and arginine residues which are often scarce in membrane proteins, whereas chymotrypsin is able to better cleave the more hydrophobic regions of the protein. Cross-linked peptides were analysed by LC-MS/MS and data searched with pLink software [102], before manual validation of spectra. After this analysis 117 cross-links were identified of which 50 were inter-subunit cross-links (Figure 5.22, Appendix Tables C.4 and C.5).

Amongst the cross-links identified were 23 cross-links involving the supernumerary and mammalian-specific subunits *e*, *g*, *DAPIT*, and *6.8PL* (Figure 5.23). Cross-links were identified between subunit *A6L* and subunits *d* and F_6, which would place *A6L*

on the matrix side of the membrane, with extensive contact with the peripheral stalk. Subunit *b1*, another constituent of the peripheral stalk cross-links to subunit *g*, one of the supernumerary subunits, which also cross-links to subunit *e*. The location of subunit *b1*, with its transmembrane helices, places subunit *g* in the membrane intrinsic region. Subunit *DAPIT* cross-links to itself, and also with subunit *6.8PL*. These cross-links indicate that *DAPIT*, while being located in close proximity to *6.8PL*, does not appear to be near other residues available for cross-linking. In addition to *DAPIT*, *6.8PL* cross-links to the N-terminus of

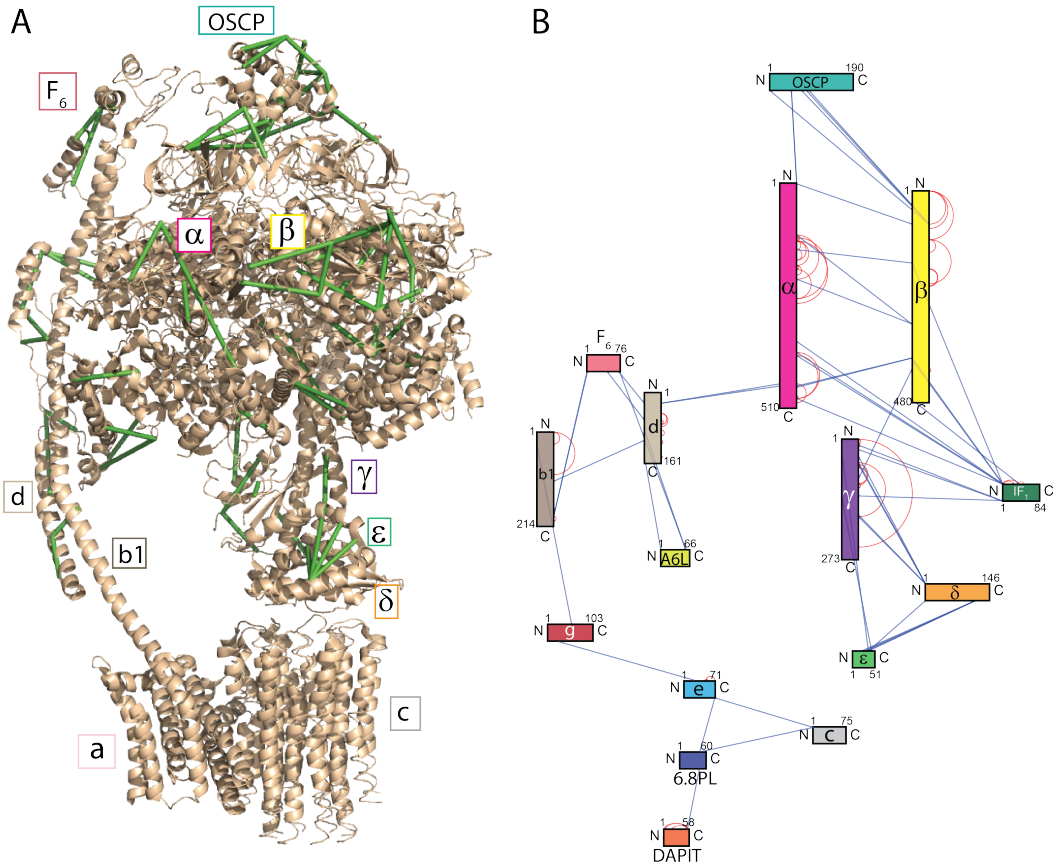


Figure 5.22: BS3 and DSS cross-links of the bovine F-type ATP synthase. (A) Cross-links (green) are displayed on the cryo-EM structure of bovine F_oF_1 (PDB ID: 5ARA [278]). (B) Cross-links are visualised as a network. Protein subunits are indicated by bars representing the amino acid sequence, with lines indicating intra-protein (red) and inter-protein (blue) cross-links. Figure produced with xVis software [247].

A

Protein 1	Residue 1	Protein 2	Residue 2	Peptide 1	Peptide 2	BS3/T	DSS/T	DSS/C
b1	28	*g*	55	LYPK*TGVTGPY	K*IINSAK			x
d	85	*A6L*	57	VPIPEDK*YTAQVDAEEKEDVK	WTK*IYLPLLLPL	x		
d	144	*A6L*	1	NIIPFDQMTIEDLNEVFPETK*L	M*PQL			x
e	71	*c*	1	ELAEAQEDTILK*	D*IDTAAK		x	
e	48	*e*	55	RLAAEEK*K	DEQK*R	x		
e	49	*e*	55	LAAEEKK*KR	RDEQK*R	x		
e	50	*e*	48	K*RDEQKR	RLAAEEK*K	x		
e	50	*e*	48	K*RDEQK	LAAEEK*K			x
e	55	*e*	48	KRDEQK*R	LAAEEK*K	x		
e	55	*e*	48	DEQK*RIER	LAAEEK*K		x	
e	71	*e*	50	ELAEAQEDTILK*	K*RDEQKR	x		
e	71	*6.8PL*	49	ELAEAQEDTILK*	K*ASSAAPAHGHH			x
e	71	*6.8PL*	49	ELAEAQEDTILK*	ALK*ASSAAPAHGHH		x	
g	24	*e*	2	APALVNAAVTYSK*PR	V*PPVQVSPLIK	x		
A6L	43	*d*	117	HNFYHNPELTPTK*MLK	IQEYEK*ELEK	x		
A6L	46	*d*	25	MLK*QNTPWETK	NQK*AVANSLK	x	x	
A6L	46	*F6*	73	MLK*QNTPWETK	FEVVEK*PQS	x		
A6L	43	*A6L*	46	HNPELTPTK*ML	K*QNTPWETK			x
6.8PL	49	*c*	1	ALK*ASSAAPAHGHH	D*IDTAAK	x		
6.8PL	49	*c*	1	K*ASSAAPAHGHH	D*IDTAAKF			x
6.8PL	49	*DAPIT*	56	ALK*ASSAAPAHGHH	TPAVK*AT			x
DAPIT	2	*DAPIT*	17	A*GPEADAQFHFTGIK	K*YFNSYTLTGR	x	x	
DAPIT	2	*DAPIT*	56	A*GPEADAQFHFTGIKK	TPAVK*AT		x	
DAPIT	17	*DAPIT*	56	K*YFNSYTLTGR	TPAVK*AT	x		

B

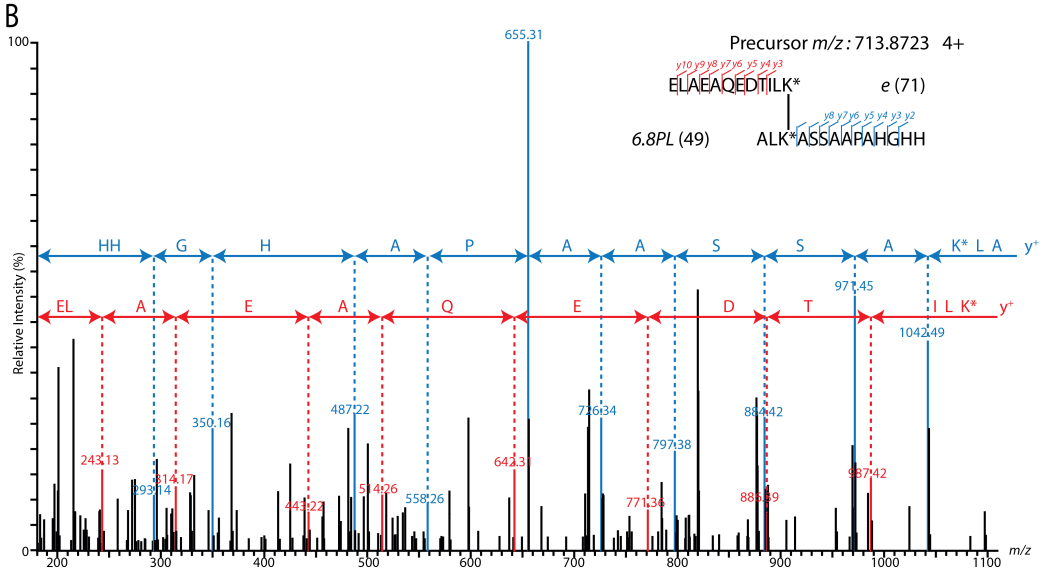


Figure 5.23: Cross-linking of membrane subunits of bovine mitochondrial F-type ATP synthase. (A) Table of identified BS3 and DSS cross-linked peptides of protein subunits that are candidates for inclusion in the IMD. The cross-linked protein subunits, cross-linked residues, and peptide sequences are given. The experimental conditions (cross-linking reagent and digest enzyme) under which the cross-links were identified are indicated; BS3 cross-linker and trypsin (BS3/T), DSS cross-linker and trypsin (DSS/T), or DSS cross-linker and chymotrypsin (DSS/C). (B) An example MS/MS spectrum of cross-linked peptides from subunits *e* and *6.8PL*, with b and y ions indicated (red and blue, respectively).

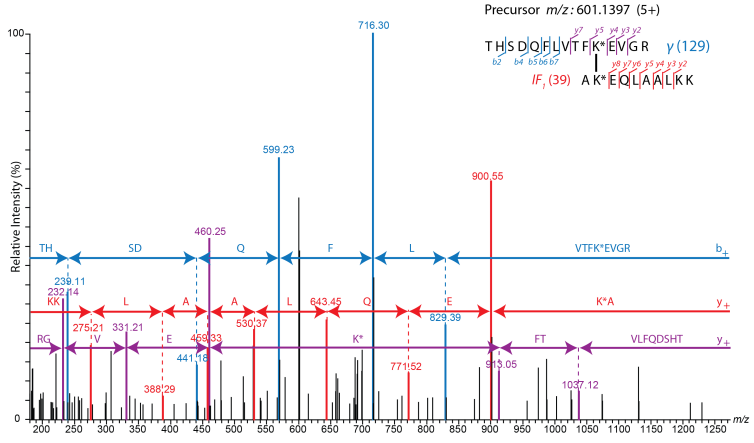
subunit *c*, located on the lower surface of the ATP synthase. This indicates a location for *6.8PL* at the lower solvent-accessible side of the membrane, near to the IMD, with an additional cross-link between *6.8PL* and subunit *e* identified. The location of subunit *e* is somewhat more difficult to place on account of the multiple cross-links identified, placing it close to *6.8PL* with also a single cross-link identified between *c* and *e*, suggesting a location on the periphery of the membrane. Together these cross-links give a picture of the localisation of these different subunits and suggest that it is subunits *6.8PL* and *DAPIT* that make up the IMD, with subunit *e* in close proximity, contacting subunit *6.8PL* but not *DAPIT*.

In addition to information on the identity of the IMD, multiple other cross-links were also identified. Of particular interest were the cross-links identified between IF_1 and other subunits; namely subunits α, β, and γ (Figure 5.24). In order to display these cross-links on the F_1 structure, a structure from X-ray crystallography of the bovine $\alpha_3\beta_3$ hexameric head and γ in complex with a truncated IF_1 inhibitor was used (PDB ID: 1OHH [316]). As many of the cross-links identified were at the N-terminus of IF_1 and absent in the crystal structure, a homology model of IF_1 was generated using the Phyre2 web server [237] and aligned to the X-ray structure using Pymol. Mapping the identified cross-links on the structure of the F_1 domain (Figure 5.24C) shows that the cross-links can all be accommodated on this structure, consistent with IF_1 binding through the edge of the F_1 head. Interestingly, no cross-links indicating an IF_1 dimer were observed which is unexpected on account of the proposed binding of IF_1 as a dimer unit. One would expect inter-subunit IF_1 to IF_1 cross-links if this was the case and yet none was identified; moreover no cross-links were observed on the C-terminal half of IF_1, indicating that no other protein residues available for cross-linking were in close proximity. These results therefore suggest that IF_1 binds to the F_1 as a monomer.

A

Protein 1	Residue 1	Protein 2	Residue 2	Distance/ Å	Peptide 1	Peptide 2	BS3/T	DSS/T	DSS/C
α	359	*IF1*	1	23.3	K*GIRPAINVGL	G*SESGDNVR			x
α	384	*IF1*	1	40.9	AMK*QVAGTMK	G*SESGDNVR	x		
α	391	*IF1*	1	31.8	AMKQVAGTMK*LELAQYR	G*SESGDNVR	x		
α	391	*IF1*	1	31.8	QVAGTMK*LELAQYR	G*SESGDNVR		x	
α	488	*IF1*	1	49.0	IRTDGK*ISEESDAK	G*SESGDNVR	x		
β	164	*IF1*	1	29.9	IGLFGGAGVGK*TVLIMELINNVAK	G*SESGDNVR	x	x	
β	379	*IF1*	1	25.1	GVQK*ILQDYK	G*SESGDNVR		x	
β	384	*IF1*	1	15.9	ILQDYK*SLQDIIAILGMDELSEEDKLTVSR	G*SESGDNVR	x		
β	474	*IF1*	39	13.1	ADK*LAEEHS	AK*EQLAALK	x		
γ	14	*IF1*	1	6.8	LKSIK*NIQK	G*SESGDNVR		x	
γ	24	*IF1*	1	20.9	SMK*MVAAAK	G*SESGDNVR	x	x	
γ	129	*IF1*	39	23.0	THSDQFLVTFK*EVGR	AK*EQLAALKK	x	x	
IF1	1	*IF1*	24	13.1	G*SESGDNVR	DAGGAFGK*R	x	x	
IF1	24	*IF1*	1	13.1	DAGGAFGK*REQAEEERYFR	G*SESGDNVR		x	
IF1	39	*IF1*	1	33.0	AK*EQLAALKK	G*SESGDNVR		x	
IF1	47	*IF1*	39	12.3	K*HHENEISHHAK	AK*EQLAALKK	x	x	x

B



C



Figure 5.24: Identified cross-links between the inhibitor IF_1 with subunits α, β, γ, and IF_1. (A) The cross-linked protein subunits, cross-linked residues, peptide sequences, and measured Cα–Cα distances are given. The experimental conditions (cross-linking reagent and digest enzyme) under which the cross-links were identified are indicated; BS3 cross-linker and trypsin (BS3/T), DSS cross-linker and trypsin (DSS/T), or DSS cross-linker and chymotrypsin (DSS/C). (B) Example MS/MS spectrum for identification of the BS3 cross-link γ(K129)–IF_1(K39). Assigned b (blue) and y (red and purple) ions are highlighted. (C) Green lines indicate identified cross-links between the bovine soluble F_1 domain and IF_1. Cross-links are identified from IF_1 (brown) to subunits α (blue), β (purple) and γ (gold). The structure of IF_1, modelled with the Phyre2 web server [237], is shown aligned to the IF_1-bound F_1 X-ray crystallography structure (PDB ID: 1OHH [316]), with a single copy each of subunits α and β shown for simplicity.

5.6.2 Discussion

Although numerous PTMs were identified across 14 subunits of the bovine F-type ATP synthase, in most cases the subunit masses determined by denaturing MS could be explained by NME and loss of transit peptides, indicating that many of the identified PTMs are substoichiometric. It is of particular interest how prevalent PTMs are in the ATP synthase enzymes, suggesting a high level of regulation and control imparted by PTMs on ATP synthesis. The importance of PTMs of mammalian mitochondrial ATP synthase from heart tissue in heart disease and diabetes has been recognised and with cardiovascular complications being a major global killer [332], understanding of these PTMs is imperative.

In some cases, expected PTMs were not identified on some subunits, for example the expected trimethylation of Lys43 on subunit *c* [333]. Subunit *c* had low sequence coverage in the proteomics experiments and this is most likely due to low enzymatic digest efficiency on account of the highly-hydrophobic nature of subunit *c*, an issue which has previously been highlighted in proteomics studies of ATP synthases [334]. The mass of subunit *c* determined from the denaturing MS analysis was 7650 Da, 42 Da higher than the sequence mass and corresponds to trimethylation of Lys43. A second less intense species is also detected in the spectrum at 7692 Da, 42 Da larger again. It is possible that this could be trimethylation of Lys7 or acetylation of Lys7, Arg38, or the N-terminus. Modifications to one or more lysine and arginine residues would further rationalise the difficulty in detecting subunit *c* with bottom-up proteomics as tryptic digest efficiency would be reduced by uncleavable modified residues.

The IF_1 inhibitor was co-purified with ATP synthase and identified by proteomics and denaturing MS. The mass of the major IF_1 species detected by denaturing MS

corresponded to intact IF_1 with loss of the 25 amino acid residue transit peptide. Minor species were detected with sequential loss of the first five N-terminal residues. No other protein subunits exhibited this behaviour in the denaturing MS analysis, indicating that loss of these N-terminal amino acids does not occur during analysis. As IF_1 is present in this sample on account of its binding to F_1, this means that loss of the first five amino acids does not appear to prevent IF_1 binding to IF_1 and likely occurs *in vivo*. *In vivo*, the N-terminus of IF_1 is disordered, gaining its α-helix structure upon binding to F_1, and therefore the remaining 55 amino acids comprising the active inhibitory protein must be capable of binding to IF_1 in the absence of residues 1–5. Intrinsically-disordered regions of proteins are known to be more susceptible to proteases [335], and therefore the N-terminal amino acid loss observed here may be due to this, or to low specificity in transit peptide cleavage. The pH-dependent oligomeric state of the IF_1 inhibitor is key to its function, yet there has been uncertainty of whether the active state is monomeric or dimeric. IF_1 was found to be bound to the F_1 domain by native MS and cross-linking placed it in the $\alpha_3\beta_3$ head, contacting subunit γ. Importantly, it was monomeric IF_1 that was observed binding by native MS and no cross-links were observed on the C-terminal half of IF_1, implying that no other protein is nearby. This does not support the model of an active IF_1 dimer, with each IF_1 binding to separate ATP synthase F_1 domains.

Comparison of the ATP-bound and ATP-unbound states of complexes deriving from the F_1 and F_1–IF_1 precursors revealed an increase in ATP-bound species when IF_1 is bound to the complex, contradicting the proposed IF_1 binding mechanism in which hydrolysis of two ATP molecules is required. IF_1 binds between subunit α and β, preventing rotation of $\alpha_3\beta_3$ relative to γ and the F_o, halting ATP synthesis. The increase in one ATP bound to the complex when IF_1 is present could be on

account of IF_1 trapping ADP + P_i in the catalytic site of the β subunit when binding, although it is difficult to distinguish the location of nucleotide in either the catalytic or non-catalytic sites.

CID of the F_1 complex revealed a bound adduct $\approx$ 1466 Da. This mass was thought to be cardiolipin, a lipid known to be important for the function of ATP synthase [336] and lipidomics indicated the presence of cardiolipin (1448 Da) along with the presence of a second lipid, PE (767 Da). Disrupting the synthesis of mitochondrial PE has been shown to alter mammalian mitochondrial morphology and proved lethal to embryos [337]. While further studies have identified that PE deficiency impairs oxidative phosphorylation but does not reduce the activity of mitochondrial ATP synthase [338]. The role of PE on ATP synthase is therefore far from clear, however it is likely the two interact closely as PE is a constituent of biological membranes [339].

The role of cardiolipin on ATP synthase has been studied more extensively. In computer simulations, the conserved trimethylated Lys43 position selectively associates with cardiolipin, as does the second lysine residue Lys7 located at the opposite end of the *c*-ring [340]. Furthermore it has been proposed that negatively-charged cardiolipin bound to trimethylated Lys43 near the proton exit half-channel could act as a proton-acceptor [340, 341], preventing build-up of positive charge at the half-channel exit. It was surprising that cardiolipin was found to bind to the F_1 complex, without the *c*-ring present. The exact location of cardiolipin binding is unclear, yet the most-likely candidates would be subunits in close proximity to the membrane and *c*-ring including subunits γ, δ, and ϵ. This could be the result of close contact between the cardiolipin residing in the *c*-ring lumen and the lower surface of one of these three subunits.

The location of the small supernumerary subunits has remained uncertain for some time now, with their function also unclear. In addition, the identification of an additional density by cryo-EM [278, 286], termed the IMD, is currently unassigned. Cross-linking analysis placed subunit *A6L* in close proximity to subunits *d* and F_6 of the peripheral stalk, indicating a location on the intermembrane side. A recent cryo-EM based structure of the isolated F_o domain from *S. cerevisiae* published by Guo *et al.* (PDB ID: 6B2Z [273]) since this cross-linking analysis was performed, shows that the *A6L* homologue (subunit *8*) is closely associated with subunit *d* and serves to anchor the peripheral stalk to the membrane F_o domain. Furthermore homologues of subunit *f* and *g* are both located in the vicinity of the stalk anchoring region, near to subunits *b1* and *e* as identified by the cross-linking data. Subunit *e* in the yeast enzyme is longer than in the bovine enzyme, yet its location in the cryo-EM density map is particularly interesting, extending out of the membrane into the intermembrane space. The cross-linking data suggest that subunit *e* in bovine is located in a similar place as in the yeast enzyme with a cross-link between the N-terminus of subunit *c* and Lys71 of subunit *e* placing subunit *e* in the intermembrane space. Yet, the size of subunit *e* on its own, particularly in the bovine enzyme where subunit *e* is only 71 amino acids, cannot account for the IMD as visualised by cryo-EM. The cross-linking data however further place subunit *6.8PL*, absent in yeast, in close proximity to both subunits *e* and *c*. In addition, cross-linking places *DAPIT* in close proximity to *6.8PL* with no cross-links to other subunits identified. Interestingly, Guo *et al.* found that the *b*-*e*-*g* subcomplex was required to induce the membrane-bending kink, yet in the bovine enzyme the absence of the IMD does not affect the membrane-bending kink, conflicting with an IMD comprised of subunits *e*, *f*, and *g*. It appears therefore from these studies combined with the cross-linking data identified here that subunits *DAPIT* and *6.8PL* are located between subunits *e* and *c* and thus remain

strong candidates for contributing to the IMD density observed by cryo-EM of the bovine F_oF_1.

5.7 Conclusions

The research presented in this chapter has examined various aspects of structure, interactions, and regulation of ATP synthases. First the V-type ATP synthase from the yeast *S. cerevisiae* was studied. Proteomics analysis of V_oV_1 containing the Stv1p isoform of subunit *a* revealed the presence of the same two subunits, *e* and *f*, known to associate with the vacuolar enzyme via the Vph1p isoform of subunit *a*. This was surprising on account of the differences in Stv1p and Vph1p subunits and leads to the question of whether subunits *e* and *f* have similar binding sites to Stv1p as they do in Vph1p. As the location of subunit *e* and *f* binding sites on Vph1p have been uncovered through cryo-EM analysis of the Vph1p-V_o domain, cryo-EM of the Stv1p-V_o domain would be the next step in determining differences between the two enzymes.

Next, a potential switch for starvation-induced autoinhibition by V_1/V_o dissociation was investigated. It was thought that a phosphorylation site at Ser677 could be the control point for this inhibitory mechanism due to its conservation and location within the enzyme. Proteomics analysis showed that while Ser677 was phosphorylated in intact V_oV_1, it was also phopshorylated in the dissociation-inhibited V_o domain. Furthermore, comparison of all phosphorylation sites identified in the two samples did not indicate that phosphorylation of subunit *a* was responsible for dissociation. The switch for dissociation is therefore still unknown, yet phosphorylation may still be the key. To further explore this, denatured MS of all V-type ATPase subunits in the intact and dissociation-inhibited

complexes could be used to assess the global PTM status of each subunit. Together with bottom-up proteomics analysis of all subunits, any discrepancies in PTM status between the intact and dissociated complexes could be identified, potentially allowing identification of a PTM responsible for V-type ATPase autoinhibition.

Having considered the role of phosphorylation in regulation of the yeast V-type ATPase, the effect of PTMs on the F-type ATP synthase from spinach chloroplasts was investigated. By utilising comparative cross-linking to assess changes in protein-protein interactions after deacetylation, a previously identified conformational change on subunit ϵ was shown to be sensitive to acetylation, with deacetylation favouring the inhibition-competent conformation. Furthermore, when deacetylation was combined with dephosphorylation, there was significant disruption to protein-protein interactions reflecting the importance of these two PTMs in complex stability. Further study is needed to determine the exact acetylation sites which affect the inhibiting conformation of subunit ϵ. Site-directed mutagenesis of identified acetylation sites may assist in this regard, while investigating the activity of acetylase/deacetylase enzymes in response to environmental pressures may yield useful insights into the role of acetylation in F_oF_1 regulation.

The final ATP synthase to be investigated in this chapter was the mammalian F-type ATP synthase from bovine heart mitochondria. The ATP synthase enzyme was found to be highly modified by phosphorylation, acetylation, and methylation suggesting a sophisticated network of regulatory interactions present in the enzyme. The role of PTMs is already the target of intensive study, particularly due to its implication in heart disease, and the PTMs identified here will contribute towards this field by providing targets for further study. More work is still needed to further understand the role of PTMs. One approach could be comparative cross-linking, as

utilised for the cATP synthase from *S. oloracea*, to investigate the effect of *in vitro* dephosphorylation and deacetylation on protein-protein interactions, potentially identifying PTM-controlled conformational changes.

Native MS revealed the presence of high mass species attributed to the dissociated F_1 head and a dimer lacking the F_1 head, indicating possible stronger interactions at the dimer interface than the F_1-F_o interface. Chemical cross-linking provided insight into the organisation of the small membrane subunits, particularly the supernumerary subunits *e*, *g*, *6.8PL*, and *DAPIT*. Subunit *A6L* was found to be located in close proximity to the peripheral stalk, as was subunit *g*. Subunit *e* extends to the intermembrane space, cross-linking with the N-terminus of subunit *c*. Subunits *6.8PL* and *DAPIT*, for which little is known, were found to be located in the intermembrane space, with cross-linking locating them in the vicinity of the unassigned IMD. Subunits *6.8PL* and *DAPIT*, in close association with subunit *e*, are therefore proposed to constitute the IMD. Quantification of subunits DAPIT and 6.8PL relative to the rest of the ATP synthase complex through the use of proteomics and isotopically-labelled peptides would provide further evidence to either support or contradict their proposed placement in the IMD.

Proteomics analysis of the purified ATP synthase revealed the presence of the IF_1 inhibitor with native MS revealing a high proportion of IF_1 bound to the complex. Comparison of IF_1-bound and -unbound species revealed discrepancies in nucleotide binding with the IF_1-bound state retaining a higher level of nucleotide, indicating IF_1 locks the $\alpha_3\beta_3$ head in place, relative to γ, and prevents nucleotide release. Furthermore, the IF_1 was observed to bind with a stoichiometry of one, as opposed to the expected two, indicating that the active state of bovine IF_1 is monomeric. Cross-linking further supported this monomeric IF_1 interaction, with all cross-links

accommodated by a monomeric IF_1-F_1 structure and no IF_1 dimer interactions detected. To better understand IF_1 binding, native MS instrument parameters could be optimised in an attempt to allow detection of intact F_oF_1 dimers, as the stoichiometry of IF_1, together with the distribution of IF_1 between monomeric and dimeric F_oF_1 would clarify the role, if any, of IF_1 in promoting F_oF_1 dimerisation, as proposed previously [342].

5.8 Methods

5.8.1 Protein Purification

5.8.1.1 Vacuolar and Golgi V-type ATPases from *S. cerevisiae*

Purified V-type ATPase samples from *S. cerevisiae* were provided by Stephanie Bueler, Thamiya Vasanthakumar, and Professor John Rubinstein (The University of Toronto, Canada) as excised gel bands, purified as follows. Intact V_oV_1 used for phosphosite analysis was purified from yeast strain SABY31 as described [291]. Yeast cells were grown in YPD medium in an 11 L BioFlo fermentor (New Brunswick Scientific) until an OD_{600} of 0.8–1.2 was reached. Cells were harvested by centrifugation at 4,000 × g for 15 minutes and lysed with a bead beater in lysis buffer (PBS (pH 7.4), 8 % (w/v) sucrose, 2 % (w/v) sorbitol, 2 % (w/v) glucose, 5 mM EDTA, 0.001 % (w/v) phenylmethylsulfonyl fluoride (PMSF), 5 mM ϵ-aminocaproic acid, and 5 mM p-aminobenzoic acid). The cell lysate was centrifuged at 3,000 × g for 10 minutes and membranes harvested at 38,000 × g for 40 minutes. The harvested membranes were resuspended in lysis buffer and supplemented with 1 % n-Dodecyl β-D-maltoside (DDM) followed by centrifugation at 42,000 × g. The V_oV_1 complex, located in the supernatant, was loaded on an M2-Sepharose column (Sigma) equilibrated with buffer (50 mM Tris-HCl (pH 7.4), 150 mM NaCl, 0.03 % (w/v) DDM) and eluted with buffer supplemented with 150 μg mL^{-1} 3×FLAG peptide (Sigma).

The vacuolar form of the V_o complex for phosphosite identification was purified from yeast strain CACY1, which encodes Vph1p (subunit *a*) with a C-terminal 3×FLAG tag [14]. Yeast was grown in an 11 L BioFlo fermenter (New Brunswick Scientific) for 60 hours, with the extended growth time used to starve the yeast of nutrients for dissociation of the V_o and V_1 subcomplexes. The V_o complex was

solubilised with DDM and purified on an M2-Sepharose column (Sigma) in elution buffer (50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.02 % (w/v) DDM, 50 μg mL^{-1} 3×FLAG peptide), as described [14, 291]. Fractions containing the V_o complex were pooled and concentrated to 2.5 mg mL^{-1} using an Amicon concentrator (100 kDa MWCO).

The Golgi-form of the V_o complex was purified from yeast strain SABY67, which includes STV1 with a C-terminal 3×FLAG tag and VPH1 knockout, in the background BY4741 strain. Yeast was grown in 10 L of YPD medium for 48 hours. Cells were harvested by centrifugation and then lysed using a bead beater. The membranes were harvested and solubilised with 1 % DDM. The V_o complex was purified by FLAG-affinity chromatography with M2 sepharose anti-FLAG resin (Sigma) in a final buffer (50 mM Tris-HCl (pH 7.4), 150 mM NaCl, 0.025 % (w/v) DDM). The sample was further purified by SEC on a Superdex 200 column (GE Healthcare). V_o-containing fractions were concentrated and proteins separated by SDS-PAGE.

5.8.1.2 Chloroplast F-type ATP Synthase from *S. oleracea*

Purified cATP synthase from *S. oleracea* was provided by Professor Carla Schmidt (The Martin Luther University of Halle-Wittenberg), purified as follows. The cATP synthase was purified as described previously [99, 343]. Fault-free leaves from *S. oleracea* were stored in the dark overnight at 4 ^{o}C to reduce starch content and then homogenised in lysis buffer (0.4 M sucrose, 100 mM Tricine-NaOH (pH 8.0), 2 mM $MgCl_2$, 1 mM PMSF, supplemented with Complete Protease Inhibitor Tablets (Roche)). The homogenised leaves were filtered through four layers of synthetic nylon and centrifuged at 21,590 × g for 25 minutes at 4 ^{o}C. The centrifuged pellet, containing chloroplasts and thylakoid membranes, was

resuspended in buffer (10 mM Tris-HCl (pH 8.0), 0.5 mM $MgCl_2$, 0.1 mM PMSF) and stirred on ice for 15 minutes. The resuspension was centrifuged at 21,595 × g for 25 minutes and the pellet washed with buffer (0.4 M sucrose, 10 mM Tris-HCl (pH 8.0), 150 mM NaCl, 0.5 mM $MgCl_2$, 0.1 mM PMSF) before flash-freezing in liquid N_2.

The pellet was thawed on ice and resuspended in resuspension buffer (0.4 M sucrose, 50 mM Tricine-NaOH (pH 8.0), 2 mM $MgCl_2$). The chlorophyll concentration was determined as described [344] and resuspension buffer added to adjust to a final chlorophyll concentration of 5 mg mL^{-1}. DTT was added to a final concentration of 50 mM and the suspension stirred on ice for 15 minutes. The suspension was diluted 1:1 with extraction buffer (20 mM Tricine-NaOH (pH 8.0), 200 mM sucrose, 5 mM $MgCl_2$, 400 mM $(NH_4)_2SO_4$, 2 mM Na_2-ATP, 25 mM sodium cholate, 60 mM β-D-octylglucoside, 50 mM DTT) and stirred on ice for 30 minutes. The suspension was centrifuged at 218,000 × g for 60 minutes. The supernatant was supplemented with $(NH_4)_2SO_4$ to a final concentration of 1.2 M and centrifuged at 10,000 × g for 15 minutes. The cATP synthase complex was precipitated with $(NH_4)_2SO_4$ at a final concentration of 1.8 M and collected by centrifugation at 10,000 × g for 15 minutes. The pellet was resuspended in buffer (30 mM NaH_2PO_4-NaOH (pH 7.2), 200 mM sucrose, 2 mM $MgCl_2$, 0.5 mM Na_2-EDTA, 4 mM DDM) and the suspension flash-frozen in liquid N_2.

The sample was applied to a sucrose gradient (30 mM NaH_2PO_4-NaOH (pH 7.2), 2 mM $MgCl_2$, 0.5 mM Na_2-EDTA, 4 mM DDM, 12–30 % sucrose) and centrifuged at 182,000 × g for 22 hours. Fractions containing cATP synthase, as determined by SDS-PAGE, were pooled and purified by SEC on a Superdex 200 column (GE Healthcare) equilbrated in SEC buffer (30 mM NaH_2PO_4-NaOH (pH 7.2),

2 mM $MgCl_2$, 0.5 mM Na_2-EDTA, 4 mM DDM). Fractions containing cATP synthase were pooled and flash-frozen in liquid N_2.

5.8.1.3 Mitochondrial F-type ATP synthase from *B. taurus*

Purified mitochondrial F-type ATP synthase from *B. taurus* was provided by Professor Christoph Gerle (The University of Osaka, Japan), purified as follows. Mitochondrial F-type ATP synthase was purified from *B. taurus* as previously described [345, 346] with modifications for LMNG solubilisation [347]. Inner mitochondrial membranes were extracted from heart muscle from *B. taurus* as previously described [348]. Fat and connective tissue were removed from heart muscle and minced. Minced heart was homogenised in 23 mM sodium phosphate buffer (pH 7.4), at 0 ^{o}C for 5 minutes at 13,000 rpm in a homogeniser (Nihon Seiki). The homogenised material was centrifuged at 2,800 rpm for 20 minutes in a refrigerated Kubota Model 9810 centrifuge, using an RS-6600 rotor. The precipitate was resuspended in 22.2 mM sodium phosphate buffer (pH 7.4), and homogenised as described. The supernatants were pooled and centrifuged at 10,000 rpm for 30 minutes in a refrigerated Beckman Model Avanti HP-30I centrifuge using a JLA-10.500 rotor. The precipitate was resuspended in 50 mM Tris-HCl buffer (pH 8.0) and centrifuged at 30,000 rpm for 30 minutes in a Beckman Model-7 ultracentrifuge, with a 45 Ti rotor. The precipitate was resuspended in buffer (50 mM Tris-HCl (pH 8.0), 660 mM sucrose) to a final protein concentration of 23 mg mL^{-1}.

Sub-mitochondrial particles were solubilised in buffer (40 mM HEPES (pH 7.8), 2 mM $MgCl_2$, 0.1 mM EDTA, 0.1 mM DTT) anad supplemented with 0.7 % (w/v) deoxycholate and 0.4 % (w/v) decyl maltoside (DM). The solubilised particles were centrifuged at 176,000 $\times$ g for 50 minutes. ATP synthase was purified

from the supernatant through a sucrose gradient (40 mM HEPES (pH 7.8), 0.1 mM EDTA, 0.1 mM DTT, 0.2 % (w/v) DM and steps of 2.0 M, 1.1 M, 1.0 M, or 0.9 M sucrose) at 176,000 × g for 15 hours 30 minutes. Fractions were tested for ATPase activity using an ATP-regenerating enzyme-coupled assay [349]. The fractions exhibiting activity were pooled and the detergent exchanged to LMNG using a Poros-20HQ ion-exchange column with a double gradient of 0.2 %–0 % DM and 0 %–0.05 % LMNG over 1 hour 20 minutes at a flow rate of 1 mL min^{-1}. ATP synthase was eluted in buffer (40 mM HEPES (pH 7.8), 150 mM sucrose, 2 mM $MgCl_2$, 0.1 mM EDTA, 0.1 mM DTT, 0.05 % (w/v) LMNG) with a linear KCl gradient of 0–240 mM KCl. Eluted fractions were tested for native phospholipids using ammonium molybdenate complexation, and the fractions containing high amounts of native phospholipids were flash-frozen in liquid N_2.

5.8.2 Proteomics Analysis

For protein identification and PTM analysis, proteins were digested with trypsin and/or chymotrypsin as described [182]. To aid in the identification of phosphosites, phosphopeptides were enriched with TiO_2-affinity chromatography as described [92]. Peptides were dried in a Speedvac and analyzed by LC coupled to a Q Exactive Plus Orbitrap, as described previously (Methods, Chapter 4).

Data were searched against the UniProt database (taxonomy filter *S cerevisiae* or *B. taurus*, as appropriate) using the Mascot v2.4.1 search engine (Matrix Science). The enzyme was set to trypsin, chymotrypsin, or trypsin/chymotrypsin as appropriate with up to two missed cleavage sites. For protein identification, the following parameters were used: peptide tolerance of 7 ppm and MS/MS tolerance of 0.1 Da; fixed modification of carbamidomethylation of cysteines; variable modification of oxidation of methionines. In addition, for phosphosite identification,

variable modifications of phosphorylation of serines, threonines and tyrosines were additionally allowed. For PTM analysis of bovine ATP synthase, additional variable modifications of acetylation of the protein N-terminus and lysines, methylation of arginines or lysines, and trimethylation of lysines were allowed.

Data for phosphosite analysis of subunit *a* from *S. cerevisiae* were also searched using MaxQuant software against a database containing subunit *a* (Vph1p). Bovine F_oF_1 data were also searched using MaxQuant against a database containing sequences minus transit peptides for all 17 subunits ($\alpha, \beta, \gamma, \epsilon$, *a*, *b1*, *c*, *d*, *e*, *f*, *g*, *A6L*, F_6, *OSCP*, *DAPIT*, *6.8PL* and IF_1). Mass accuracy filters were set to 20 ppm for precursor ions and 0.5 Da for MS/MS. Enzyme was set to trypsin with up to two missed cleavage sites. Carbamidomethylation of cysteines was a fixed modification and variable modifications of acetylation of protein N-terminus and lysines, oxidation of methionines, phosphorylation of serines, threonines and tyrosines, methylation of arginines or lysines, and trimethylation of lysines were allowed.

5.8.3 Cross-Linking Mass Spectrometry

5.8.3.1 Comparative Cross-linking of the Chloroplast F-type ATP Synthase from *S. oloracea*

Comparative cross-linking experiments on cATP synthase from *S. oleracea* were performed by Professor Carla Schmidt (The Martin Luther University of Halle-Wittenberg), as follows. For deacetylation, an aliquot of F-type ATP synthase from *S. oleracea* (20 μL of 10 μM) was deacetylated by incubation with 1.25 μg human recombinant SIRT3 (Cambridge Bioscience) at 37 ^{o}C for 2 hours. For simultaneous deacetylation and dephosphorylation, F-type ATP synthase was incubated with SIRT3 for 1 hour, followed by addition of 100–250 units of calf alkaline phosphatase (New England Biolabs) and further incubation for 1 hour at

37 ^{o}C. Control samples were processed in parallel using buffer instead of SIRT3/phosphatase addition. The deacetylated and deacetylated/dephosphorylated samples were cross-linked by incubation with 5 μL BS3-d4 (2.5 mM), while the control samples were incubated with 5 μL BS3-d0 (2.5 mM) at 25 ^{o}C for 1 hour before quenching with 1 μL Tris (1M). Each BS3-d4 cross-linked sample was mixed 1:1 with the cross-linked control sample, precipitated with ethanol, and solubilised in Rapigest surfactant (Waters) according to the manufacturer's protocol. The protein was reduced with DTT, alkylated with iodoacetamide, and digested with trypsin before analysis by LC-MS/MS, as described above.

5.8.3.2 Cross-linking of the Mitochondrial F-type ATP Synthase from *B. taurus*

Mitochondrial F-type ATP synthase from *B. taurus* was cross-linked with either BS3-d0/d4 or DSS-d0/d12. Equimolar mixtures of deuterated and non-deuterated cross-linkers BS3-d0/d4 (4.2 mM) or DSS-d0/d12 (17 mM) were prepared in water and DMSO, respectively. Aliquots of F-type ATP synthase (12 μL of 10 μM) were cross-linked separately with 2 μL BS3-d0/d4 (4.2 mM) and 0.5 μL DSS-d0/d12 (17 mM) for 90 minutes at room temperature, followed by quenching with 1 μL Tris (1 M). Proteins were separated by SDS-PAGE and gel bands were excised and digested with trypsin, as described above. A second aliquot of DSS-cross-linked protein was digested with both trypsin and chymotrypsin overnight at 26 ^{o}C. Peptides were analysed by LC-MS/MS as described above.

5.8.3.3 Cross-linking Data Analysis

Raw data were converted into mgf format using pXtract software (http://www.pfindstudio.com/software/pXtract/index.html). Cross-links were identified using pLink software [102], and comparative cross-links quantified, as

described [97]. Search parameters were as follows: instrument spectra, HCD; enzyme, trypsin or trypsin/chymotrypsin; maximum number of cleavage sites, 3; variable modifications of methionine oxidation and carbamidomethylation of cysteine; cross-linkers, BS3-d0/d4 or DSS-d0/d12; minimum peptide length, 4; maximum peptide length, 100; minimum peptide mass, 400 Da; maximum peptide mass, 10,000 Da; FDR, 1 %. Cross-link identifications were manually validated by visual inspection of each spectrum. For comparative cross-linking data, ion chromatograms were extracted and ratios reflecting changes in protein interactions were calculated and averaged across several peptide mass spectra. Cross-links were visualised with xVis software (https://xvis.genzentrum.lmu.de [247]).

5.8.4 Homology Modelling

For cATP synthase subunit ϵ, a previously-reported homology model was used [99] generated with MODELLER [350], using the template PDB ID: 2RQ7 [301]. The homology model of bovine IF_1 was generated with the Phyre2 server [237] in intensive mode using the template PDB ID: 1GMJ [315] and alignment to the structure of IF_1-bound $\alpha_3\beta_3$ (PDB ID: 1OHH [316]).

5.8.5 Denaturing Mass Spectrometry

For denaturing LC-MS analysis, bovine F-Type ATP synthase was analysed by LC coupled to an LTQ Orbitrap XL Mass Spectrometer. Proteins were loaded onto an RP monolithic trap column (200 μm $\times$ 5 mm, PepSwift PS-DVB, Thermo Fisher Scientific) with buffer A (0.1% (v/v) FA) and separated on a RP monolithic analytical column (100 μm $\times$ 25 cm ProSwift RP 4H, Thermo Fisher Scientific) with a 60 minute linear gradient to buffer B (0.1% (v/v) FA/ 50% (v/v) isopropanol (IPA)/50% (v/v) ACN) with a flow rate of 500 nL min^{-1}. The LTQ Orbitrap XL spray voltage was 2.2 kV and the capillary temperature 275 ^{o}C. Positive ion mode

was used and the LTQ Orbitrap XL was operated in survey full-scan MS mode. The MS spectra were acquired in the Orbitrap (m/z 350–2,000) with a resolution of 60,000.

Alternatively, 13 μL protein sample was acidified with 0.2 % TFA (v/v) and loaded onto a ZipTip C4 column (Millipore) pre-washed with 50 % ACN (v/v)/ 0.1 % (v/v) TFA and 0.1 % (v/v) TFA. The protein was washed with 0.1% (v/v) TFA, followed by water. Bound proteins were eluted with 1.3 μL 75 % (v/v) ACN/ 0.1 % (v/v) TFA and analysed on a Q Exactive Plus Mass Spectrometer, as described below.

5.8.6 Lipidomics Analysis

The lipidomics experimental procedure was performed by Dr Di Wu and Dr Dror Chorev (The University of Oxford, UK), as follows. An aliquot of bovine F-type ATP synthase was digested with trypsin overnight at 37 ^{o}C and then dried in a vacuum concentrator (SpeedVac, Thermo Fisher Scientific). The sample was resuspended 60% ACN and sonicated for 10 minutes. The sample was analysed by LC-MS/MS with a Dionex UltiMate 3000 RSLC nano System connected to an LTQ Orbitrap XL Hybrid Mass Spectrometer (Thermo Fisher Scientific) via a dynamic nanospray source. Lipids were separated on an analytical column (Acclaim PepMap 100, C18, 75 μm $\times$ 15 cm; Thermo Fisher Scientific). Buffers were adapted from [351]; buffer A 60 % (v/v) ACN/ 10 mM ammonium formate/ 0.1 % (v/v) FA; buffer B 90 % IPA/ 10 % ACN/ 10 mM ammonium formate/ 0.1 % FA. Lipids were separated with a linear gradient of 32 % to 99 % buffer B at a temperature of 40 ^{o}C and flow rate of 300 nL min^{-1} over 30 minutes. The spray voltage was 1.8 kV and the capillary temperature 175 ^{o}C. Negative ion mode was used and the LTQ Orbitrap XL was operated in data-dependent mode. Survey full-scan MS

spectra were acquired in the Orbitrap (m/z 350–2,000) with a resolution of 60,000. The top five most intense ions were selected for CID fragmentation in the linear ion trap at an AGC target of 30,000 and a normalized collision energy of 38 % at an activation of q = 0.25, with an activation time of 30 ms [352].

Identified masses were searched in the LIPID MAPS database (www.lipidmaps.org/tools/ [330]) for potential assignment and MS/MS spectra were assigned manually.

5.8.7 Native Mass Spectrometry

Native mass spectra were acquired on a Q Exactive Plus Mass Spectrometer with EMR (Thermo Fisher Scientific), modified for transmission of high-mass complexes as described [44]. F-type ATP synthase from *B. taurus* was exchanged to 50 mM ammonium acetate (pH 7.5), 0.03 % LMNG with micro Bio-spin columns (Bio-Rad) twice, prior to MS analysis. 3 μL protein complex was loaded into a gold-coated borosilicate capillary, manufactured in-house [165]. Instrument parameters were optimised for each spectrum with typical values as follows: capillary voltage 1.6 kV; capillary temperature 30 ^{o}C; desolvation voltage -100 V; HCD voltage 0-200 V; C-trap entrance lens 2; inject flatapole voltage 8 V; interflatapole voltage 4 V; bent flatapole voltage 4 V; trapping gas parameter 4–8; noise threshold 3.

For proteins denatured on a ZipTip C4 column, instrument parameters were as follows: capillary voltage 1.6 kV; capillary temperature 350 ^{o}C; HCD voltage 5 V; desolvation voltage -200 V; C-trap entrance lens 2; inject flatapole voltage 8 V; interflatapole voltage 4 V; bent flatapole voltage 4 V; trapping gas parameter 4; noise threshold 3.

Chapter 6

Conclusions

In this thesis, MS-based proteomics and native MS analyses have been combined to study three protein systems; Hsp70 chaperones, eukaryotic translation initiation factors, and F- and V-type ATP synthases, in order to elucidate their structure, organisation, and regulatory mechanisms.

In Chapter 3, it was shown that while Hsp70 can form dimers, the extent of dimerisation depends on expression system, with Sf9-expressed Hsp70 showing a greater propensity to dimerise than *E. coli*-expressed Hsp70. It was expected that a phosphorylation site, *p*-T504, identified in $Hsp70_{Sf9}$ could be involved in promoting dimerisation. Increasing the ionic strength decreased $Hsp70_{Sf9}$ dimerisation, rationalised by a shielding of ionic interactions. Incubating both Hsp70 proteins with a phosphatase enzyme to remove phosphosites decreased dimerisation in $Hsp70_{Sf9}$ but not $Hsp70_{E.\ coli}$. A phosphomimic of the T504 phosphosite, produced in *E. coli*, showed an increase in dimerisation, relative to the wild type $Hsp70_{E.\ coli}$. Finally, a potential model for dimerisation was considered, with dimerisation of a substrate-binding deficient mutant ruling out a substrate-binding type dimerisation interaction. Together these results suggest a functional role for an Hsp70 dimer and

a regulatory role for phosphorylation in the context of protein complex assembly.

The effect of nucleotide binding on Hsp70 dimerisation could be investigated with a number of experimental approaches. Native MS of Hsp70 in the presence on non-hydrolysable ATP analogues, such as AMP-PNP and ATP-γS would allow investigation of Hsp70 in the ATP-bound conformation while removing the effects of ATP hydrolysis. Furthermore, to compare the effect of the ADP and ATP-bound states of Hsp70, an Hsp70 mutant with impaired ATP hydrolysis, such as T204A [179], purified in the presence of either ADP or ATP could be used. Native MS of the two samples would then allow ADP and ATP-bound Hsp70 to be compared, while minimising ATP hydrolysis and its associated conformational changes. Newly-developed high-resolution mass spectrometers [44], have recently allowed more accurate determination of protein complex masses and higher resolution allowing nucleotide-binding states to be more clearly resolved, as demonstrated in the study of eIFs and ATP synthases in Chapters 3 and 4. Applying this new instrumentation to the study of Hsp70 may therefore allow the distribution of both PTMs and nucleotide-binding states between Hsp70 monomer and dimer populations.

This research has highlighted the effect of Hsp70 PTMs, but the role of acetylation in particular remains relatively unexplored. Native MS of acetylation mimics (i.e. K108Q, K159Q), in analogous experiments to those performed on the T504E phosphomimic, would help clarify the effect of acetylation on Hsp70 interactions with other protein complexes. The effect of PTMs on Hsp70's interactions with its co-chaperones and clients, including a potential role in targeting to the antagonistic CHIP-mediated protein degradation and HOP-mediated protein folding pathways, would be a worthwhile target for future investigations. By incubating Hsp70, and

its variants such as the phosphomimic Hsp70$_{T504E}$, with CHIP and HOP proteins, native MS could be used to investigate the competition in Hsp70-CHIP/HOP complex formation. Furthermore, the use of recently-developed high-resolution instrumentation may lead to additional insights into the effects of nucleotide binding and PTM status on protein complex formation, providing a more complete understanding of Hsp70 interactions.

In Chapter 4, eukaryotic translation initiation factors were studied by MS. Eukaryotic translation initiation is a tightly-regulated process, with PTMs playing an important role. Multiple phosphorylation and acetylation sites were identified on eIF2 and eIF2B. In some cases these modifications, when combined with previously published sites, occurred in clusters such as on the linker to the catalytic domain of eIF2Bϵ and the N-terminal region of eIF2β both of which are involved in multiple protein-protein interactions. These phosphorylation sites are highly conserved across multiple eukaryotic species and therefore likely play a role in regulation. Cross-linking of yeast eIF2B revealed multiple intersubunit interactions, with cross-links supporting a structure for eIF2B similar to the reported *S. pombe* eIF2B crystal structure with a $\beta_2\delta_2$ core, peripheral $\gamma\epsilon$ subunits, and an α dimer. Additional cross-linking experiments could be used to further investigate the role of PTMs in the protein-protein interactions of eIF2 and eIF2B. Together with native MS, a comparative cross-linking strategy, as utilised in the study of cATP synthase in Chapter 5, could be employed. Treatment of eIF2 and eIF2B with phosphatase and deacetylase enzymes followed by analysis by native MS and comparative cross-linking could give insights into the effect of PTMs on protein complex stability and conformations. This study could be further extended to investigate the effect of PTMs on interactions between eIFs by incubating eIF2, eIF2B, and eIF5 together after treatment with phosphatase and deacetylase enzymes and observing any

changes in protein complex formation by native MS. To study the effects of specific PTMs such as conserved phosphosites identified in flexible protein regions, site-directed mutagenesis could also be used to produce phosphomimic variants of eIF2 and eIF2B in a similar approach as that used in Chapter 3 to investigate $Hsp70_{E.coli}$.

Native MS analysis of human eIF2B corroborated the proposed structure of eIF2B in a pseudo-symmetric decameric assembly, with CID and in-solution disruption generating subcomplexes providing further experimental support for the model, including the formation of eIF2Bα dimers. Upon addition of the drug ISRIB, known to enhance cognition, MS analysis showed a decrease in the dissociation of eIF2Bα, with a concomitant decrease in the detected eIF2B$\beta\gamma\delta\epsilon$ subcomplex. Together, this indicates an ISRIB-mediated stabilisation of the intact, active eIF2B decameric complex and thus rationalises the abrogation of the ISR on ISRIB treatment and consequently its reported therapeutic effects. As PTMs are known to affect eIF2B-eIF2 interactions, the influence of PTMs on ISRIB treatment is an important area to explore. To understand the effects of PTMs, eIF2B could be purified from human cells such as HEK293 cells followed by ISRIB treatment and native MS analysis to determine if the effects of ISRIB match those observed here in *E. coli*-expressed eIF2B. Understanding these interactions is a critical step on the pathway to allowing ISRIB treatment for human patients, and therefore such an experiment could yield important insights into human eIF2B-ISRIB interactions.

In Chapter 5, a study of the rotary ATP synthases was presented, with enzymes from three species studied here. Firstly, the identity of two small membrane-associated subunits; *e* and *f*, from the yeast Golgi V-type ATPase were identified through proteomics analysis and found to be the same form as those

associated with the vacuolar complex. It is not known if subunit e and f interact in the same way with the two isoforms of subunit a, and therefore cryo-EM analysis of the Golgi V_o domain would allow this to be clarified. Next, a phosphorylation hypothesis was tested for the dissociative autoinhibition of V-type ATPase in yeast, with phosphorylation of subunit a not found to influence autoinhibition. There is still potential for PTMs, including phosphorylation, to be responsible for V_o/V_1 dissociation. To determine any role phosphorylation may play in V-type ATPase autoinhibition a comprehensive analysis of each subunit could be undertaken. By subjecting subunits from the intact V_oV_1 and the dissociated V_o/V_1 complexes to both denatured MS and proteomics analyses, both the global PTM status and the identification of specific PTM sites could be compared between the two samples, potentially resolving this long-unanswered question.

The effect of PTMs on the conformation and protein-protein interactions of the spinach F-type ATP synthase was explored using comparative cross-linking mass spectrometry. When the enzyme was deacetylated, evidence suggested that a conformational change on subunit ϵ occurred which is controlled by acetylation state. Secondly, when the enzyme was simultaneously dephosphorylated and deacetylated, a significant decrease in cross-linking was observed arising from extensive disruption to protein-protein interfaces on account of the strongly-destabilising effect of simultaneous removal of acetylation and phosphorylation sites. These results highlight the significant impact that global removal of modifications can impart on protein complexes. The effect of individual PTMs on protein conformation and complex stability is an area where current understanding is limited and therefore examination of these effects should be the next research objective. This could be achieved by using site-directed mutagenesis to produce acetylation mimics, focussing particularly on acetylation sites identified

on subunit ϵ (i.e. K112Q [266]).

Finally, the bovine mitochondrial F-type ATP synthase was investigated through a combination of proteomics and native MS. The presence and PTM status of all protein subunits was analysed, indicating a strong degree of modification and identifying co-purification of the IF_1 inhibitor. To explore the role of PTMs in modulating protein-protein interactions, a combination of native MS and comparative cross-linking, in conjunction with dephosphorylation and deacetylation, could be utilised, analogous to the approach used for the cATP synthase. This would not only demonstrate if these PTMs are important in strengthening subunit interfaces but may also reveal PTM-influenced conformational changes, as identified in the cATP synthase.

Chemical cross-linking mass spectrometry revealed subunit interactions between the small membrane-bound subunits, placing subunits *DAPIT* and *6.8PL* in the vicinity of the IMD in close proximity to subunit *e*. Determining the stoichiomtery of subunits *DAPIT* and *6.8PL* to the F-type synthase complex could provide further evidence to guide assignment of the constituents of the IMD. Isotopically-labelled peptides could therefore be used to perform a relative quantification of subunits *DAPIT* and *6.8PL* to determine their ratios to both one another and to the F-type ATP synthase complex.

Cross-linking confirmed the binding of IF_1 to the centre of the $\alpha_3\beta_3$ head, contacting the γ central stalk with the cross-links indicating binding of an IF_1 monomer as opposed to the expected dimer. Native MS analysis provided further support for the binding of monomeric IF_1, with multiple IF_1-bound complexes observed each incorporating a single IF_1 molecule. Optimising experimental

conditions and instrument parameters for native MS analysis may allow the intact bovine F-type ATP synthase dimer to be detected as well-resolved peaks. Not only would this allow the stoichiometry of IF_1-binding to the dimer to be determined, but would also give information on any effect IF_1 has on stabilising ATP synthase dimers. Optimisation of native MS analysis may also allow the F_o domain of ATP sythase to be studied in greater detail, including the supernumerary subunits and the IMD, and reveal the presence of any specific lipid binding. The potential for native MS to uncover such a wealth of information means that developing the ideal conditions for MS analysis of the bovine F-Type ATP synthase should be prioritised in further studies.

In summary, the research presented in this thesis has provided numerous insights into protein PTMs, protein-protein interactions, protein complex stoichiometry, organisation, and structure of the three protein systems investigated and contributes to our understanding in the field of protein regulation. Furthermore, the data presented herein opens up new questions and provides the stimulus for multiple exciting areas of investigation with the aim of building a more complete understanding of the many carefully-tuned regulatory mechanisms effected by proteins in biological organisms.

Appendix A

Supplemental Information for Chapter 3

Protein subunits	Theoretical mass (Da)
^{13}C-Hsp70$_{E.\ coli}$	73,774
(^{13}C-Hsp70$_{E.\ coli}$)$_2$	147,548
Hsp70$_{Sf9}$	70,009
(Hsp70$_{Sf9}$)$_2$	140,018
(^{13}C-Hsp70$_{E.\ coli}$)$_1$(Hsp70$_{Sf9}$)$_1$	143,783
Hsp70$_{T504E}$	70,667
(Hsp70$_{T504E}$)$_2$	141,334
^{15}N-Hsp70$_{E.\ coli}$	71,500
(^{15}N-Hsp70$_{E.\ coli}$)$_2$	142,999
(Hsp70$_{T504E}$)$_1$(^{15}N-Hsp70$_{E.\ coli}$)$_1$	142,167
^{15}N-Hsp70$_{V438F\ T504E}$	71,576
(^{15}N-Hsp70$_{V438F\ T504E}$)$_2$	143,151
(Hsp70$_{T504E}$)$_1$(^{15}N-Hsp70$_{V438F\ T504E}$)$_1$	142,243

Table A.1: Theoretical masses of monomers and dimers of Hsp70 variants calculated from the protein sequences.

Human Hsp70

MAKAAAIGIDLGTTYSCVGVFQHGKVEIIANDQGNRTTPSYVAFTDTERLIGDAAKNQVALNPQNTVFDAKRLIGRKFGDPVV
QSDMKHWPFQVINDGDKPKVQVSYKGETKAFYPEEISSMVLTKMKEIAEAYLGYPVTNAVITVPAYFNDSQRQATKDAGVIA
GLNVLRIINEPTAAAIAYGLDRTGKGERNVLIFDLGGGTFDVSILTIDDGIFEVKATAGDTHLGGEDFDNRLVNHFVEEFKRKHKK
DISQNKRAVRRLRTACERAKRTLSSSTQASLEIDSLFEGIDFYTSITRARFEELCSDLFRSTLEPVEKALRDAKLDKAQIHDLVLVGG
STRIPKVQKLLQDFFNGRDLNKSINPDEAVAYGAAVQAAILMGDKSENVQDLLLLDVAPLSLGLETAGGVMTALIKRNSTIPTKQ
TQIFTTYSDNQPGVLIQVYEGERAMTKDNNLLGRFELSGIPPAPRGVPQIEVTFDIDANGILNVTATDKSTGKANKITITNDKGRL
SKEEIERMVQEAEKYKAEDEVQRERVSAKNALESYAFNMKSAVEDEGLKGKISEADKKKVLDKCQEVISWLDANTLAEKDEFEH
KRKELEQVCNPIISGLYQGAGGPGPGGFGAQGPKGGSGSGPTIEED

Human Hsp70 His-tagged

HHHHHHGGGSENLYFQ*GIDPFTMAKAAAIGIDLGTTYSCVGVFQHGKVEIIANDQGNRTTPSYVAFTDTERLIGDAAKNQVA
LNPQNTVFDAKRLIGRKFGDPVVQSDMKHWPFQVINDGDKPKVQVSYKGETKAFYPEEISSMVLTKMKEIAEAYLGYPVTNA
VITVPAYFNDSQRQATKDAGVIAGLNVLRIINEPTAAAIAYGLDRTGKGERNVLIFDLGGGTFDVSILTIDDGIFEVKATAGDTHL
GGEDFDNRLVNHFVEEFKRKHKKDISQNKRAVRRLRTACERAKRTLSSSTQASLEIDSLFEGIDFYTSITRARFEELCSDLFRSTLEP
VEKALRDAKLDKAQIHDLVLVGGSTRIPKVQKLLQDFFNGRDLNKSINPDEAVAYGAAVQAAILMGDKSENVQDLLLLDVAPLS
LGLETAGGVMTALIKRNSTIPTKQTQIFTTYSDNQPGVLIQVYEGERAMTKDNNLLGRFELSGIPPAPRGVPQIEVTFDIDANGIL
NVTATDKSTGKANKITITNDKGRLSKEEIERMVQEAEKYKAEDEVQRERVSAKNALESYAFNMKSAVEDEGLKGKISEADKKKVL
DKCQEVISWLDANTLAEKDEFEHKRKELEQVCNPIISGLYQGAGGPGPGGFGAQGPKGGSGSGPTIEEVD

Figure A.1: Protein sequences of recombinantly-expressed human Hsp70 1A. For *E. coli* expression, an N-terminal 6×His-tag and TEV cleavage site are included. *Denotes TEV cleavage site.

Appendix B

Supplemental Information for Chapter 4

Protein subunit	Modified residue	Highest Score	Enrichment	Direct analysis	Database	Conservation
eIF2α	*N-term-ac-* S2	46		x	*	
	p S194	16	x		*	
	p S292	24		x	A, B, C	*S. pombe*
	p S294	17		x	A, B, C	*S. pombe, C. elegans*
	p S301	42		x	B, C	*C. elegans*
eIF2β	*N-term-ac-* S2	63		x	*	
	p S40	76	x	x	A, B, C	*S. pombe, C. albicans*
	ac K89	49		x	*	
	p S91	39	x		C	*S. pombe*
	p S92	43	x	x	A, C	*S. pombe*
	p S112	64	x	x	A, B, C	*D. melanogaster*
eIF2γ	*p* T60	88	x		A, C	
	ac K129	37		x	*	
	p T137	16	x		*	
	p S258	41	x		A, C	
	p S265	27	x		*	
	p S312	32	x		C	*C. albicans*
	ac K401	17		x	*	
	ac K519	47		x	*	
eIF2Bα	*N-term-ac-* S2	52		x	A	
	p Y166	19	x		*	
eIF2Bβ	*N-term-ac-* S2	79		x	*	
eIF2Bγ	*p* S296	39	x		A, C	
	ac K346	58		x	*	
eIF2Bδ	*p* S131	49	x		*	
eIF2Bε	*p* S435	66	x		C	
	p S449	72	x		*	
	p S525	67	x	x	A, C	*H. sapiens*
	p S538	57	x	x	A, B, C	*H. sapiens, M. musculus, S. pombe*

Table B.1: PTMs identified in eIF2 and eIF2B from *S. cerevisiae*. The eIF protein subunit and the identified modified residues (p, phosphorylation; ac, acetylation) with the highest observed Mascot score for each modified site are listed. The method of determination (phosphopeptide enrichment with TiO_2 or direct analysis without enrichment) is indicated. The presence of the identified PTMs in databases (A, UniProtKB [260]; B, PHOSIDA [261]; C, PTM*func* [170]) along with site conservation and their respective species is given.

Human eIF2α
MDDKELIEYFKSQMKEDPDMASAVAAIRTLLEFLKRDKGETIQGLRANLTSAIETLCGVDSSVAVSSGGELFLRFISLASLEYSDYS
KCKKIMIERGELFLRRISLSRNKIADLCHTFIKDGATILTHAYSRVVLRVLEAAVAAKKRFSVYVTESQPDLSGKKMAKALCHLNVP
VTVVLDAAVGYIMEKADLVIVGAEGVVENGGIINKIGTNQMAVCAKAQNKPFYVVAESFKFVRLFPLNQQDVPDKFKYKADTL
KVAQTGQDLKEEHPWVDYTAPSLITLLFTDLGVLTPSAVSDELIKLYL

Human eIF2α His-tagged version
MHHHHHHGGGSENLYFQ*SDDKELIEYFKSQMKEDPDMASAVAAIRTLLEFLKRDKGETIQGLRANLTSAIETLCGVDSSVAVS
SGGELFLRFISLASLEYSDYSKCKKIMIERGELFLRRISLSRNKIADLCHTFIKDGATILTHAYSRVVLRVLEAAVAAKKRFSVYVTESQ
PDLSGKKMAKALCHLNVPVTVVLDAAVGYIMEKADLVIVGAEGVVENGGIINKIGTNQMAVCAKAQNKPFYVVAESFKFVRLF
PLNQQDVPDKFKYKADTLKVAQTGQDLKEEHPWVDYTAPSLITLLFTDLGVLTPSAVSDELIKLYL

Human eIF2β His-tagged Version
MHHHHHHGGGSENLYFQ*SPGSAAKGSELSERIESFVETLKRGGGPRSSEEMARETLGLLRQIITDHRWSNAGELMELIRREGR
RMTAAQPSETTVGNMVRRVLKIIREEYGRLHGRSDESDQQESLHKLLTSGGLNEDFSFHYAQLQSNIIEAINELLVELEGTMENI
AAQALEHIHSNEVIMTIGFSRTVEAFLKEAARKRKFHVIVAECAPFCQGHEMAVNLSKAGIETTVMTDAAIFAVMSRVNKVIIGT
KTILANGALRAVTGTHTLALAAKHHSTPLIVCAPMFKLSPQFPNEEDSFHKFVAPEEVLPFTEGDILEKVSVHCPVFDYVPPELITL
FISNIGGNAPSYIYRLMSELYHPDDHVL

Human eIF2γ
MEFQAVVMAVGGGSRMTDLTSSIPKPLLPVGNKPLIWYPLNLLERVGFEEVIVVTTRDVQKALCAEFKMKMKPDIVCIPDDAD
MGTADSLRYIYPKLKTDVLVLSCDLITDVALHEVVDLFRAYDASLAMLMRKGQDSIEPVPGQKGKKKAVEQRDFIGVDSTGKRL
LFMANEADLDEELVIKGSILQKHPRIRFHTGLVDAHLYCLKKYIVDFLMENGSITSIRSELIPYLVRKQFSSASSQQGQEEKEEDLKK
KELKSLDIYSFIKEANTLNLAPYDACWNACRGDRWEDLSRSQVRCYVHIMKEGLCSRVSTLGLYMEANRQVPKLLSALCPEEPP
VHSSAQIVSKHLVGVDSLIGPETQIGEKSSIKRSVIGSSCLIKDRVTITNCLLMNSVTVEEGSNIQGSVICNNAVIEKGADIKDCLIGS
GQRIEAKAKRVNEVIVGNDQLMEI

Human eIF2δ
MAAVAVAVREDSGSGMKAELPPGPGAVGREMTKEEKLQLRKEKKQQKKKRKEEKGAEPETGSAVSAAQCQVGPTRELPESGI
QLGTPREKVPAGRSKAELRAERRAKQEAERALKQARKGEQGGPPPKASPSTAGETPSGVKRLPEYPQVDDLLLRRLVKKPERQ
QVPTRKDYGSKVSLFSHLPQYSRQNSLTQFMSIPSSVIHPAMVRLGLQYSQGLVSGSNARCIALLRALQQVIQDYTTPPNEELSR
DLVNKLKPYMSFLTQCRPLSASMHNAIKFLNKEITSVGSSKREEEAKSELRAAIDRYVQEKIVLAAQAISRFAYQKISNGDVILVYG
CSSLVSRILQEAWTEGRRFRVVVVDSRPWLEGRHTLRSLVHAGVPASYLLIPAASYVLPEVSKVLLGAHALLANGSVMSRVGTA
QLALVARAHNVPVLVCCETYKFCERVQTDAFVSNELDDPDDLQCKRGEHVALANWQNHASLRLLNLVYDVTPPELVDLVITEL
GMIPCSSVPVVLRVKSSDQ

Human eIF2Bε
MAAPVVAPPGVVVSRANKRSGAGPGGSGGGGARGAEEEPPPPLQAVLVADSFDRRFFPISKDQPRVLLPLANVALIDYTLEFLT
ATGVQETFVFCCWKAAQIKEHLLKSKWCRPTSLNVVRIITSELYRSLGDVLRDVDAKALVRSDFLLVYGDVISNINITRALEEHRLR
RKLEKNVSVMTMIFKESSPSHPTRCHEDNVVVAVDSTTNRVLHFQKTQGLRRFAFPLSLFQGSSDGVEVRYDLLDCHISICSPQV
AQLFTDNFDYQTRDDFVRGLLVNEEILGNQIHMHVTAKEYGARVSNLHMYSAVCADVIRRWVYPLTPEANFTDSTTQSCTHSR
HNIYRGPEVSLGHGSILEENVLLGSGTVIGSNCFITNSVIGPGCHIGDNVVLDQTYLWQGVRVAAGAQIHQSLLCDNAEVKERV
TLKPRSVLTSQVVVGPNITLPEGSVISLHPPDAEEDEDDGEFSDDSGADQEKDKVKMKGYNPAEVGAAGKGYLWKAAGMNM
EEEEELQQNLWGLKINMEEESESESEQSMDSEEPDSRGGSPQMDDIKVFQNEVLGTLQRGKEENISCDNLVLEINSLKYAYNVS
LKEVMQVLSHVVLEFPLQQMDSPLDSSRYCALLLPLLKAWSPVFRNYIKRAADHLEALAAIEDFFLEHEALGISMAKVLMAFYQL
EILAEETILSWFSQRDTTDKGQQLRKNQQLQRFIQWLKEAEEESSEDD

Figure B.1: Protein sequences of recombinantly-expressed human eIF2B subunits. The five subunits (α, 6 $\times$ His-β, γ, δ, ϵ) of human eIF2B were expressed together to produce the intact eIF2B complex. Separately, eIF2Bα subunit incorporating an N-terminal 6$\times$His-tag and TEV cleavage site was expressed to produce isolated eIF2Bα. *Denotes TEV cleavage site.

Protein subunits	Theoretical mass (Da)
α	33,712
γ	50,240
α_2	67,424
His-α_2	71,350
ε	80,380
γε	130,620
βγδε	229,130
$\gamma_2\varepsilon_2$	261,240
$\alpha_2\beta_2\gamma_2\delta_2\varepsilon_1$	445,304
$\alpha_2\beta_2\gamma_1\delta_2\varepsilon_2$	475,444
$\alpha_1\beta_2\gamma_2\delta_2\varepsilon_2$	491,972
$\alpha_2\beta_2\gamma_2\delta_2\varepsilon_2$	525,684
$\alpha_4\beta_4\gamma_4\delta_4\varepsilon_4$	1,051,368

Table B.2: Theoretical masses of decameric human eIF2B and subcomplexes calculated from the protein sequences.

Appendix C

Supplemental Information for Chapter 5

Protein 1	Protein 2	Residue 1	Residue 2	d0/d4
α	α	11	25	0.94
		11	114	0.32
		11	466	1.21
		114	25	1.76
		378	374	1.09
		378	384	0.86
		456	466	1.31
		456	469	0.56
		456	491	0.92
		469	491	0.64
α	β	11	17	1.71
		11	392	1.40
		374	426	1.17
		382	495	0.77
		456	359	1.04
		469	359	1.33
α	δ	2	101	0.99
α	I	2	125	0.67
		114	135	0.68
		466	116	0.71
α	II	456	164	0.59
β	β	18	50	1.10
		392	359	0.74
		359	399	0.52
		426	359	0.73
		392	399	0.44
		426	392	1.52

Protein 1	Protein 2	Residue 1	Residue 2	d0/d4
β	δ	18	205	1.09
		359	210	1.38
β	γ	359	55	1.05
		392	55	2.00
		426	71	1.20
		426	149	0.91
β	I	426	80	1.02
β	II	426	131	2.56
δ	δ	210	231	1.33
ε	ε	2	20	7.17
		105	112	204.21
ε	III	112	48	1.22
γ	γ	55	260	2.15
		55	334	0.76
		55	335	0.95
		59	334	0.76
		140	163	1.65
		140	222	1.26
		140	245	1.15
		140	263	1.46
		222	245	0.71
		245	260	0.94
I	I	80	87	1.49
I	II	73	131	0.79
		73	138	0.73
		107	161	1.00
		107	173	1.33

Table C.1: Quantification of comparatively cross-linked peptides of cATP synthase, after treatment with SIRT3 deacetylase enzyme. The cross-linked protein subunits, the cross-linked residues, and the ratio of naturally modified (d0) versus deacetylated (d4) cATP synthase cross-linked peptides are listed. Cross-link ratios that change by more than a factor of two are highlighted in grey.

Protein 1	Protein 2	Residue 1	Residue 2	d0/d4
α	α	378	491	2.07
α	β	382	495	4.17
α	δ	11	101	2.77
α	I	2	125	2.96
β	β	18	50	5.45
		359	17	0.25
		359	392	5.76
δ	δ	101	200	3.86
		200	210	1.88
γ	γ	55	335	1.20
		140	222	3.17
		140	245	8.23
		222	245	1.77
I	I	80	87	3.55
		87	107	1.00
I	II	73	131	5.01
		73	138	3.28
		107	161	4.27
		107	173	1.71

Table C.2: Quantification of comparatively cross-linked peptides of cATP synthase, after treatment with SIRT3 deacetylase and phosphatase enzymes. The cross-linked protein subunits, the cross-linked residues, and the ratio of naturally modified (d0) versus deacetylated (d4) cATP synthase cross-linked peptides are listed. Cross-link ratios that change by more than a factor of two are highlighted in grey.

Protein subunits	Expected mass (Da)
ε	5,650
δ	15,063
OSCP	20,929
β	51,756
$\alpha_2\beta_2\gamma_1\delta_1\varepsilon_1$*OSCP*$_1$ATP-Mg$_1$ADP-Mg$_1$	286,811
$\alpha_3\beta_3\gamma_1$CDL$_1$ATP-Mg$_1$ADP-Mg$_1$	353,642
$\alpha_3\beta_3\gamma_1\varepsilon_1CDL_1$ATP-Mg$_1$ADP-Mg$_1$	359,292
$\alpha_3\beta_3\gamma_1$*IF1*$_1$CDL$_1$ATP-Mg$_1$ADP-Mg$_1$	363,222
$\alpha_3\beta_3\gamma_1\varepsilon_1$*IF1*$_1CDL_1$ATP-Mg$_1$ADP-Mg$_1$	368,872
$\alpha_3\beta_3\gamma_1$*OSCP*$_1$CDL$_1$ATP-Mg$_1$ADP-Mg$_1$	374,571
$\alpha_3\beta_3\gamma_1\varepsilon_1$*OSCP*$_1CDL_1$ATP-Mg$_1$ADP-Mg$_1$	380,221
$\alpha_3\beta_3\gamma_1$*IF1*$_1$*OSCP*$_1$CDL$_1$ATP-Mg$_1$ADP-Mg$_1$	384,151
$\alpha_3\beta_3\gamma_1\delta_1$*OSCP*$_1CDL_1$ATP-Mg$_1$ADP-Mg$_1$	389,634
$\alpha_3\beta_3\gamma_1\delta_1\varepsilon_1$*OSCP*$_1CDL_1$ATP-Mg$_1$ADP-Mg$_1$	395,284
$\alpha_3\beta_3\gamma_1\delta_1\varepsilon_1$*IF1*$_1$*OSCP*$_1CDL_1$ATP-Mg$_1$ADP-Mg$_1$	404,864
$\alpha_2\beta_1\gamma_1\delta_1\varepsilon_1$*OSCP*$_2$*IF1*$_2$*F6*$_2a_2b_2c_{16}d_2e_2f_2g_2$ *DAPIT*$_2$*6.8PL*$_2$A6L$_2$CDL$_2$ATP-Mg$_1$ADP-Mg$_1$	656,240
$\alpha_3\beta_3\gamma_1\delta_1\varepsilon_1$*OSCP*$_2$*IF1*$_2$*F6*$_2a_2b_2c_{16}d_2e_2f_2g_2$ *DAPIT*$_2$*6.8PL*$_2$A6L$_2$CDL$_2$ATP-Mg$_2$ADP-Mg$_2$	815,916

Table C.3: Expected masses of bovine F-type ATP synthase complexes calculated from the subunit masses determined by denatured MS.

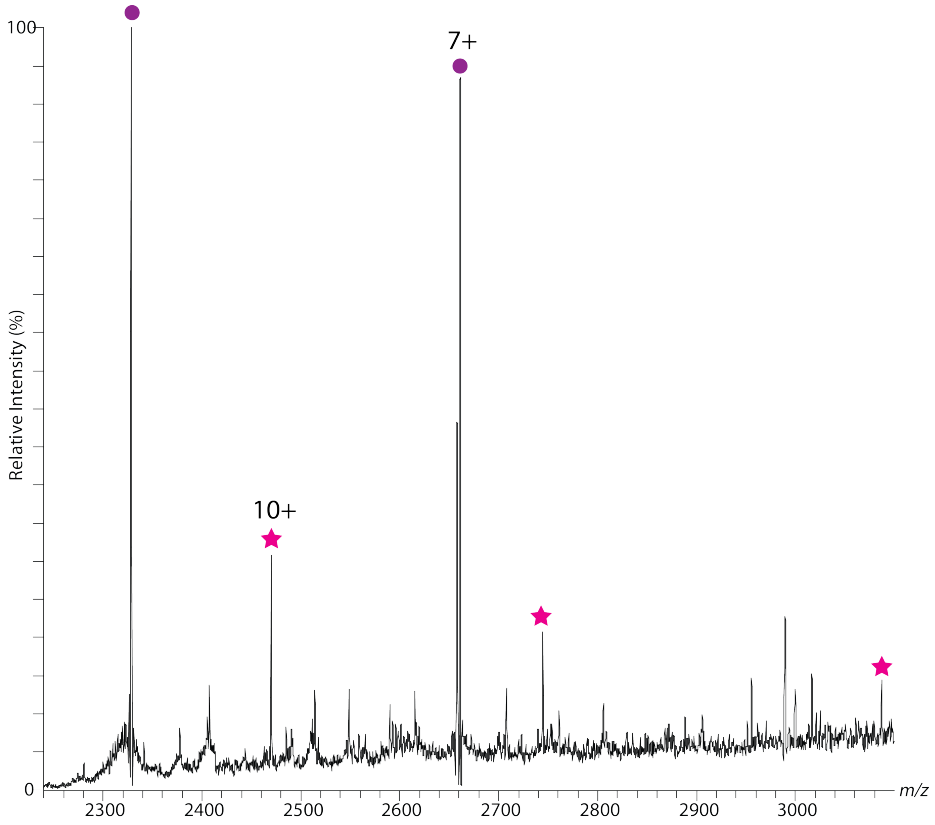


Figure C.1: Native mass spectrum of bovine F_oF_1 subjected to CID. Dissociated subunits *b1* (24,691 ±1 Da, pink stars) and *d* (18,620 ±1 Da, purple circles) are detected.

Protein 1	Residue 1	Protein 2	Residue 2	Peptide 1	Peptide 2	B/T	D/T	D/C
α	118	α	132	VVDALGNAIDGKGPIGSK(12)	KAPGIIPR(1)			x
α	118	α	218	VVDALGNAIDGKGPIGSK(12)	STVAQLVKR(8)	x		
α	124	α	132	VVDALGNAIDGKGPIGSKAR(18)	KAPGIIPR(1)			x
α	124	α	132	VVDALGNAIDGKGPIGSKAR(18)	VGLKAPGIIPR(4)	x		
α	124	α	262	VVDALGNAIDGKGPIGSKAR(18)	DNGKHALIIYDDLSK(4)	x	x	
α	132	α	124	KAPGIIPR(1)	GPIGSKAR(6)			x
α	151	α	132	EPMQTGIKAVDSLVPIGR(8)	VGLKAPGIIPR(4)	x		
α	151	α	132	EPMQTGIKAVDSLVPIGR(8)	KAPGIIPR(1)			x
α	151	α	175	EPMQTGIKAVDSLVPIGR(8)	QTGKTSIAIDTIINQK(4)	x		
α	175	α	132	QTGKTSIAIDTIINQK(4)	KAPGIIPR(1)			x
α	187	α	132	TSIAIDTIINQKR(12)	KAPGIIPR(1)			x
α	187	α	196	TSIAIDTIINQKR(12)	FNDGTDEKK(8)			x
α	187	α	209	TSIAIDTIINQKR(12)	VAIGQKR(6)			x
α	262	α	124	DNGKHALIIYDDLSK(4)	GPIGSKAR(6)	x	x	
α	262	α	196	DNGKHALIIYDDLSK(4)	FNDGTDEKKK(8)	x		
α	273	α	132	HALIIYDDLSKQAVAYR(11)	VGLKAPGIIPR(4)	x		
α	384	α	488	AMKQVAGTMKLELAQYR(3)	TDGKISEESDAKLK(4)	x		
α	384	α	488	AMKQVAGTMKLELAQYR(3)	TDGKISEESDAK(4)		x	
α	391	α	488	AMKQVAGTMKLELAQYR(10)	IRTDGKISEESDAK(6)		x	
α	482	α	488	FENAFLSHVISQHQALLSKIR(19)	TDGKISEESDAKLK(4)	x		
α	482	α	488	FENAFLSHVISQHQALLSKIR(19)	TDGKISEESDAK(4)		x	
α	488	α	384	IRTDGKISEESDAKLK(6)	AMKQVAGTMK(3)	x		
α	488	α	384	IRTDGKISEESDAK(6)	AMKQVAGTMK(3)		x	
α	498	α	455	LKEIVTNFLAGFEA(2)	GYLDKLEPSKITK(5)	x		
α	498	α	455	ISEESDAKLKEIVTNFLAGFEA(10)	GYLDKLEPSKITK(5)		x	
α	2	β	76	QKTGTAEVSSILEER(2)	GQKVLDSGAPIR(3)	x		
α	218	β	302	RSTVAQLVKR(9)	ITTTKK(5)	x		
α	460	β	379	GYLDKLEPSKITKFENAFLSHVISQHQALLSK(10)	GVQKILQDYK(4)		x	
α	463	β	379	ITKFENAFLSHVISQHQALLSK(3)	GVQKILQDYK(4)		x	
β	76	α	2	GQKVLDSGAPIRIPVGPETLGR(3)	QKTGTAEVSSILEER(2)	x		
β	164	α	151	IGLFGGAGVGKTVLIMELINNVAK(11)	EPMQTGIKAVDSLVPIGR(8)	x		
β	303	α	118	ITTTKKGSITSVQAIYVPADDLTDPAPATTFAHLDATTVLSR(5)	VVDALGNAIDGKGPIGSK(12)		x	
α	1	*OSCP*	50	QKTGTAEVSSILEER(1)	EPKMAASLLNPYVK(3)	x		
α	2	*OSCP*	50	QKTGTAEVSSILEER(2)	EPKMAASLLNPYVK(3)	x		
OSCP	3	α	83	FAKLVRPPVQIYGIEGR(3)	LIKEGDIVKR(3)	x		
OSCP	50	α	1	VGQILKEPKMAASLLNPYVKR(9)	QKTGTAEVSSILEER(1)	x		
OSCP	50	α	1	VGQILKEPKMAASLLNPYVK(9)	QKTGTAEVSSILEER(1)		x	
α	455	*d*	25	GYLDKLEPSKITK(5)	NQKAVANSLK(3)		x	
α	460	*d*	25	GYLDKLEPSKITK(10)	NQKAVANSLK(3)	x		
β	76	β	1	GQKVLDSGAPIR(3)	QASPSPK(1)	x		
β	86	β	1	GQKVLDSGAPIKIPVGPETLGR(12)	AQTSPSPK(1)		x	
β	111	β	1	IMNVIGEPIDERGPIKTK(16)	QASPSPK(1)	x		
β	140	β	111	QFAAIHAEAPEFVEMSVEQEILVTGIKVVDLLAPYAKGGK(27)	IMNVIGEPIDERGPIKTK(16)	x		
β	177	β	216	TVLIMELINNVAKAHGGYSVFAGVGER(13)	DATSKVALVYGQMNEPPGAR(5)	x	x	
β	178	β	217	TVLIMELINNVAKAHGGYSVFAGVGER(13)	DATSKVALVYGQMNEPPGAR(5)		x	
β	211	β	111	TREGNDLYHEMIESGVINLKDATSK(20)	GPIKTK(4)	x		
β	403	β	411	SLQDIIAILGMDELSEEDKLTVSR(19)	KIQR(1)	x		
β	403	γ	90	SLQDIIAILGMDELSEEDKLTVSR(19)	QMKSEAANLAAAGK(3)	x	x	
β	76	*OSCP*	1	GQKVLDSGAPIRIPVGPETLGR(3)	FAKLVRPPVQIYGIEGR(1)	x		
β	76	*OSCP*	73	GQKVLDSGAPIRIPVGPETLGR(3)	VGQILKEPKMAASLLNPYVK(9)	x		
β	76	*OSCP*	84	GQKVLDSGAPIRIPVGPETLGR(3)	MAASLLNPYVKR(11)	x		
β	76	*OSCP*	90	GQKVLDSGAPIRIPVGPETLGR(3)	VKSLSDMTAK(2)	x		
OSCP	3	β	76	FAKLVRPPVQIYGIEGR(3)	GQKVLDSGAPIR(3)	x		
OSCP	61	β	76	MAASLLNPYVKR(11)	GQKVLDSGAPIR(3)	x		

Table C.4: Cross-links identified in bovine F_oF_1. The cross-linked protein subunits, the cross-linked residues, and the peptide sequences are given. The experimental conditions (cross-linking reagent and digest enzyme) under which the cross-links were identified are indicated; BS3 cross-linker and trypsin (B/T), DSS cross-linker and trypsin (D/T), or DSS cross-linker and chymotrypsin (D/C).

Protein 1	Residue 1	Protein 2	Residue 2	Peptide 1	Peptide 2	B/T	D/T	D/C
γ	1	γ	11	ATLKDITRR(1)	LKSIKNIQK(2)		x	
γ	18	γ	11	NIQKITK(4)	KSIK(1)			x
γ	54	γ	176	EKADIKTPEDK(2)	KTEEKPIF(5)			x
γ	65	γ	58	KHLIIGVSSDR(1)	ADIKTPEDKK(4)			x
γ	65	γ	63	KHLIIGVSSDR(1)	TPEDKK(5)			x
γ	65	γ	104	KHLIIGVSSDR(1)	EVKIIGVGDK(3)			x
γ	101	γ	58	SEAANLAAAGKEVK(11)	ADIKTPEDK(4)		x	
γ	101	γ	111	SEAANLAAAGKEVK(11)	IIGVGDKIR(7)			x
γ	245	γ	1	NASEMIDKLTLTFNR(8)	ATLKDITR(1)		x	
γ	172	δ	1	SVISYKTEEKPIFSLDTISSAESMSIYDDIDADVLR(6)	AEAAAAQAPAAGPGQMSFTFASPTQVFFNSANVR(1)	x		
γ	176	δ	1	SVISYKTEEKPIFSLDTISSAESMSIYDDIDADVLR(10)	AEAAAAQAPAAGPGQMSFTFASPTQVFFNSANVR(1)	x		
δ	1	γ	39	AEAAAAQAPAAGPGQMSFTFASPTQVFFNSANVR(1)	ELKPARVYGVGSLALYEK(3)	x		
δ	1	γ	39	AEAAAAQAPAAGPGQMSF(1)	ELKPAR(3)			x
δ	1	γ	39	AEAAAAQAPAAGPGQMSFTFASPTQVFFNSANVR(1)	AERELKPAR(6)		x	
δ	1	γ	54	AEAAAAQAPAAGPGQMSF(1)	EKADIKTPEDK(2)			x
δ	1	γ	58	AEAAAAQAPAAGPGQMSFTFASPTQVFFNSANVR(1)	ADIKTPEDK(4)	x		
δ	1	γ	176	AEAAAAQAPAAGPGQMSFTFASPTQVFFNSANVR(1)	TEEKPIFSLDTISSAESMSIYDDIDADVLR(4)	x		
γ	129	ε	44	THSDQFLVTFKEVGR(11)	TSGSTIKIVK(7)	x		
ε	37	γ	30	ANAMKTSGSTIK(5)	MVAAAKYAR(6)	x		
δ	1	ε	37	AEAAAAQAPAAGPGQMSFTFASPTQVFFNSANVR(1)	ANAMKTSGSTIK(5)	x		
δ	114	ε	21	ANLEKAQSELLGAADEATR(5)	YSQICAKAVR(7)	x	x	
δ	114	ε	28	ANLEKAQSELLGAADEATR(5)	DALKTEFKANAMK(4)	x		
δ	114	ε	32	ANLEKAQSELLGAADEATR(5)	DALKTEFKANAMK(8)	x	x	
δ	114	ε	37	ANLEKAQSELLGAADEATR(5)	ANAMKTSGSTIK(5)	x		
b1	28	*b1*	1	LYPKTGVTGPY(4)	PVPPLPEHGGK(1)			x
b1	97	*b1*	1	IAQLEEVKQASIK(8)	PVPPLPEHGGK(1)			x
b1	191	*b1*	202	VVQSISAQQEKETIAK(11)	CIADLKLLSK(6)	x	x	
b1	112	*d*	109	QIQDAIDMEKSQQALVQKR(10)	SCAEFLTQSKTR(10)	x	x	
b1	28	*g*	55	LYPKTGVTGPY(4)	KIINSAK(1)			x
b1	191	*F6*	1	VVQSISAQQEKETIAK(11)	NKELDPVQK(1)	x	x	x
b1	191	*F6*	2	VVQSISAQQEKETIAK(11)	NKELDPVQK(2)	x	x	x
d	48	*d*	73	LATLPEKPPAIDWAYYK(7)	KFNALKVPIPEDK(1)	x		
d	48	*d*	78	LATLPEKPPAIDWAYYK(7)	FNALKVPIPEDK(5)	x	x	x
d	63	*d*	73	ANVAKAGLVDDFEK(5)	KFNALK(1)	x		
d	73	*d*	63	KFNALKVPIPEDK(1)	ANVAKAGLVDDFEK(5)	x	x	
d	78	*d*	48	FNALKVPIPEDKYTAQVDAEEKEDVK(5)	LATLPEKPPAIDWAYYK(7)	x	x	
d	85	*d*	99	FNALKVPIPEDKYTAQVDAEEK(12)	EDVKSCAEFLTQSK(4)		x	
d	109	*d*	117	SCAEFLTQSKTR(10)	IQEYEKELEK(6)	x		
d	144	*d*	148	NIIPFDQMTIEDLNEVFPETKLDK(21)	KKYPYWPHRPIETL(1)	x		
d	144	*d*	149	NIIPFDQMTIEDLNEVFPETKLDKK(21)	KYPYWPHRPIETL(1)	x		
d	25	*F6*	73	NQKAVANSLK(3)	FEVVEKPQS(6)	x		
d	63	*F6*	47	ANVAKAGLVDDFEK(5)	LKQMYGK(2)	x	x	
F6	45	*d*	63	QTSGGPVDAGPEYQQDLDRELFKLKQMYGK(23)	ANVAKAGLVDDFEK(5)	x		
F6	67	*d*	85	ADMNTFPNFTFEDPKFEVVEKPQS(15)	VPIPEDKYTAQVDAEEKEDVK(7)	x		
OSCP	28	*OSCP*	77	YATALYSAASKQNKLEQVEK(11)	EKFSPLTSNLINLLAENGR(2)	x	x	
OSCP	31	*OSCP*	77	YATALYSAASKQNKLEQVEK(14)	EKFSPLTSNLINLLAENGR(2)	x		
OSCP	50	*OSCP*	67	VGQILKEPKMAASLLNPYVK(9)	VKSLSDMTAK(2)	x	x	
OSCP	50	*OSCP*	75	VGQILKEPKMAASLLNPYVK(9)	SLSDMTAKEK(8)	x		
OSCP	61	*OSCP*	67	MAASLLNPYVKR(11)	VKSLSDMTAK(2)	x	x	
OSCP	77	*OSCP*	31	EKFSPLTSNLINLLAENGR(2)	QNKLEQVEKELLR(3)		x	
OSCP	77	*OSCP*	37	EKFSPLTSNLINLLAENGR(2)	LEQVEKELLR(6)	x		
OSCP	77	*OSCP*	37	EKFSPLTSNLINLLAENGR(2)	QNKLEQVEKELLR(9)		x	
OSCP	77	*OSCP*	67	EKFSPLTSNLINLLAENGR(2)	VKSLSDMTAK(2)	x		

Table C.5: Cross-links identified in bovine F_oF_1. The cross-linked protein subunits, the cross-linked residues, and peptide sequences are given. The experimental conditions (cross-linking reagent and digest enzyme) under which the cross-links were identified are indicated; BS3 cross-linker and trypsin (B/T), DSS cross-linker and trypsin (D/T), or DSS cross-linker and chymotrypsin (D/C).

References

[1] D. Eisenberg, E. M. Marcotte, I. Xenarios, and T. O. Yeates, "Protein function in the post-genomic era," *Nature*, vol. 405, no. 6788, pp. 823–826, 2000.

[2] P. E. Wright and H. J. Dyson, "Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm," *Journal of Molecular Biology*, vol. 293, no. 2, pp. 321–331, 1999.

[3] F. C. Bernstein, T. F. Koetzle, G. J. Williams, E. F. Meyer, M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi, "The protein data bank," *The FEBS Journal*, vol. 80, no. 2, pp. 319–324, 1977.

[4] S. Monaco, E. Gordon, M. W. Bowler, S. Delageniere, M. Guijarro, D. Spruce, O. Svensson, S. M. McSweeney, A. A. McCarthy, G. Leonard, and M. Hanao, "Automatic processing of macromolecular crystallography X-ray diffraction data at the ESRF," *Journal of Applied Crystallography*, vol. 46, no. 3, pp. 804–810, 2013.

[5] K. R. Acharya and M. D. Lloyd, "The advantages and limitations of protein crystal structures," *Trends in Pharmacological Sciences*, vol. 26, no. 1, pp. 10–14, 2005.

[6] G. T. Montelione, D. Zheng, Y. J. Huang, K. C. Gunsalus, and T. Szyperski, "Protein NMR spectroscopy in structural genomics," *Nature Structural & Molecular Biology*, vol. 7, pp. 982–985, 2000.

[7] M. Carroni and H. R. Saibil, "Cryo electron microscopy to determine the structure of macromolecular complexes," *Methods*, vol. 95, pp. 78–85, 2016.

[8] M. K. Higgins and S. M. Lea, "On the state of crystallography at the dawn of the electron microscopy revolution," *Current Opinion in Structural Biology*, vol. 46, pp. 95–101, 2017.

[9] A. Micsonai, F. Wien, L. Kernya, Y.-H. Lee, Y. Goto, M. Réfrégiers, and J. Kardos, "Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy," *Proceedings of the National Academy of Sciences*, vol. 112, no. 24, pp. E3095–E3103, 2015.

[10] T. D. Craggs and A. N. Kapanidis, "Six steps closer to FRET-driven structural biology," *Nature Methods*, vol. 9, no. 12, pp. 1157–1158, 2012.

[11] E. M. Puchner and H. E. Gaub, "Force and function: probing proteins with AFM-based force spectroscopy," *Current Opinion in Structural Biology*, vol. 19, no. 5, pp. 605–614, 2009.

[12] J. M. Plitzko, B. Schuler, and P. Selenko, "Structural biology outside the box–inside the cell," *Current Opinion in Structural Biology*, vol. 46, pp. 110–121, 2017.

[13] C. Schmidt and H. Urlaub, "Combining cryo-electron microscopy (cryo-EM) and cross-linking mass spectrometry (CX-MS) for structural elucidation of large protein assemblies," *Current Opinion in Structural Biology*, vol. 46, pp. 157–168, 2017.

[14] M. T. Mazhab-Jafari, A. Rohou, C. Schmidt, S. A. Bueler, S. Benlekbir, C. V. Robinson, and J. L. Rubinstein, "Atomic model for the membrane-embedded V_o motor of a eukaryotic V-ATPase," *Nature*, vol. 539, no. 7627, pp. 118–122, 2016.

[15] M. Mann, C. K. Meng, and J. B. Fenn, "Interpreting mass spectra of multiply charged ions," *Analytical Chemistry*, vol. 61, no. 15, pp. 1702–1708, 1989.

[16] M. Mann and O. N. Jensen, "Proteomic analysis of post-translational modifications," *Nature Biotechnology*, vol. 21, no. 3, pp. 255–261, 2003.

[17] J. Rappsilber, "The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes," *Journal of Structural Biology*, vol. 173, no. 3, pp. 530–540, 2011.

[18] A. J. Heck, "Native mass spectrometry: a bridge between interactomics and structural biology," *Nature Methods*, vol. 5, no. 11, pp. 927–933, 2008.

[19] N. P. Barrera, N. Di Bartolo, P. J. Booth, and C. V. Robinson, "Micelles protect membrane complexes from solution to vacuum," *Science*, vol. 321, no. 5886, pp. 243–246, 2008.

[20] V. Katta and B. T. Chait, "Observation of the heme-globin complex in native myoglobin by electrospray-ionization mass spectrometry," *Journal of the American Chemical Society*, vol. 113, no. 22, pp. 8534–8535, 1991.

[21] J. L. Benesch, B. T. Ruotolo, D. A. Simmons, and C. V. Robinson, "Protein complexes in the gas phase: technology for structural genomics and proteomics," *Chemical Reviews*, vol. 107, no. 8, pp. 3544–3567, 2007.

[22] R. H. van den Heuvel and A. J. Heck, "Native protein mass spectrometry: from intact oligomers to functional machineries," *Current Opinion in Chemical Biology*, vol. 8, no. 5, pp. 519–526, 2004.

[23] P. G. Wolynes, "Biomolecular folding in vacuo!!!(?)," *Proceedings of the National Academy of Sciences*, vol. 92, no. 7, pp. 2426–2427, 1995.

[24] M. Zhou, A. M. Sandercock, C. S. Fraser, G. Ridlova, E. Stephens, M. R. Schenauer, T. Yokoi-Fong, D. Barsky, J. A. Leary, J. W. Hershey, J. A. Doudna, and C. V. Robinson, "Mass spectrometry reveals modularity and a complete subunit interaction map of the eukaryotic translation factor eIF3," *Proceedings of the National Academy of Sciences*, vol. 105, no. 47, pp. 18139–18144, 2008.

[25] J. Snijder, R. J. Rose, D. Veesler, J. E. Johnson, and A. J. Heck, "Studying 18 MDa virus assemblies with native mass spectrometry," *Angewandte Chemie International Edition*, vol. 52, no. 14, pp. 4020–4023, 2013.

[26] A. Laganowsky, E. Reading, T. M. Allison, M. B. Ulmschneider, M. T. Degiacomi, A. J. Baldwin, and C. V. Robinson, "Membrane proteins bind lipids selectively to modulate their structure and function," *Nature*, vol. 510, no. 7503, pp. 172–175, 2014.

[27] J. A. Loo, "Studying noncovalent protein complexes by electrospray ionization mass spectrometry," *Mass Spectrometry Reviews*, vol. 16, no. 1, pp. 1–23, 1997.

[28] B. T. Ruotolo and C. V. Robinson, "Aspects of native proteins are retained in vacuum," *Current Opinion in Chemical Biology*, vol. 10, no. 5, pp. 402–408, 2006.

[29] B. T. Ruotolo, K. Giles, I. Campuzano, A. M. Sandercock, R. H. Bateman, and C. V. Robinson, "Evidence for macromolecular protein rings in the absence of bulk water," *Science*, vol. 310, no. 5754, pp. 1658–1661, 2005.

[30] M. Zhou, N. Morgner, N. P. Barrera, A. Politis, S. C. Isaacson, D. Matak-Vinković, T. Murata, R. A. Bernal, D. Stock, and C. V. Robinson, "Mass spectrometry of intact V-type ATPases reveals bound lipids and the effects of nucleotide binding," *Science*, vol. 334, no. 6054, pp. 380–385, 2011.

[31] T. Wyttenbach and M. T. Bowers, "Structural stability from solution to the gas phase: native solution structure of ubiquitin survives analysis in a solvent-free ion mobility–mass spectrometry environment," *The Journal of Physical Chemistry B*, vol. 115, no. 42, pp. 12266–12275, 2011.

[32] S. C. Wang, A. Politis, N. Di Bartolo, V. N. Bavro, S. J. Tucker, P. J. Booth, N. P. Barrera, and C. V. Robinson, "Ion mobility mass spectrometry of two tetrameric membrane protein complexes reveals compact structures and differences in stability and packing," *Journal of the American Chemical Society*, vol. 132, no. 44, pp. 15468–15470, 2010.

[33] Z. Ouyang, Z. Takáts, T. A. Blake, B. Gologan, A. J. Guymon, J. M. Wiseman, J. C. Oliver, V. J. Davisson, and R. G. Cooks, "Preparing protein microarrays by soft-landing of mass-selected ions," *Science*, vol. 301, no. 5638, pp. 1351–1354, 2003.

[34] J. Fernandez-Martinez, S. J. Kim, Y. Shi, P. Upla, R. Pellarin, M. Gagnon, I. E. Chemmama, J. Wang, I. Nudelman, W. Zhang, R. Williams, W. Rice, D. L. Stokes, D. Zenklusen, B. T. Chait, A. Sali, and M. P. Rout, "Structure and function of the nuclear pore complex cytoplasmic mRNA export platform," *Cell*, vol. 167, no. 5, pp. 1215–1228, 2016.

[35] M. M. Jore, S. Johnson, D. Sheppard, N. M. Barber, Y. I. Li, M. A. Nunn, H. Elmlund, and S. M. Lea, "Structural basis for therapeutic inhibition of complement C5," *Nature Structural & Molecular Biology*, vol. 23, no. 5, pp. 378–386, 2016.

[36] C. Rouillon, M. Zhou, J. Zhang, A. Politis, V. Beilsten-Edmands, G. Cannone, S. Graham, C. V. Robinson, L. Spagnolo, and M. F. White, "Structure of the CRISPR interference complex CSM reveals key similarities with cascade," *Molecular Cell*, vol. 52, no. 1, pp. 124–134, 2013.

[37] K. Lasker, F. Förster, S. Bohn, T. Walzthoeni, E. Villa, P. Unverdorben, F. Beck, R. Aebersold, A. Sali, and W. Baumeister, "Molecular architecture of the 26S proteasome holocomplex determined by an integrative approach," *Proceedings of the National Academy of Sciences*, vol. 109, no. 5, pp. 1380–1387, 2012.

[38] R. B. Fenwick, H. van den Bedem, J. S. Fraser, and P. E. Wright, "Integrated description of protein dynamics from room-temperature X-ray crystallography and NMR," *Proceedings of the National Academy of Sciences*, vol. 111, no. 4, pp. E445–E454, 2014.

[39] M. Aznauryan, L. Delgado, A. Soranno, D. Nettels, J.-r. Huang, A. M. Labhardt, S. Grzesiek, and B. Schuler, "Comprehensive structural and dynamical view of an unfolded protein from the combination of single-molecule FRET, NMR, and SAXS," *Proceedings of the National Academy of Sciences*, vol. 113, no. 37, pp. E5389–E5398, 2016.

[40] M. G. Iadanza, A. J. Higgins, B. Schiffrin, A. N. Calabrese, D. J. Brockwell, A. E. Ashcroft, S. E. Radford, and N. A. Ranson, "Lateral opening in the intact β-barrel assembly machinery captured by cryo-EM," *Nature Communications*, vol. 7, 2016.

[41] V. C. Wasinger, S. J. Cordwell, A. Cerpa-Poljak, J. X. Yan, A. A. Gooley, M. R. Wilkins, M. W. Duncan, R. Harris, K. L. Williams, and I. Humphery-Smith, "Progress with gene-product mapping of the Mollicutes: *Mycoplasma genitalium*," *Electrophoresis*, vol. 16, no. 1, pp. 1090–1094, 1995.

[42] P. James, "Protein identification in the post-genome era: the rapid rise of proteomics," *Quarterly Reviews of Biophysics*, vol. 30, no. 4, pp. 279–331, 1997.

[43] A. Shevchenko, O. N. Jensen, A. V. Podtelejnikov, F. Sagliocco, M. Wilm, O. Vorm, P. Mortensen, A. Shevchenko, H. Boucherie, and M. Mann, "Linking

genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels," *Proceedings of the National Academy of Sciences*, vol. 93, no. 25, pp. 14440–14445, 1996.

[44] M. van de Waterbeemd, K. L. Fort, D. Boll, M. Reinhardt-Szyba, A. Routh, A. Makarov, and A. J. Heck, "High-fidelity mass analysis unveils heterogeneity in intact ribosomal particles," *Nature Methods*, vol. 14, no. 3, pp. 283–286, 2017.

[45] G. Taylor, "Disintegration of water drops in an electric field," in *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 280, pp. 383–397, The Royal Society, 1964.

[46] M. Dole, L. Mack, R. Hines, R. Mobley, L. Ferguson, and M. d. Alice, "Molecular beams of macroions," *The Journal of Chemical Physics*, vol. 49, no. 5, pp. 2240–2249, 1968.

[47] C. Meng, M. Mann, and J. Fenn, "Of protons or proteins," *Zeitschrift für Physik D Atoms, Molecules and Clusters*, vol. 10, no. 2-3, pp. 361–368, 1988.

[48] C. Whitehouse, R. Dreyer, M. Yamashita, and J. Fenn, "Electrospray ionization for mass-spectrometry of large biomolecules," *Science*, vol. 246, no. 4926, pp. 64–71, 1989.

[49] P. Kebarle and U. H. Verkerk, "Electrospray: from ions in solution to ions in the gas phase, what we know now," *Mass Spectrometry Reviews*, vol. 28, no. 6, pp. 898–917, 2009.

[50] D. P. Smith, "The electrohydrodynamic atomization of liquids," *IEEE Transactions on Industry Applications*, no. 3, pp. 527–535, 1986.

[51] J. Iribarne and B. Thomson, "On the evaporation of small ions from charged droplets," *The Journal of Chemical Physics*, vol. 64, no. 6, pp. 2287–2294, 1976.

[52] L. Konermann, E. Ahadi, A. D. Rodriguez, and S. Vahidi, "Unraveling the mechanism of electrospray ionization," *Analytical Chemistry*, vol. 85, no. 1, pp. 2–9, 2013.

[53] T. M. Larsen, L. T. Laughlin, H. M. Holden, I. Rayment, and G. H. Reed, "Structure of rabbit muscle pyruvate kinase complexed with Mn^{2+}, K^{+}, and pyruvate," *Biochemistry*, vol. 33, no. 20, pp. 6301–6309, 1994.

[54] J. F. De La Mora, "Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism," *Analytica Chimica Acta*, vol. 406, no. 1, pp. 93–104, 2000.

[55] M. S. Wilm and M. Mann, "Electrospray and Taylor-cone theory, Dole's beam of macromolecules at last?," *International Journal of Mass Spectrometry and Ion Processes*, vol. 136, no. 2-3, pp. 167–180, 1994.

[56] F. Sobott and C. V. Robinson, "Characterising electrosprayed biomolecules using tandem-MS–the noncovalent GroEL chaperonin assembly," *International Journal of Mass Spectrometry*, vol. 236, no. 1, pp. 25–32, 2004.

[57] A. R. McKay, B. T. Ruotolo, L. L. Ilag, and C. V. Robinson, "Mass measurements of increased accuracy resolve heterogeneous populations of intact ribosomes," *Journal of the American Chemical Society*, vol. 128, no. 35, pp. 11433–11442, 2006.

[58] F. Sobott, H. Hernández, M. G. McCammon, M. A. Tito, and C. V. Robinson, "A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies," *Analytical Chemistry*, vol. 74, no. 6, pp. 1402–1407, 2002.

[59] D. J. Douglas, A. J. Frank, and D. Mao, "Linear ion traps in mass spectrometry," *Mass Spectrometry Reviews*, vol. 24, no. 1, pp. 1–29, 2005.

[60] W. Paul and H. Steinwedel, "A new mass spectrometer without a magnetic field," *Zeitschrift fur Naturforschung - Section A Journal of Physical Sciences*, vol. 8, 1953.

[61] S. A. Hofstadler, R. Bakhtiar, and R. D. Smith, "Electrospray ionization mass spectroscopy: Part I. Instrumentation and spectral interpretation," *Journal of Chemical Education*, vol. 73, no. 4, p. A82, 1996.

[62] R. J. Cotter, "Time-of-flight mass spectrometry for the structural analysis of biological molecules," *Analytical Chemistry*, vol. 64, no. 21, pp. 1027A–1039A, 1992.

[63] A. Makarov, "Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis," *Analytical Chemistry*, vol. 72, no. 6, pp. 1156–1162, 2000.

[64] K. Kingdon, "A method for the neutralization of electron space charge by positive ionization at very low gas pressures," *Physical Review*, vol. 21, no. 4, p. 408, 1923.

[65] Q. Hu, R. J. Noll, H. Li, A. Makarov, M. Hardman, and R. Graham Cooks, "The Orbitrap: a new mass spectrometer," *Journal of Mass Spectrometry*, vol. 40, no. 4, pp. 430–443, 2005.

[66] M. Scigelova and A. Makarov, "Orbitrap mass analyzer–overview and applications in proteomics," *Proteomics*, vol. 6, no. S2, pp. 16–21, 2006.

[67] J. L. Benesch, J. A. Aquilina, B. T. Ruotolo, F. Sobott, and C. V. Robinson, "Tandem mass spectrometry reveals the quaternary organization of macromolecular assemblies," *Chemistry & Biology*, vol. 13, no. 6, pp. 597–605, 2006.

[68] K. Breuker and F. W. McLafferty, "Native electron capture dissociation for the structural characterization of noncovalent interactions in native cytochrome c," *Angewandte Cchemie International Edition*, vol. 42, no. 40, pp. 4900–4904, 2003.

[69] J. E. Syka, J. J. Coon, M. J. Schroeder, J. Shabanowitz, and D. F. Hunt, "Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry," *Proceedings of the National Academy of Sciences*, vol. 101, no. 26, pp. 9528–9533, 2004.

[70] V. H. Wysocki, C. M. Jones, A. S. Galhena, and A. E. Blackwell, "Surface-induced dissociation shows potential to be more informative than collision-induced dissociation for structural studies of large systems," *Journal of the American Society for Mass Spectrometry*, vol. 19, no. 7, pp. 903–913, 2008.

[71] J. L. Benesch, "Collisional activation of protein complexes: picking up the pieces," *Journal of the American Society for Mass Spectrometry*, vol. 20, no. 3, pp. 341–348, 2009.

[72] J. V. Olsen, B. Macek, O. Lange, A. Makarov, S. Horning, and M. Mann, "Higher-energy C-trap dissociation for peptide modification analysis," *Nature Methods*, vol. 4, no. 9, pp. 709–713, 2007.

[73] P. Roepstorff and J. Fohlman, "Proposal for a common nomenclature for sequence ions in mass spectra of peptides," *Biological Mass Spectrometry*, vol. 11, no. 11, pp. 601–601, 1984.

[74] K. Biemann, "Contributions of mass spectrometry to peptide and protein structure," *Biological Mass Spectrometry*, vol. 16, no. 1-12, pp. 99–111, 1988.

[75] I. V. Chernushevich, A. V. Loboda, and B. A. Thomson, "An introduction to quadrupole–time-of-flight mass spectrometry," *Journal of Mass Spectrometry*, vol. 36, no. 8, pp. 849–865, 2001.

[76] M. Bantscheff, M. Schirle, G. Sweetman, J. Rick, and B. Kuster, "Quantitative mass spectrometry in proteomics: a critical review," *Analytical and Bioanalytical Chemistry*, vol. 389, no. 4, pp. 1017–1031, 2007.

[77] A.-C. Gingras, R. Aebersold, and B. Raught, "Advances in protein complex analysis using mass spectrometry," *The Journal of Physiology*, vol. 563, no. 1, pp. 11–21, 2005.

[78] S. A. Gerber, J. Rush, O. Stemman, M. W. Kirschner, and S. P. Gygi, "Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS," *Proceedings of the National Academy of Sciences*, vol. 100, no. 12, pp. 6940–6945, 2003.

[79] D. S. Kirkpatrick, S. A. Gerber, and S. P. Gygi, "The absolute quantification strategy: a general procedure for the quantification of proteins and post-translational modifications," *Methods*, vol. 35, no. 3, pp. 265–273, 2005.

[80] Y. Oda, K. Huang, F. Cross, D. Cowburn, and B. Chait, "Accurate quantitation of protein expression and site-specific phosphorylation," *Proceedings of the National Academy of Sciences*, vol. 96, no. 12, pp. 6591–6596, 1999.

[81] S.-E. Ong, B. Blagoev, I. Kratchmarova, D. B. Kristensen, H. Steen, A. Pandey, and M. Mann, "Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics," *Molecular & Cellular Proteomics*, vol. 1, no. 5, pp. 376–386, 2002.

[82] M. Thevis, A. A. Makarov, S. Horning, and W. Schänzer, "Mass spectrometry of stanozolol and its analogues using electrospray ionization and collision-induced dissociation with quadrupole-linear ion trap and linear ion trap-orbitrap hybrid mass analyzers," *Rapid Communications in Mass Spectrometry*, vol. 19, no. 22, pp. 3369–3378, 2005.

[83] J. Gault, J. A. Donlan, I. Liko, J. T. Hopper, K. Gupta, N. G. Housden, W. B. Struwe, M. T. Marty, T. Mize, C. Bechara, Y. Zhu, B. Wu, C. Kleanthous, M. Belov, E. Damoc, A. Makarov, and C. V. Robinson, "High-resolution mass spectrometry of small molecules bound to membrane proteins," *Nature Methods*, vol. 13, no. 4, pp. 333–336, 2016.

[84] A. Dobo and I. A. Kaltashov, "Detection of multiple protein conformational ensembles in solution via deconvolution of charge-state distributions in ESI MS," *Analytical Chemistry*, vol. 73, no. 20, pp. 4763–4773, 2001.

[85] U. A. Mirza and B. T. Chait, "Effects of anions on the positive ion electrospray ionization mass spectra of peptides and proteins," *Analytical Chemistry*, vol. 66, no. 18, pp. 2898–2904, 1994.

[86] A. T. Iavarone, J. C. Jurchen, and E. R. Williams, "Effects of solvent on the maximum charge state and charge state distribution of protein ions produced by electrospray ionization," *Journal of the American Society for Mass Spectrometry*, vol. 11, no. 11, pp. 976–985, 2000.

[87] H. J. Sterling, C. A. Cassou, A. C. Susa, and E. R. Williams, "Electrothermal supercharging of proteins in native electrospray ionization," *Analytical Chemistry*, vol. 84, no. 8, pp. 3795–3801, 2012.

[88] Y. Li and R. B. Cole, "Shifts in peptide and protein charge state distributions with varying spray tip orifice diameter in nanoelectrospray fourier transform ion cycltron resonance mass spectrometry," *Analytical Chemistry*, vol. 75, no. 21, pp. 5739–5746, 2003.

[89] A. T. James and u. A. Martin, "Gas-liquid partition chromatography: the separation and micro-estimation of volatile fatty acids from formic acid to dodecanoic acid," *Biochemical Journal*, vol. 50, no. 5, p. 679, 1952.

[90] K. Dettmer-Wilde and W. Engewald, *Practical Gas Chromatography*. Springer, 2016.

[91] J. J. Van Deemter, F. Zuiderweg, and A. v. Klinkenberg, "Longitudinal diffusion and resistance to mass transfer as causes of nonideality in chromatography," *Chemical Engineering Science*, vol. 5, no. 6, pp. 271–289, 1956.

[92] T. E. Thingholm, T. J. Jørgensen, O. N. Jensen, and M. R. Larsen, "Highly selective enrichment of phosphorylated peptides using titanium dioxide," *Nature Protocols*, vol. 1, no. 4, p. 1929, 2006.

[93] C. Hage, C. Iacobucci, A. Rehkamp, C. Arlt, and A. Sinz, "The first zero-length mass spectrometry-cleavable cross-linker for protein structure analysis," *Angewandte Chemie*, vol. 129, no. 46, pp. 14743–14747, 2017.

[94] C. Arlt, C. H. Ihling, and A. Sinz, "Structure of full-length p53 tumor suppressor probed by chemical cross-linking and mass spectrometry," *Proteomics*, vol. 15, no. 16, pp. 2746–2755, 2015.

[95] Z. A. Chen, A. Jawhari, L. Fischer, C. Buchen, S. Tahir, T. Kamenski, M. Rasmussen, L. Lariviere, J.-C. Bukowski-Wills, M. Nilges, P. Cramer, and J. Rappsilber, "Architecture of the RNA polymerase II–TFIIF complex revealed by cross-linking and mass spectrometry," *The EMBO Journal*, vol. 29, no. 4, pp. 717–726, 2010.

[96] C. Schmidt, J. A. Macpherson, A. M. Lau, K. W. Tan, F. Fraternali, and A. Politis, "Surface accessibility and dynamics of macromolecular assemblies probed by covalent labeling mass spectrometry and integrative modeling," *Analytical Chemistry*, vol. 89, no. 3, pp. 1459–1468, 2017.

[97] C. Schmidt and C. V. Robinson, "A comparative cross-linking strategy to probe conformational changes in protein complexes," *Nature Protocols*, vol. 9, no. 9, pp. 2224–2236, 2014.

[98] Z. Chen, L. Fischer, S. Tahir, J.-C. Bukowski-Wills, P. Barlow, and J. Rappsilber, "Quantitative cross-linking/mass spectrometry reveals subtle protein conformational changes," *Wellcome Open Research*, vol. 1, 2016.

[99] C. Schmidt, M. Zhou, H. Marriott, N. Morgner, A. Politis, and C. V. Robinson, "Comparative cross-linking and mass spectrometry of an intact F-type ATPase suggest a role for phosphorylation," *Nature Communications*, vol. 4, 2013.

[100] N. Morgner, C. Schmidt, V. Beilsten-Edmands, I.-o. Ebong, N. A. Patel, E. M. Clerico, E. Kirschke, S. Daturpalli, S. E. Jackson, D. Agard, and C. V.

Robinson, "Hsp70 forms antiparallel dimers stabilized by post-translational modifications to position clients for transfer to Hsp90," *Cell Reports*, vol. 11, no. 5, pp. 759–769, 2015.

[101] H. Xu, P.-H. Hsu, L. Zhang, M.-D. Tsai, and M. A. Freitas, "Database search algorithm for identification of intact cross-links in proteins and peptides using tandem mass spectrometry," *Journal of Proteome Research*, vol. 9, no. 7, pp. 3384–3393, 2010.

[102] B. Yang, Y.-J. Wu, M. Zhu, S.-B. Fan, J. Lin, K. Zhang, S. Li, H. Chi, Y.-X. Li, H.-F. Chen, S.-K. Luo, Y.-H. Ding, L.-H. Wang, Z. Hao, L.-Y. Xiu, S. Chen, K. Ye, S.-M. He, and M.-Q. Dong, "Identification of cross-linked peptides from complex samples," *Nature Methods*, vol. 9, no. 9, pp. 904–906, 2012.

[103] M. Götze, J. Pettelkau, S. Schaks, K. Bosse, C. H. Ihling, F. Krauth, R. Fritzsche, U. Kühn, and A. Sinz, "StavroX–a software for analyzing crosslinked products in protein interaction studies," *Journal of the American Society for Mass Spectrometry*, vol. 23, no. 1, pp. 76–87, 2012.

[104] C. Schmidt, V. Beilsten-Edmands, and C. V. Robinson, "The joining of the Hsp90 and Hsp70 chaperone cycles yields transient interactions and stable intermediates: insights from mass spectrometry," *Oncotarget*, vol. 6, no. 21, p. 18276, 2015.

[105] C. B. Anfinsen, "Principles that govern the folding of protein chains," *Science*, vol. 181, no. 4096, pp. 223–230, 1973.

[106] F. U. Hartl, A. Bracher, and M. Hayer-Hartl, "Molecular chaperones in protein folding and proteostasis," *Nature*, vol. 475, no. 7356, pp. 324–332, 2011.

[107] C. Levinthal, "Are there pathways for protein folding?," *Journal de Chimie Physique*, vol. 65, pp. 44–45, 1968.

[108] C. Levinthal, "How to fold graciously," *Mossbauer Spectroscopy in Biological Systems*, vol. 67, pp. 22–24, 1969.

[109] M. Karplus, "Behind the folding funnel diagram," *Nature Chemical Biology*, vol. 7, no. 7, pp. 401–404, 2011.

[110] S. S. Plotkin and J. N. Onuchic, "Understanding protein folding with energy landscape theory part i: basic concepts," *Quarterly Reviews of Biophysics*, vol. 35, no. 2, pp. 111–167, 2002.

[111] R. Zwanzig, A. Szabo, and B. Bagchi, "Levinthal's paradox," *Proceedings of the National Academy of Sciences*, vol. 89, no. 1, pp. 20–22, 1992.

[112] G. Srinivas and B. Bagchi, "Study of the dynamics of protein folding through minimalistic models," *Theoretical Chemistry Accounts*, vol. 109, no. 1, pp. 8–21, 2003.

[113] R. Herbst, U. Schäfer, and R. Seckler, "Equilibrium intermediates in the reversible unfolding of firefly *Photinus pyralis* luciferase," *Journal of Biological Chemistry*, vol. 272, no. 11, pp. 7099–7105, 1997.

[114] P. L. Privalov, "Intermediate states in protein folding," *Journal of Molecular Biology*, vol. 258, no. 5, pp. 707–725, 1996.

[115] F. Chiti and C. M. Dobson, "Protein misfolding, functional amyloid, and human disease," *Annual Review of Biochemistry*, vol. 75, pp. 333–366, 2006.

[116] R. J. Ellis and A. P. Minton, "Protein aggregation in crowded environments," *Biological Chemistry*, vol. 387, no. 5, pp. 485–497, 2006.

[117] S. R. McGuffee and A. H. Elcock, "Atomically detailed simulations of concentrated protein solutions: the effects of salt, pH, point mutations, and protein concentration in simulations of 1000-molecule systems," *Journal of the American Chemical Society*, vol. 128, no. 37, pp. 12098–12110, 2006.

[118] J. Hong and L. M. Gierasch, "Macromolecular crowding remodels the energy landscape of a protein by favoring a more compact unfolded state," *Journal of the American Chemical Society*, vol. 132, no. 30, pp. 10445–10452, 2010.

[119] R. I. Morimoto, "Proteotoxic stress and inducible chaperone networks in neurodegenerative disease and aging," *Genes & Development*, vol. 22, no. 11, pp. 1427–1438, 2008.

[120] R. Brookmeyer, S. Gray, and C. Kawas, "Projections of Alzheimer's disease in the United States and the public health impact of delaying disease onset," *American Journal of Public Health*, vol. 88, no. 9, pp. 1337–1342, 1998.

[121] Y. E. Kim, M. S. Hipp, A. Bracher, M. Hayer-Hartl, and F. Ulrich Hartl, "Molecular chaperone functions in protein folding and proteostasis," *Annual Review of Biochemistry*, vol. 82, pp. 323–355, 2013.

[122] M. J. Vos, J. Hageman, S. Carra, and H. H. Kampinga, "Structural and functional diversities between members of the human HSPB, HSPH, HSPA, and DNAJ chaperone families," *Biochemistry*, vol. 47, no. 27, pp. 7001–7011, 2008.

[123] H. Saibil, "Chaperone machines for protein folding, unfolding and disaggregation," *Nature Reviews Molecular Cell Biology*, vol. 14, no. 10, pp. 630–642, 2013.

[124] E. B. Bertelsen, L. Chang, J. E. Gestwicki, and E. R. Zuiderweg, "Solution conformation of wild-type *E. coli* Hsp70 (DnaK) chaperone complexed with ADP and substrate," *Proceedings of the National Academy of Sciences*, vol. 106, no. 21, pp. 8471–8476, 2009.

[125] T. R. Alderson, J. H. Kim, and J. L. Markley, "Dynamical structures of Hsp70 and Hsp70-Hsp40 complexes," *Structure*, vol. 24, no. 7, pp. 1014–1030, 2016.

[126] A. A. Michels, B. Kanon, O. Bensaude, and H. H. Kampinga, "Heat shock protein (Hsp) 40 mutants inhibit Hsp70 in mammalian cells," *Journal of Biological Chemistry*, vol. 274, no. 51, pp. 36757–36763, 1999.

[127] T. H. Nguyen, D. Law, and D. B. Williams, "Binding protein BiP is required for translocation of secretory proteins into the endoplasmic reticulum in *Saccharomyces cerevisiae*," *Proceedings of the National Academy of Sciences*, vol. 88, no. 4, pp. 1565–1569, 1991.

[128] Y. Shi and J. Thomas, "The transport of proteins into the nucleus requires the 70-kilodalton heat shock protein or its cytosolic cognate.," *Molecular and Cellular Biology*, vol. 12, no. 5, pp. 2186–2192, 1992.

[129] S.-H. Park, N. Bolender, F. Eisele, Z. Kostova, J. Takeuchi, P. Coffino, and D. H. Wolf, "The cytoplasmic Hsp70 chaperone machinery subjects misfolded and endoplasmic reticulum import-incompetent proteins to degradation via the ubiquitin–proteasome system," *Molecular Biology of the Cell*, vol. 18, no. 1, pp. 153–165, 2007.

[130] M. Mayer and B. Bukau, "Hsp70 chaperones: cellular functions and molecular mechanism," *Cellular and Molecular Life Sciences*, vol. 62, no. 6, p. 670, 2005.

[131] K. D. Dittmar, K. A. Hutchison, J. K. Owens-Grillo, and W. B. Pratt, "Reconstitution of the steroid receptor· hsp90 heterocomplex assembly system of rabbit reticulocyte lysate," *Journal of Biological Chemistry*, vol. 271, no. 22, pp. 12833–12839, 1996.

[132] H. Kosano, B. Stensgard, M. C. Charlesworth, N. McMahon, and D. Toft, "The assembly of progesterone receptor-hsp90 complexes using purified proteins," *Journal of Biological Chemistry*, vol. 273, no. 49, pp. 32973–32979, 1998.

[133] F. A. Aprile, A. Dhulesia, F. Stengel, C. Roodveldt, J. L. Benesch, P. Tortora, C. V. Robinson, X. Salvatella, C. M. Dobson, and N. Cremades, "Hsp70 oligomerization is mediated by an interaction between the interdomain linker and the substrate-binding domain," *PLoS One*, vol. 8, no. 6, p. e67961, 2013.

[134] C. E. Angelidis, I. Lazaridis, and G. N. Pagoulatos, "Aggregation of hsp70 and hsc70 *in vivo* is distinct and temperature-dependent and their chaperone function is directly related to non-aggregated forms," *The FEBS Journal*, vol. 259, no. 1-2, pp. 505–512, 1999.

[135] T. K. Nemoto, Y. Fukuma, H. Itoh, T. Takagi, and T. Ono, "A disulfide bridge mediated by cysteine 574 is formed in the dimer of the 70-kDa heat shock protein," *Journal of Biochemistry*, vol. 139, no. 4, pp. 677–687, 2006.

[136] J. V. Anderson, D. W. Haskell, and C. L. Guy, "Differential influence of ATP on native spinach 70-kilodalton heat-shock cognates," *Plant Physiology*, vol. 104, no. 4, pp. 1371–1380, 1994.

[137] N. Benaroudj, F. Triniolles, and M. M. Ladjimi, "Effect of nucleotides, peptides, and unfolded proteins on the self-association of the molecular chaperone HSC70," *Journal of Biological Chemistry*, vol. 271, no. 31, pp. 18471–18476, 1996.

[138] N. Benaroudj, B. Fouchaq, and M. M. Ladjimi, "The COOH-terminal peptide binding domain is essential for self-association of the molecular chaperone HSC70," *Journal of Biological Chemistry*, vol. 272, no. 13, pp. 8744–8751, 1997.

[139] B. Fouchaq, N. Benaroudj, C. Ebel, and M. M. Ladjimi, "Oligomerization of the 17-kDa peptide-binding domain of the molecular chaperone HSC70," *The FEBS Journal*, vol. 259, no. 1-2, pp. 379–384, 1999.

[140] P. T. Guidon Jr and L. E. Hightower, "Purification and initial characterization of the 71-kilodalton rat heat-shock protein and its cognate as fatty acid binding proteins," *Biochemistry*, vol. 25, no. 11, pp. 3231–3239, 1986.

[141] D. Kim, Y. J. Lee, and P. M. Corry, "Constitutive Hsp70: oligomerization and its dependence on ATP binding," *Journal of Cellular Physiology*, vol. 153, no. 2, pp. 353–361, 1992.

[142] M. Ohno, N. Kitabatake, and F. Tani, "Role of the C-terminal region of mouse inducible Hsp72 in the recognition of peptide substrate for chaperone activity," *FEBS Letters*, vol. 576, no. 3, pp. 381–386, 2004.

[143] D. R. Palleros, W. J. Welch, and A. L. Fink, "Interaction of hsp70 with unfolded proteins: effects of temperature and nucleotides on the kinetics of binding," *Proceedings of the National Academy of Sciences*, vol. 88, no. 13, pp. 5719–5723, 1991.

[144] M. Chevalier, L. King, C. Wang, M.-J. Gething, E. Elguindi, and S. Y. Blond, "Substrate binding induces depolymerization of the C-terminal peptide binding domain of murine GRP78/BiP," *Journal of Biological Chemistry*, vol. 273, no. 41, pp. 26827–26835, 1998.

[145] S. Blond-Elguindi, A. Fourie, J. Sambrook, and M. Gething, "Peptide-dependent stimulation of the ATPase activity of the molecular chaperone BiP is the result of conversion of oligomers to active monomers," *Journal of Biological Chemistry*, vol. 268, no. 17, pp. 12730–12735, 1993.

[146] A. Carlino, H. Toledo, D. Skaleris, R. DeLisio, H. Weissbach, and N. Brot, "Interactions of liver Grp78 and *Escherichia coli* recombinant Grp78 with ATP: multiple species and disaggregation," *Proceedings of the National Academy of Sciences*, vol. 89, no. 6, pp. 2081–2085, 1992.

[147] P. J. Freiden, J. Gaut, and L. Hendershot, "Interconversion of three differentially modified and assembled forms of BiP," *The EMBO Journal*, vol. 11, no. 1, p. 63, 1992.

[148] B. Gao, E. Eisenberg, and L. Greene, "Effect of constitutive 70-kDa heat shock protein polymerization on its interaction with protein substrate," *Journal of Biological Chemistry*, vol. 271, no. 28, pp. 16792–16797, 1996.

[149] A. D. Thompson, S. M. Bernard, G. Skiniotis, and J. E. Gestwicki, "Visualization and functional analysis of the oligomeric states of *Escherichia coli* heat shock protein 70 (Hsp70/DnaK)," *Cell Stress and Chaperones*, vol. 17, no. 3, pp. 313–327, 2012.

[150] H.-J. Schönfeld, D. Schmidt, H. Schröder, and B. Bukau, "The DnaK chaperone system of *Escherichia coli*: quaternary structures and interactions of the DnaK and GrpE components," *Journal of Biological Chemistry*, vol. 270, no. 5, pp. 2183–2189, 1995.

[151] E. B. Sarbeng, Q. Liu, X. Tian, J. Yang, H. Li, J. L. Wong, L. Zhou, and Q. Liu, "A functional DnaK dimer is essential for the efficient interaction with Hsp40 heat shock protein," *Journal of Biological Chemistry*, vol. 290, no. 14, pp. 8849–8862, 2015.

[152] R. Qi, E. B. Sarbeng, Q. Liu, K. Q. Le, X. Xu, H. Xu, J. Yang, J. L. Wong, C. Vorvis, W. A. Hendrickson, L. Zhou, and L. Qinglian, "Allosteric opening of the polypeptide-binding site when an Hsp70 binds ATP," *Nature Structural & Molecular Biology*, vol. 20, no. 7, pp. 900–907, 2013.

[153] N. M. Bonini, "Chaperoning brain degeneration," *Proceedings of the National Academy of Sciences*, vol. 99, no. suppl 4, pp. 16407–16411, 2002.

[154] J. Klucken, Y. Shin, E. Masliah, B. T. Hyman, and P. J. McLean, "Hsp70 reduces α-synuclein aggregation and toxicity," *Journal of Biological Chemistry*, vol. 279, no. 24, pp. 25497–25502, 2004.

[155] R. Ambra, E. Mocchegiani, R. Giacconi, R. Canali, A. Rinna, M. Malavolta, and F. Virgili, "Characterization of the hsp70 response in lymphoblasts from aged and centenarian subjects and differential effects of in vitro zinc supplementation," *Experimental Gerontology*, vol. 39, no. 10, pp. 1475–1484, 2004.

[156] A. R. Goloudina, O. N. Demidov, and C. Garrido, "Inhibition of HSP70: a challenging anti-cancer strategy," *Cancer Letters*, vol. 325, no. 2, pp. 117–124, 2012.

[157] J. Nylandsted, M. Rohde, K. Brand, L. Bastholm, F. Elling, and M. Jäättelä, "Selective depletion of heat shock protein 70 (Hsp70) activates a tumor-specific death program that is independent of caspases and bypasses Bcl-2," *Proceedings of the National Academy of Sciences*, vol. 97, no. 14, pp. 7871–7876, 2000.

[158] J.-J. Leu, J. Pimkina, A. Frank, M. E. Murphy, and D. L. George, "A small molecule inhibitor of inducible heat shock protein 70," *Molecular Cell*, vol. 36, no. 1, pp. 15–27, 2009.

[159] P. Cloutier and B. Coulombe, "Regulation of molecular chaperones through post-translational modifications: decrypting the chaperone code," *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, vol. 1829, no. 5, pp. 443–454, 2013.

[160] A. Knowlton, M. Grenier, S. Kirchhoff, and M. Salfity, "Phosphorylation at tyrosine-524 influences nuclear accumulation of HSP72 with heat stress," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 278, no. 6, pp. H2143–H2149, 2000.

[161] P. Muller, E. Ruckova, P. Halada, P. Coates, R. Hrstka, D. Lane, and B. Vojtesek, "C-terminal phosphorylation of Hsp70 and Hsp90 regulates alternate binding to co-chaperones CHIP and HOP to determine cellular protein folding/degradation balances," *Oncogene*, vol. 32, no. 25, pp. 3101–3110, 2013.

[162] I. Grummt, "Life on a planet of its own: regulation of RNA polymerase I transcription in the nucleolus," *Genes & Development*, vol. 17, no. 14, pp. 1691–1702, 2003.

[163] H.-S. Cho, T. Shimazu, G. Toyokawa, Y. Daigo, Y. Maehara, S. Hayami, A. Ito, K. Masuda, N. Ikawa, H. I. Field, E. Tsuchiya, O. Shin-ichi, B. A. Ponder, M. Yoshida, Y. Nakamura, and R. Hamamoto, "Enhanced HSP70 lysine methylation promotes proliferation of cancer cells through activation of Aurora kinase B," *Nature Communications*, vol. 3, p. 1072, 2012.

[164] J. H. Seo, J.-H. Park, E. J. Lee, T. T. L. Vo, H. Choi, J. Y. Kim, J. K. Jang, H.-J. Wee, H. S. Lee, S. H. Jang, *et al.*, "ARD1-mediated Hsp70 acetylation balances stress-induced protein refolding and degradation," *Nature Communications*, vol. 7, p. 12882, 2016.

[165] H. Hernández and C. V. Robinson, "Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry," *Nature Protocols*, vol. 2, no. 3, p. 715, 2007.

[166] B. M. Babior, "NADPH oxidase: an update," *Blood*, vol. 93, no. 5, pp. 1464–1476, 1999.

[167] H. Nishi, K. Hashimoto, and A. R. Panchenko, "Phosphorylation in protein-protein binding: effect on stability and function," *Structure*, vol. 19, no. 12, pp. 1807–1815, 2011.

[168] J. Seo and K.-J. Lee, "Post-translational modifications and their biological functions: proteomic analysis and systematic approaches," *Journal of Biochemistry and Molecular Biology*, vol. 37, no. 1, pp. 35–44, 2004.

[169] V. van Noort, J. Seebacher, S. Bader, S. Mohammed, I. Vonkova, M. J. Betts, S. Kühner, R. Kumar, T. Maier, M. O'Flaherty, V. Rybin, A. Schmeisky, E. Yus, J. Stlke, L. Serrano, R. B. Russell, A. J. Heck, P. Bork, and A.-C. Gavin, "Cross-talk between phosphorylation and lysine acetylation in a genome-reduced bacterium," *Molecular Systems Biology*, vol. 8, no. 1, p. 571, 2012.

[170] P. Beltrao, V. Albanèse, L. R. Kenner, D. L. Swaney, A. Burlingame, J. Villén, W. A. Lim, J. S. Fraser, J. Frydman, and N. J. Krogan, "Systematic functional prioritization of protein posttranslational modifications," *Cell*, vol. 150, no. 2, pp. 413–425, 2012.

[171] M. T. Marty, A. J. Baldwin, E. G. Marklund, G. K. Hochberg, J. L. Benesch, and C. V. Robinson, "Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles," *Analytical Chemistry*, vol. 87, no. 8, pp. 4370–4376, 2015.

[172] M. P. Mayer, H. Schröder, S. Rüdiger, K. Paal, T. Laufen, and B. Bukau, "Multistep mechanism of substrate binding determines chaperone activity of Hsp70," *Nature Structural & Molecular Biology*, vol. 7, no. 7, pp. 586–593, 2000.

[173] P. Zhang, I. Julia, J. Leu, M. E. Murphy, D. L. George, and R. Marmorstein, "Crystal structure of the stress-inducible human heat shock protein 70 substrate-binding domain in complex with peptide substrate," *PLoS One*, vol. 9, no. 7, p. e103518, 2014.

[174] A. L. Lai, E. M. Clerico, M. E. Blackburn, N. A. Patel, C. V. Robinson, P. P. Borbat, J. H. Freed, and L. M. Gierasch, "Key features of an Hsp70 chaperone allosteric landscape revealed by ion-mobility native mass spectrometry and double electron-electron resonance," *Journal of Biological Chemistry*, vol. 292, no. 21, pp. 8773–8785, 2017.

[175] Q. Liu and W. A. Hendrickson, "Insights into Hsp70 chaperone activity from a crystal structure of the yeast Hsp110 Sse1," *Cell*, vol. 131, no. 1, pp. 106–120, 2007.

[176] M. Zahn, N. Berthold, B. Kieslich, D. Knappe, R. Hoffmann, and N. Sträter, "Structural studies on the forward and reverse binding modes of peptides to the chaperone DnaK," *Journal of Molecular Biology*, vol. 425, no. 14, pp. 2463–2479, 2013.

[177] E. Kirschke, D. Goswami, D. Southworth, P. R. Griffin, and D. A. Agard, "Glucocorticoid receptor function regulated by coordinated action of the Hsp90 and Hsp70 chaperone cycles," *Cell*, vol. 157, no. 7, pp. 1685–1697, 2014.

[178] L. Kundrat and L. Regan, "Balance between folding and degradation for Hsp90-dependent client proteins: a key role for CHIP," *Biochemistry*, vol. 49, no. 35, pp. 7428–7438, 2010.

[179] J. Amick, S. E. Schlanger, C. Wachnowsky, M. A. Moseng, C. C. Emerson, M. Dare, W.-I. Luo, S. S. Ithychanda, J. C. Nix, J. A. Cowan, *et al.*, "Crystal structure of the nucleotide-binding domain of mortalin, the mitochondrial Hsp70 chaperone," *Protein Science*, vol. 23, no. 6, pp. 833–842, 2014.

[180] K. Kamieniarz and R. Schneider, "Tools to tackle protein acetylation," *Chemistry & Biology*, vol. 16, no. 10, pp. 1027–1029, 2009.

[181] D. R. Southworth and D. A. Agard, "Client-loading conformation of the Hsp90 molecular chaperone revealed in the cryo-EM structure of the human Hsp90: Hop complex," *Molecular Cell*, vol. 42, no. 6, pp. 771–781, 2011.

[182] A. Shevchenko, M. Wilm, O. Vorm, and M. Mann, "Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels," *Analytical Chemistry*, vol. 68, no. 5, pp. 850–858, 1996.

[183] J. V. Olsen, L. M. de Godoy, G. Li, B. Macek, P. Mortensen, R. Pesch, A. Makarov, O. Lange, S. Horning, and M. Mann, "Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap," *Molecular & Cellular Proteomics*, vol. 4, no. 12, pp. 2010–2021, 2005.

[184] R. Kityk, J. Kopp, I. Sinning, and M. P. Mayer, "Structure and dynamics of the ATP-bound open conformation of Hsp70 chaperones," *Molecular Cell*, vol. 48, no. 6, pp. 863–874, 2012.

[185] V. Beilsten-Edmands, Y. Gordiyenko, J. C. Kung, S. Mohammed, C. Schmidt, and C. V. Robinson, "eIF2 interactions with initiator tRNA and eIF2B are regulated by post-translational modifications and conformational dynamics," *Cell Discovery*, vol. 1, p. 15020, 2015.

[186] C. Schmidt, V. Beilsten-Edmands, and C. V. Robinson, "Insights into eukaryotic translation initiation from mass spectrometry of macromolecular protein assemblies," *Journal of Molecular Biology*, vol. 428, no. 2, pp. 344–356, 2016.

[187] F. H. C. Crick, "The biological replication of macromolecules," *Symposium of the Society for Experimental Biology*, vol. XII, no. 138, 1958.

[188] F. Crick, "Central dogma of molecular biology," *Nature*, vol. 227, no. 5258, pp. 561–563, 1970.

[189] N. Sonenberg and A. G. Hinnebusch, "Regulation of translation initiation in eukaryotes: mechanisms and biological targets," *Cell*, vol. 136, no. 4, pp. 731–745, 2009.

[190] C. E. Aitken and J. R. Lorsch, "A mechanistic overview of translation initiation in eukaryotes," *Nature Structural & Molecular Biology*, vol. 19, no. 6, pp. 568–576, 2012.

[191] M. V. Rodnina and W. Wintermeyer, "Recent mechanistic insights into eukaryotic ribosomes," *Current Opinion in Cell Biology*, vol. 21, no. 3, pp. 435–443, 2009.

[192] F. Mignone, C. Gissi, S. Liuni, and G. Pesole, "Untranslated regions of mRNAs," *Genome Biology*, vol. 3, no. 3, pp. reviews0004–1, 2002.

[193] C. R. Singh, B. Lee, T. Udagawa, S. S. Mohammad-Qureshi, Y. Yamamoto, G. D. Pavitt, and K. Asano, "An eIF5/eIF2 complex antagonizes guanine nucleotide exchange by eIF2B during translation initiation," *The EMBO Journal*, vol. 25, no. 19, pp. 4537–4546, 2006.

[194] M. D. Jennings and G. D. Pavitt, "eIF5 has GDI activity necessary for translational control by eIF2 phosphorylation," *Nature*, vol. 465, no. 7296, pp. 378–381, 2010.

[195] M. D. Jennings, Y. Zhou, S. S. Mohammad-Qureshi, D. Bennett, and G. D. Pavitt, "eIF2B promotes eIF5 dissociation from eIF2 GDP to facilitate guanine nucleotide exchange for translation initiation," *Genes & Development*, vol. 27, no. 24, pp. 2696–2707, 2013.

[196] C. R. Singh, T. Udagawa, B. Lee, S. Wassink, H. He, Y. Yamamoto, J. T. Anderson, G. D. Pavitt, and K. Asano, "Change in nutritional status modulates the abundance of critical pre-initiation intermediate complexes during translation initiation *in vivo*," *Journal of Molecular Biology*, vol. 370, no. 2, pp. 315–330, 2007.

[197] R. J. Jackson, C. U. Hellen, and T. V. Pestova, "The mechanism of eukaryotic translation initiation and principles of its regulation," *Nature Reviews Molecular Cell Biology*, vol. 11, no. 2, pp. 113–127, 2010.

[198] J. Nika, S. Rippel, and E. M. Hannig, "Biochemical analysis of the eIF2$\beta\gamma$ complex reveals a structural function for eIF2α in catalyzed nucleotide exchange," *Journal of Biological Chemistry*, vol. 276, no. 2, pp. 1051–1056, 2001.

[199] M. Naveau, C. Lazennec-Schurdevin, M. Panvert, E. Dubiez, Y. Mechulam, and E. Schmitt, "Roles of yeast eIF2α and eIF2β subunits in the binding of the initiator methionyl-tRNA," *Nucleic Acids Research*, vol. 41, no. 2, pp. 1047–1057, 2012.

[200] T. Krishnamoorthy, G. D. Pavitt, F. Zhang, T. E. Dever, and A. G. Hinnebusch, "Tight binding of the phosphorylated α subunit of initiation factor 2 (eIF2α) to the regulatory subunits of guanine nucleotide exchange factor eIF2B is required for inhibition of translation initiation," *Molecular and Cellular Biology*, vol. 21, no. 15, pp. 5018–5030, 2001.

[201] K. Asano, J. Clayton, A. Shalev, and A. G. Hinnebusch, "A multifactor complex of eukaryotic initiation factors, eIF1, eIF2, eIF3, eIF5, and initiator tRNAMet is an important translation initiation intermediate *in vivo*," *Genes & Development*, vol. 14, no. 19, pp. 2534–2546, 2000.

[202] C. R. Singh, H. He, M. Ii, Y. Yamamoto, and K. Asano, "Efficient incorporation of eukaryotic initiation factor 1 into the multifactor complex is critical for formation of functional ribosomal preinitiation complexes *in vivo*," *Journal of Biological Chemistry*, vol. 279, no. 30, pp. 31910–31920, 2004.

[203] D. S. Olsen, E. M. Savner, A. Mathew, F. Zhang, T. Krishnamoorthy, L. Phan, and A. G. Hinnebusch, "Domains of eIF1A that mediate binding to eIF2, eIF3 and eIF5B and promote ternary complex recruitment *in vivo*," *The EMBO Journal*, vol. 22, no. 2, pp. 193–204, 2003.

[204] E. Stolboushkina, S. Nikonov, A. Nikulin, U. Bläsi, D. J. Manstein, R. Fedorov, M. Garber, and O. Nikonov, "Crystal structure of the intact archaeal translation initiation factor 2 demonstrates very high conformational flexibility in the α- and β-subunits," *Journal of Molecular Biology*, vol. 382, no. 3, pp. 680–691, 2008.

[205] K. Asano, T. Krishnamoorthy, L. Phan, G. D. Pavitt, and A. G. Hinnebusch, "Conserved bipartite motifs in yeast eIF5 and eIF2Bϵ, GTPase-activating and GDP–GTP exchange factors in translation initiation, mediate binding to their common substrate eIF2," *The EMBO Journal*, vol. 18, no. 6, pp. 1673–1688, 1999.

[206] S. S. Mohammad-Qureshi, R. Haddad, E. J. Hemingway, J. P. Richardson, and G. D. Pavitt, "Critical contacts between the eukaryotic initiation factor 2B (eIF2B) catalytic domain and both eIF2β and 2γ mediate guanine nucleotide exchange," *Molecular and Cellular Biology*, vol. 27, no. 14, pp. 5225–5234, 2007.

[207] P. V. Alone and T. E. Dever, "Direct binding of translation initiation factor eIF2γ-G domain to its GTPase-activating and GDP-GTP exchange factors eIF5 and eIF2Bϵ," *Journal of Biological Chemistry*, vol. 281, no. 18, pp. 12636–12644, 2006.

[208] G. D. Pavitt, K. V. Ramaiah, S. R. Kimball, and A. G. Hinnebusch, "eIF2 independently binds two distinct eIF2B subcomplexes that catalyze and regulate guanine–nucleotide exchange," *Genes & Development*, vol. 12, no. 4, pp. 514–526, 1998.

[209] N. C. Wortham, M. Martinez, Y. Gordiyenko, C. V. Robinson, and C. G. Proud, "Analysis of the subunit organization of the eIF2B complex reveals new insights into its structure and regulation," *The FASEB Journal*, vol. 28, no. 5, pp. 2225–2237, 2014.

[210] Y. Gordiyenko, C. Schmidt, M. D. Jennings, D. Matak-Vinkovic, G. D. Pavitt, and C. V. Robinson, "eIF2B is a decameric guanine nucleotide exchange factor with a $\gamma 2\varepsilon 2$ tetrameric core," *Nature Communications*, vol. 5, 2014.

[211] G. Pavitt, "eif2b, a mediator of general and gene-specific translational control," 2005.

[212] A. M. Cigan, J. L. Bushman, T. R. Boal, and A. G. Hinnebusch, "A protein complex of translational regulators of GCN4 mRNA is the guanine nucleotide-exchange factor for translation initiation factor 2 in yeast," *Proceedings of the National Academy of Sciences*, vol. 90, no. 11, pp. 5350–5354, 1993.

[213] T. B. Hiyama, T. Ito, H. Imataka, and S. Yokoyama, "Crystal structure of the α subunit of human translation initiation factor 2B," *Journal of Molecular Biology*, vol. 392, no. 4, pp. 937–951, 2009.

[214] N. C. Wortham and C. G. Proud, "eIF2B: recent structural and functional insights into a key regulator of translation," *Biochemical Society Transactions*, vol. 43, no. 6, pp. 1234–1240, 2015.

[215] T. E. Dever, L. Feng, R. C. Wek, A. M. Cigan, T. F. Donahue, and A. G. Hinnebusch, "Phosphorylation of initiation factor 2α by protein kinase GCN2 mediates gene-specific translational control of GCN4 in yeast," *Cell*, vol. 68, no. 3, pp. 585–596, 1992.

[216] C. O. Brostrom, C. R. Prostko, R. J. Kaufman, and M. A. Brostrom, "Inhibition of translational initiation by activators of the glucose-regulated stress protein and heat shock protein stress response systems," *Journal of Biological Chemistry*, vol. 271, no. 40, pp. 24995–25002, 1996.

[217] D. Ron, "Translational control in the endoplasmic reticulum stress response," *The Journal of Clinical Investigation*, vol. 110, no. 10, p. 1383, 2002.

[218] H. P. Harding, Y. Zhang, H. Zeng, I. Novoa, P. D. Lu, M. Calfon, N. Sadri, C. Yun, B. Popko, R. Paules, D. F. Stojdl, J. C. Bell, T. Hettmann, J. M. Leiden, and D. Ron, "An integrated stress response regulates amino acid metabolism and resistance to oxidative stress," *Molecular Cell*, vol. 11, no. 3, pp. 619–633, 2003.

[219] R. Wek, H.-Y. Jiang, and T. Anthony, "Coping with stress: eIF2 kinases and translational control," *Biochemical Society Transactions*, vol. 34, pp. 7–11, 2006.

[220] T. Rzymski, M. Milani, L. Pike, F. Buffa, H. Mellor, L. Winchester, I. Pires, E. Hammond, I. Ragoussis, and A. Harris, "Regulation of autophagy by ATF4 in response to severe hypoxia," *Oncogene*, vol. 29, no. 31, pp. 4424–4435, 2010.

[221] J. Ye, M. Kumanova, L. S. Hart, K. Sloane, H. Zhang, D. N. De Panis, E. Bobrovnikova-Marjon, J. A. Diehl, D. Ron, and C. Koumenis, "The GCN2-ATF4 pathway is critical for tumour cell survival and proliferation in response to nutrient deprivation," *The EMBO Journal*, vol. 29, no. 12, pp. 2082–2096, 2010.

[222] M. Garcia, E. Meurs, and M. Esteban, "The dsRNA protein kinase PKR: virus and cell control," *Biochimie*, vol. 89, no. 6, pp. 799–811, 2007.

[223] D. Scheuner, B. Song, E. McEwen, C. Liu, R. Laybutt, P. Gillespie, T. Saunders, S. Bonner-Weir, and R. J. Kaufman, "Translational control is required for the unfolded protein response and *in vivo* glucose homeostasis," *Molecular Cell*, vol. 7, no. 6, pp. 1165–1176, 2001.

[224] H. P. Harding, I. Novoa, Y. Zhang, H. Zeng, R. Wek, M. Schapira, and D. Ron, "Regulated translation initiation controls stress-induced gene expression in mammalian cells," *Molecular Cell*, vol. 6, no. 5, pp. 1099–1108, 2000.

[225] I. Novoa, Y. Zhang, H. Zeng, R. Jungreis, H. P. Harding, and D. Ron, "Stress-induced gene expression requires programmed recovery from translational repression," *The EMBO Journal*, vol. 22, no. 5, pp. 1180–1187, 2003.

[226] A. W. Engberg and T. W. Teasdale, "Psychosocial outcome following traumatic brain injury in adults: A long-term population-based follow-up," *Brain Injury*, vol. 18, no. 6, pp. 533–545, 2004.

[227] S. T. DeKosky, M. D. Ikonomovic, and S. Gandy, "Traumatic brain injury–football, warfare, and long-term effects," *New England Journal of Medicine*, vol. 363, no. 14, pp. 1293–1296, 2010.

[228] T. Petrov, B. D. Underwood, B. Braun, S. S. Alousi, and J. A. Rafols, "Upregulation of iNOS expression and phosphorylation of eIF-2 α are paralleled by suppression of protein synthesis in rat hypothalamus in a closed head trauma model," *Journal of Neurotrauma*, vol. 18, no. 8, pp. 799–812, 2001.

[229] D. H. Smith, V. E. Johnson, and W. Stewart, "Chronic neuropathologies of single and repetitive TBI: substrates of dementia?," *Nature Reviews Neurology*, vol. 9, no. 4, pp. 211–221, 2013.

[230] P. Sullivan, D. Petitti, and J. Barbaccia, "Head trauma and age of onset of dementia of the Alzheimer type," *The Journal of the American Medical Association*, vol. 257, no. 17, pp. 2289–2290, 1987.

[231] C. Sidrauski, D. Acosta-Alvear, A. Khoutorsky, P. Vedantham, B. R. Hearn, H. Li, K. Gamache, C. M. Gallagher, K. K. Ang, C. Wilson, V. Okreglak, A. Ashkenazi, B. Hann, K. Nader, M. R. Arkin, A. R. Renslo, N. Sonenberg, and P. Walter, "Pharmacological brake-release of mRNA translation enhances cognitive memory," *Elife*, vol. 2, p. e00498, 2013.

[232] C. Sidrauski, A. M. McGeachy, N. T. Ingolia, and P. Walter, "The small molecule ISRIB reverses the effects of eIF2α phosphorylation on translation and stress granule assembly," *Elife*, vol. 4, p. e05033, 2015.

[233] C. Sidrauski, J. C. Tsai, M. Kampmann, B. R. Hearn, P. Vedantham, P. Jaishankar, M. Sokabe, A. S. Mendez, B. W. Newton, E. L. Tang, E. Verschueren, J. R. Johnson, N. J. Krogan, C. S. Fraser, J. S. Weissman, A. R. Renslo, and P. Walter, "Pharmacological dimerization and activation of the exchange factor eIF2B antagonizes the integrated stress response," *Elife*, vol. 4, p. e07314, 2015.

[234] Y. Sekine, A. Zyryanova, A. Crespillo-Casado, P. M. Fischer, H. P. Harding, and D. Ron, "Mutations in a translation initiation factor identify the target of a memory-enhancing compound," *Science*, vol. 348, no. 6238, pp. 1027–1030, 2015.

[235] T. Ly, Y. Ahmad, A. Shlien, D. Soroka, A. Mills, M. J. Emanuele, M. R. Stratton, and A. I. Lamond, "A proteomic chronology of gene expression through the cell cycle in human myeloid leukemia cells," *Elife*, vol. 3, p. e01630, 2014.

[236] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, "The protein data bank, 1999–," in *International Tables for Crystallography Volume F Crystallography of Biological Macromolecules*, pp. 675–684, Springer, 2006.

[237] L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, and M. J. Sternberg, "The Phyre2 web portal for protein modeling, prediction and analysis," *Nature Protocols*, vol. 10, no. 6, pp. 845–858, 2015.

[238] Y. Ban, Y. Kobayashi, T. Hara, T. Hamada, T. Hashimoto, S. Takeda, and T. Hattori, "α-tubulin is rapidly phosphorylated in response to hyperosmotic stress in rice and Arabidopsis," *Plant and Cell Physiology*, vol. 54, no. 6, pp. 848–858, 2013.

[239] J. Francis, W. Lin, O. Rozenblatt-Rosen, and M. Meyerson, "The menin tumor suppressor protein is phosphorylated in response to DNA damage," *PLoS One*, vol. 6, no. 1, p. e16119, 2011.

[240] E. Rollet and M. Best-Belpomme, "HSP26 and 27 are phosphorylated in response to heat shock and ecdysterone in *Drosophila melanogaster* cells," *Biochemical and Biophysical Research Communications*, vol. 141, no. 2, pp. 426–433, 1986.

[241] H. Cao, A. R. Sanguinetti, and C. C. Mastick, "Oxidative stress activates both Src-kinases and their negative regulator Csk and induces phosphorylation of two targeting proteins for Csk: caveolin-1 and paxillin," *Experimental Cell Research*, vol. 294, no. 1, pp. 159–171, 2004.

[242] C. P. Albuquerque, M. B. Smolka, S. H. Payne, V. Bafna, J. Eng, and H. Zhou, "A multidimensional chromatography technology for in-depth phosphoproteome analysis," *Molecular & Cellular Proteomics*, vol. 7, no. 7, pp. 1389–1396, 2008.

[243] X. Li, S. A. Gerber, A. D. Rudner, S. A. Beausoleil, W. Haas, J. Villen, J. E. Elias, and S. P. Gygi, "Large-scale phosphorylation analysis of α-factor-arrested *Saccharomyces cerevisiae*," *Journal of Proteome Research*, vol. 6, no. 3, pp. 1190–1197, 2007.

[244] B. Kuhle, N. K. Eulig, and R. Ficner, "Architecture of the eIF2B regulatory subcomplex and its implications for the regulation of guanine nucleotide exchange on eIF2," *Nucleic Acids Research*, vol. 43, no. 20, pp. 9994–10014, 2015.

[245] A. M. Bogorad, B. Xia, D. G. Sandor, A. B. Mamonov, T. R. Cafarella, S. Jehle, S. Vajda, D. Kozakov, and A. Marintchev, "Insights into the architecture of the eIF2B$\alpha/\beta/\delta$ regulatory subcomplex," *Biochemistry*, vol. 53, no. 21, pp. 3432–3445, 2014.

[246] Y. Zhou, *Structural Study of eIF2B by Electron Microscopy*. PhD thesis, The University of Manchester, 2016.

[247] M. Grimm, T. Zimniak, A. Kahraman, and F. Herzog, "xVis: a web server for the schematic visualization and interpretation of crosslink-derived spatial restraints," *Nucleic Acids Research*, vol. 43, no. W1, pp. W362–W369, 2015.

[248] K. Kashiwagi, M. Takahashi, M. Nishimoto, T. B. Hiyama, T. Higo, T. Umehara, K. Sakamoto, T. Ito, and S. Yokoyama, "Crystal structure of eukaryotic translation initiation factor 2B," *Nature*, vol. 531, no. 7592, pp. 122–125, 2016.

[249] R. Elsby, J. F. Heiber, P. Reid, S. R. Kimball, G. D. Pavitt, and G. N. Barber, "The α subunit of eukaryotic initiation factor 2B (eIF2B) is required for eIF2-mediated translational suppression of vesicular stomatitis virus," *Journal of Virology*, vol. 85, no. 19, pp. 9716–9725, 2011.

[250] K. F. Geoghegan, H. B. Dixon, P. J. Rosner, L. R. Hoth, A. J. Lanzetti, K. A. Borzilleri, E. S. Marr, L. H. Pezzullo, L. B. Martin, P. K. LeMotte, A. S. McColl, A. V. Kamath, and J. G. Stroha, "Spontaneous α-N-6-phosphogluconoylation of a 'his tag' in *Escherichia coli*: the cause of extra mass of 258 or 178 Da in fusion proteins," *Analytical Biochemistry*, vol. 267, no. 1, pp. 169–184, 1999.

[251] L. Yatime, E. Schmitt, S. Blanquet, and Y. Mechulam, "Functional molecular mapping of archaeal translation initiation factor 2," *Journal of Biological Chemistry*, vol. 279, no. 16, pp. 15984–15993, 2004.

[252] X. Wang and C. G. Proud, "A novel mechanism for the control of translation initiation by amino acids, mediated by phosphorylation of eukaryotic initiation factor 2B," *Molecular and Cellular Biology*, vol. 28, no. 5, pp. 1429–1442, 2008.

[253] K. Kashiwagi, T. Ito, and S. Yokoyama, "Crystal structure of eIF2B and insights into eIF2–eIF2B interactions," *The FEBS Journal*, vol. 284, no. 6, pp. 868–874, 2017.

[254] A. Chou, K. Krukowski, T. Jopson, P. J. Zhu, M. Costa-Mattioli, P. Walter, and S. Rosi, "Inhibition of the integrated stress response reverses cognitive deficits after traumatic brain injury," *Proceedings of the National Academy of Sciences*, p. 201707661, 2017.

[255] J. C. Tsai, L. E. Miller-Vedam, A. A. Anand, P. Jaishankar, H. C. Nguyen, A. R. Renslo, A. Frost, and P. Walter, "Structure of the nucleotide exchange factor eIF2B reveals mechanism of memory-enhancing molecule," *Science*, vol. 359, no. 6383, p. eaaq0939, 2018.

[256] A. F. Zyryanova, F. Weis, A. Faille, A. A. Alard, A. Crespillo-Casado, Y. Sekine, H. P. Harding, F. Allen, L. Parts, C. Fromont, P. M. Fischer, A. J. Warren, and D. Ron, "Binding of ISRIB reveals a regulatory site in the nucleotide exchange factor eIF2B," *Science*, vol. 359, no. 6383, pp. 1533–1536, 2018.

[257] E. Gomez, S. S. Mohammad, and G. D. Pavitt, "Characterization of the minimal catalytic domain within eIF2B: the guanine-nucleotide exchange factor for translation initiation," *The EMBO Journal*, vol. 21, no. 19, pp. 5292–5301, 2002.

[258] S. S. Mohammad-Qureshi, R. Haddad, K. S. Palmer, J. P. Richardson, E. Gomez, and G. D. Pavitt, "Purification of FLAG-tagged eukaryotic initiation factor 2B complexes, subcomplexes, and fragments from *Saccharomyces cerevisiae*," *Methods in Enzymology*, vol. 431, pp. 1–13, 2007.

[259] L.-h. Wang, D.-Q. Li, Y. Fu, H.-P. Wang, J.-F. Zhang, Z.-F. Yuan, R.-X. Sun, R. Zeng, S.-M. He, and W. Gao, "pFind 2.0: a software package for peptide and protein identification via tandem mass spectrometry," *Rapid Communications in Mass Spectrometry*, vol. 21, no. 18, pp. 2985–2991, 2007.

[260] R. Apweiler, A. Bairoch, C. H. Wu, W. C. Barker, B. Boeckmann, S. Ferro, E. Gasteiger, H. Huang, R. Lopez, M. J. Magrane, Michele Martin, D. A. Natale, C. O'Donovan, N. Redaschi, and L.-S. L. Yeh, "UniProt: the universal protein knowledgebase," *Nucleic Acids Research*, vol. 32, no. suppl_1, pp. D115–D119, 2004.

[261] F. Gnad, S. Ren, J. Cox, J. V. Olsen, B. Macek, M. Oroshi, and M. Mann, "PHOSIDA (phosphorylation site database): management, structural and evolutionary investigation, and prediction of phosphosites," *Genome Biology*, vol. 8, no. 11, p. R250, 2007.

[262] T. Ito, A. Marintchev, and G. Wagner, "Solution structure of human initiation factor eIF2α reveals homology to the elongation factor eEF1B," *Structure*, vol. 12, no. 9, pp. 1693–1704, 2004.

[263] M. Sokabe, M. Yao, N. Sakai, S. Toya, and I. Tanaka, "Structure of archaeal translational initiation factor 2 $\beta\gamma$–GDP reveals significant conformational change of the β-subunit and switch 1 region," *Proceedings of the National Academy of Sciences*, vol. 103, no. 35, pp. 13016–13021, 2006.

[264] A. Simonetti, J. B. Querido, A. G. Myasnikov, E. Mancera-Martinez, A. Renaud, L. Kuhn, and Y. Hashem, "eIF3 peripheral subunits rearrangement after mRNA binding and start-codon recognition," *Molecular Cell*, vol. 63, no. 2, pp. 206–217, 2016.

[265] T. Boesen, S. S. Mohammad, G. D. Pavitt, and G. R. Andersen, "Structure of the catalytic fragment of translation initiation factor 2B and identification of a critically important catalytic residue," *Journal of Biological Chemistry*, vol. 279, no. 11, pp. 10584–10592, 2004.

[266] C. Schmidt, V. Beilsten-Edmands, S. Mohammed, and C. V. Robinson, "Acetylation and phosphorylation control both local and global stability of the chloroplast F_1 ATP synthase," *Scientific Reports*, vol. 7, 2017.

[267] H. Lodish, A. Berk, S. L. Zipursky, P. Matsudaira, D. Baltimore, and J. Darnell, *Molecular cell biology*, vol. 4. WH Freeman New York, 2000.

[268] J. E. Walker, "The ATP synthase: the understood, the uncertain and the unknown," 2013.

[269] S. P. Muench, J. Trinick, and M. A. Harrison, "Structural divergence of the rotary ATPases," *Quarterly Reviews of Biophysics*, vol. 44, no. 3, pp. 311–356, 2011.

[270] P. D. Boyer, "The ATP synthase–a splendid molecular machine," *Annual Review of Biochemistry*, vol. 66, no. 1, pp. 717–749, 1997.

[271] S. B. Vik and B. J. Antonio, "A mechanism of proton translocation by F_1F_o ATP synthases suggested by double mutants of the a subunit," *Journal of Biological Chemistry*, vol. 269, no. 48, pp. 30364–30369, 1994.

[272] W. Junge, H. Lill, and S. Engelbrecht, "ATP synthase: an electrochemical ransducer with rotatory mechanics," *Trends in Biochemical Sciences*, vol. 22, no. 11, pp. 420–423, 1997.

[273] H. Guo, S. A. Bueler, and J. L. Rubinstein, "Atomic model for the dimeric F_o region of mitochondrial ATP synthase," *Science*, vol. 358, no. 6365, pp. 936–940, 2017.

[274] P. D. Boyer, "The binding change mechanism for ATP synthasesome probabilities and possibilities," *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, vol. 1140, no. 3, pp. 215–250, 1993.

[275] K. Shimabukuro, R. Yasuda, E. Muneyuki, K. Y. Hara, K. Kinosita, and M. Yoshida, "Catalysis and rotation of F_1 motor: cleavage of ATP at the catalytic site occurs in 1 ms before 40 substep rotation," *Proceedings of the National Academy of Sciences*, vol. 100, no. 25, pp. 14731–14736, 2003.

[276] A. Malyan, "Noncatalytic nucleotide binding sites: Properties and mechanism of involvement in ATP synthase activity regulation," *Biochemistry*, vol. 78, no. 13, p. 1512, 2013.

[277] M. Kanekatsu, H. Saito, K. Motohashi, and T. Hisabori, "The β subunit of chloroplast ATP synthase (CF_oCF_1-ATPase) is phosphorylated by casein kinase II," *IUBMB Life*, vol. 46, no. 1, pp. 99–105, 1998.

[278] A. Zhou, A. Rohou, D. G. Schep, J. V. Bason, M. G. Montgomery, J. E. Walker, N. Grigorieff, and J. L. Rubinstein, "Structure and conformational states of the bovine mitochondrial ATP synthase by cryo-EM," *Elife*, vol. 4, p. e10180, 2015.

[279] J. P. Abrahams, A. G. Leslie, R. Lutter, and J. E. Walker, "Structure at 2.8 Å resolution of F_1-ATPase from bovine heart mitochondria," *Nature*, vol. 370, no. 6491, pp. 621–628, 1994.

[280] H. Päivärinne and H. Kainulainen, "DAPIT, a novel protein down-regulated in insulin-sensitive tissues in streptozotocin-induced diabetes," *Acta Diabetologica*, vol. 38, no. 2, pp. 83–86, 2001.

[281] E. Terzi, P. Boyot, A. Van Dorsselaer, B. Luu, and E. Trifilieff, "Isolation and amino acid sequence of a novel 6.8-kDa mitochondrial proteolipid from beef heart: Use of FAB-MS for molecular mass determination," *FEBS Letters*, vol. 260, no. 1, pp. 122–126, 1990.

[282] M. J. Runswick, J. V. Bason, M. G. Montgomery, G. C. Robinson, I. M. Fearnley, and J. E. Walker, "The affinity purification and characterization of ATP synthase complexes from mitochondria," *Open Biology*, vol. 3, no. 2, p. 120160, 2013.

[283] M.-F. Giraud, P. Paumard, V. Soubannier, J. Vaillier, G. Arselin, B. Salin, J. Schaeffer, D. Brèthes, J.-P. di Rago, and J. Velours, "Is there a relationship between the supramolecular organization of the mitochondrial ATP synthase and the formation of cristae?," *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, vol. 1555, no. 1, pp. 174–180, 2002.

[284] J. Habersetzer, I. Larrieu, M. Priault, B. Salin, R. Rossignol, D. Brèthes, and P. Paumard, "Human F_1F_o ATP synthase, mitochondrial ultrastructure and OXPHOS impairment: a (super-) complex matter?," *PLoS One*, vol. 8, no. 10, p. e75429, 2013.

[285] P. Paumard, J. Vaillier, B. Coulary, J. Schaeffer, V. Soubannier, D. M. Mueller, D. Brèthes, J.-P. di Rago, and J. Velours, "The ATP synthase is involved in

generating mitochondrial cristae morphology," *The EMBO Journal*, vol. 21, no. 3, pp. 221–230, 2002.

[286] N. Fischer, V. Beilsten-Edmands, D. Chorev, F. Hauer, C. Jiko, S. Shimada, K. Shinzawa-Itoh, C. Robinson, H. Stark, and C. Gerle, "Structure and function of the intermembrane space domain of mammalian F_oF_1 ATP synthase," *BioRxiv*, p. 193045, 2017.

[287] T. H. Stevens and M. Forgac, "Structure, function and regulation of the vacuolar (H^+)-ATPase," *Annual Review of Cell and Developmental Biology*, vol. 13, no. 1, pp. 779–808, 1997.

[288] A. N. Smith, J. Skaug, K. A. Choate, A. Nayir, A. Bakkaloglu, S. Ozen, S. A. Hulton, S. A. Sanjad, E. A. Al-Sabban, R. P. Lifton, S. W. Scherer, and F. E. Karet, "Mutations in ATP6N1B, encoding a new kidney vacuolar proton pump 116-kD subunit, cause recessive distal renal tubular acidosis with preserved hearing," *Nature Genetics*, vol. 26, no. 1, pp. 71–76, 2000.

[289] A. Frattini, P. J. Orchard, C. Sobacchi, S. Giliani, M. Abinun, J. P. Mattsson, D. J. Keeling, A.-K. Andersson, P. Wallbrandt, L. Zecca, L. D. Notarangelo, P. Vezzoni, and A. Villa, "Defects in TCIRG1 subunit of the vacuolar proton pump are responsible for a subset of human autosomal recessive osteopetrosis," *Nature Genetics*, vol. 25, no. 3, pp. 343–346, 2000.

[290] M. Forgac, "Vacuolar ATPases: rotary proton pumps in physiology and pathophysiology," *Nature Reviews Molecular Cell Biology*, vol. 8, no. 11, pp. 917–929, 2007.

[291] S. Benlekbir, S. A. Bueler, and J. L. Rubinstein, "Structure of the vacuolar-type ATPase from *Saccharomyces cerevisiae* at 11-Å resolution," *Nature Structural & Molecular Biology*, vol. 19, no. 12, pp. 1356–1362, 2012.

[292] M. F. Manolson, B. Wu, D. Proteau, B. E. Taillon, B. T. Roberts, M. A. Hoyt, and E. W. Jones, "STV1 gene encodes functional homologue of 95-kDa yeast vacuolar H^+-ATPase subunit Vph1p," *Journal of Biological Chemistry*, vol. 269, no. 19, pp. 14064–14074, 1994.

[293] S. Kawasaki-Nishi, K. Bowers, T. Nishi, M. Forgac, and T. H. Stevens, "The amino-terminal domain of the vacuolar proton-translocating ATPase a subunit controls targeting and *in vivo* dissociation, and the carboxyl-terminal domain affects coupling of proton transport and ATP hydrolysis," *Journal of Biological Chemistry*, vol. 276, no. 50, pp. 47411–47420, 2001.

[294] S. Kawasaki-Nishi, T. Nishi, and M. Forgac, "Yeast V-ATPase complexes containing different isoforms of the 100-kDa a-subunit differ in coupling efficiency and *in vivo* dissociation," *Journal of Biological Chemistry*, vol. 276, no. 21, pp. 17941–17948, 2001.

[295] J. Zhao, S. Benlekbir, and J. L. Rubinstein, "Electron cryomicroscopy observation of rotational states in a eukaryotic V-ATPase," *Nature*, vol. 521, no. 7551, p. 241, 2015.

[296] R. I. Feldman and P. D. Boyer, "The role of tightly bound ADP on chloroplast ATPase," *Journal of Biological Chemistry*, vol. 260, no. 24, pp. 13088–13094, 1985.

[297] P. M. Kane, "Disassembly and reassembly of the yeast vacuolar H^+-ATPase *in vivo*," *Journal of Biological Chemistry*, vol. 270, no. 28, pp. 17025–17032, 1995.

[298] J. Schumann, M. Richter, and R. McCarty, "Partial proteolysis as a probe of the conformation of the gamma subunit in activated soluble and membrane-bound chloroplast coupling factor 1," *Journal of Biological Chemistry*, vol. 260, no. 21, pp. 11817–11823, 1985.

[299] K. J. Parra, K. L. Keenan, and P. M. Kane, "The H subunit (Vma13p) of the yeast V-ATPase inhibits the ATPase activity of cytosolic V1 complexes," *Journal of Biological Chemistry*, vol. 275, no. 28, pp. 21761–21767, 2000.

[300] J. Qi and M. Forgac, "Cellular environment is important in controlling V-ATPase dissociation and its dependence on activity," *Journal of Biological Chemistry*, vol. 282, no. 34, pp. 24743–24751, 2007.

[301] H. Yagi, H. Konno, T. Murakami-Fuse, A. Isu, T. Oroguchi, H. Akutsu, M. Ikeguchi, and T. Hisabori, "Structural and functional analysis of the intrinsic inhibitor subunit ϵ of F_1-ATPase from photosynthetic organisms," *Biochemical Journal*, vol. 425, no. 1, pp. 85–98, 2010.

[302] P. P. Laget and J. B. Smith, "Inhibitory properties of endogenous subunit ϵ in the *Escherichia coli* F_1 ATPase," *Archives of Biochemistry and Biophysics*, vol. 197, no. 1, pp. 83–89, 1979.

[303] Y. Kato-Yamada, D. Bald, M. Koike, K. Motohashi, T. Hisabori, and M. Yoshida, "ε subunit, an endogenous inhibitor of bacterial F_1-ATPase, also inhibits F_oF_1-ATPase," *Journal of Biological Chemistry*, vol. 274, no. 48, pp. 33991–33994, 1999.

[304] M. L. Richter, W. J. Patrie, and R. E. McCarty, "Preparation of the epsilon subunit and epsilon subunit-deficient chloroplast coupling factor 1 in reconstitutively active forms," *Journal of Biological Chemistry*, vol. 259, no. 12, pp. 7371–7373, 1984.

[305] X.-B. Shi, J.-M. Wei, and Y.-K. Shen, "Effects of sequential deletions of residues from the N-or C-terminus on the functions of ε subunit of the chloroplast ATP synthase," *Biochemistry*, vol. 40, no. 36, pp. 10825–10831, 2001.

[306] J. J. Garcıa and R. A. Capaldi, "Unisite catalysis without rotation of the γ-ε domain in *Escherichia coli* F_1-ATPase," *Journal of Biological Chemistry*, vol. 273, no. 26, pp. 15940–15945, 1998.

[307] E. Cabezon, M. Runswick, A. Leslie, and J. Walker, "The structure of bovine IF_1, the regulatory subunit of mitochondrial F-ATPase," *The EMBO Journal*, vol. 20, no. 24, pp. 6990–6996, 2001.

[308] B. A. Feniouk, T. Suzuki, and M. Yoshida, "The role of subunit epsilon in the catalysis and regulation of F_oF_1-ATP synthase," *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, vol. 1757, no. 5, pp. 326–338, 2006.

[309] T. Hashimoto, Y. Negawa, and K. Tagawa, "Binding of intrinsic ATPase inhibitor to mitochondrial ATPase-stoichiometry of binding of nucleotides, inhibitor, and enzyme," *The Journal of Biochemistry*, vol. 90, no. 4, pp. 1151–1157, 1981.

[310] B. Norling, C. Tourikas, B. Hamasur, and E. Glaser, "Evidence for an endogenous ATPase inhibitor protein in plant mitochondria," *The FEBS Journal*, vol. 188, no. 2, pp. 247–252, 1990.

[311] N. Cintron and P. L. Pedersen, "A protein inhibitor of the mitochondrial adenosine triphosphatase complex of rat liver. purification and characterization.," *Journal of Biological Chemistry*, vol. 254, no. 9, pp. 3439–3443, 1979.

[312] M. E. Pullman and G. C. Monroy, "A naturally occurring inhibitor of mitochondrial adenosine triphosphatase," *Journal of Biological Chemistry*, vol. 238, no. 11, pp. 3762–3769, 1963.

[313] E. Cabezon, P. J. G. Butler, M. J. Runswick, and J. E. Walker, "Modulation of the oligomerization state of the bovine F_1-ATPase inhibitor protein, IF_1, by pH," *Journal of Biological Chemistry*, vol. 275, no. 33, pp. 25460–25464, 2000.

[314] E. Cabezón, I. Arechaga, P. Jonathan, G. Butler, and J. E. Walker, "Dimerization of bovine F_1-ATPase by binding the inhibitor protein, IF_1," *Journal of Biological Chemistry*, vol. 275, no. 37, pp. 28353–28355, 2000.

[315] E. Cabezón, P. J. G. Butler, M. J. Runswick, R. J. Carbajo, and J. E. Walker, "Homologous and heterologous inhibitory effects of ATPase inhibitor proteins on F-ATPases," *Journal of Biological Chemistry*, vol. 277, no. 44, pp. 41334–41341, 2002.

[316] E. Cabezón, M. G. Montgomery, A. G. Leslie, and J. E. Walker, "The structure of bovine F_1-ATPase in complex with its regulatory protein IF_1," *Nature Structural & Molecular Biology*, vol. 10, no. 9, pp. 744–750, 2003.

[317] T. Kouzarides, "Acetylation: a regulatory modification to rival phosphorylation?," *The EMBO Journal*, vol. 19, no. 6, pp. 1176–1179, 2000.

[318] C. Choudhary, C. Kumar, F. Gnad, M. L. Nielsen, M. Rehman, T. C. Walther, J. V. Olsen, and M. Mann, "Lysine acetylation targets protein complexes and co-regulates major cellular functions," *Science*, vol. 325, no. 5942, pp. 834–840, 2009.

[319] P. Giansanti, L. Tsiatsiani, T. Y. Low, and A. J. Heck, "Six alternative proteases for mass spectrometry-based proteomics beyond trypsin," *Nature Protocols*, vol. 11, no. 5, pp. 993–1006, 2016.

[320] M. Grabsztunowicz, M. M. Koskela, and P. Mulo, "Post-translational modifications in regulation of chloroplast function: recent advances," *Frontiers in Plant Science*, vol. 8, 2017.

[321] M. Rahman, N. K. Nirala, A. Singh, L. J. Zhu, K. Taguchi, T. Bamba, E. Fukusaki, L. M. Shaw, D. G. Lambright, J. K. Acharya, and U. R. Acharya, "*Drosophila* Sirt2/mammalian SIRT3 deacetylates ATP synthase β and regulates complex V activity," *Journal of Cell Biology*, pp. jcb–201404118, 2014.

[322] A. Vassilopoulos, J. D. Pennington, T. Andresson, D. M. Rees, A. D. Bosley, I. M. Fearnley, A. Ham, C. R. Flynn, S. Hill, K. L. Rose, H.-S. Kim, C.-X. Deng, J. E. Walker, and G. David, "SIRT3 deacetylates ATP synthase F_1 complex proteins in response to nutrient-and exercise-induced stress," *Antioxidants & Redox Signaling*, vol. 21, no. 4, pp. 551–564, 2014.

[323] I. Finkemeier, M. Laxa, L. Miguet, A. J. Howden, and L. J. Sweetlove, "Proteins of diverse function and subcellular location are lysine acetylated in *Arabidopsis*," *Plant Physiology*, vol. 155, no. 4, pp. 1779–1790, 2011.

[324] Z. Hall, A. Politis, M. F. Bush, L. J. Smith, and C. V. Robinson, "Charge-state dependent compaction and dissociation of protein complexes: insights from ion mobility and molecular dynamics," *Journal of the American Chemical Society*, vol. 134, no. 7, pp. 3429–3438, 2012.

[325] K. Pagel, S.-J. Hyung, B. T. Ruotolo, and C. V. Robinson, "Alternate dissociation pathways identified in charge-reduced protein complex ions," *Analytical Chemistry*, vol. 82, no. 12, pp. 5363–5372, 2010.

[326] S. Nesci, F. Trombetti, V. Ventrella, and A. Pagliarani, "Post-translational modifications of the mitochondrial F_1F_o-ATPase," *Biochimica et Biophysica Acta (BBA)-General Subjects*, 2017.

[327] F. Frottin, A. Martinez, P. Peynot, S. Mitra, R. C. Holz, C. Giglione, and T. Meinnel, "The proteomics of N-terminal methionine cleavage," *Molecular & Cellular Proteomics*, vol. 5, no. 12, pp. 2336–2349, 2006.

[328] J. Carroll, I. M. Fearnley, Q. Wang, and J. E. Walker, "Measurement of the molecular masses of hydrophilic and hydrophobic subunits of ATP synthase and complex I in a single experiment," *Analytical Biochemistry*, vol. 395, no. 2, pp. 249–255, 2009.

[329] N. V. Dudkina, J. Heinemeyer, W. Keegstra, E. J. Boekema, and H.-P. Braun, "Structure of dimeric ATP synthase from mitochondria: an angular association of monomers induces the strong curvature of the inner membrane," *FEBS Letters*, vol. 579, no. 25, pp. 5769–5772, 2005.

[330] E. Fahy, M. Sud, D. Cotter, and S. Subramaniam, "LIPID MAPS online tools for lipid research," *Nucleic Acids Research*, vol. 35, pp. W606–W612, 2007.

[331] T. Taverner, H. Hernández, M. Sharon, B. T. Ruotolo, D. Matak-Vinkovic, D. Devos, R. B. Russell, and C. V. Robinson, "Subunit architecture of intact protein complexes from mass spectrometry and homology modeling," *Accounts of Chemical Research*, vol. 41, no. 5, pp. 617–627, 2008.

[332] L. A. Kane and J. E. Van Eyk, "Post-translational modifications of ATP synthase in the heart: biology and function," *Journal of Bioenergetics and Biomembranes*, vol. 41, no. 2, pp. 145–150, 2009.

[333] R. Chen, I. M. Fearnley, D. N. Palmer, and J. E. Walker, "Lysine 43 is trimethylated in subunit c from bovine mitochondrial ATP synthase and in storage bodies associated with Batten disease," *Journal of Biological Chemistry*, vol. 279, no. 21, pp. 21883–21887, 2004.

[334] T. B. Walpole, D. N. Palmer, H. Jiang, S. Ding, I. M. Fearnley, and J. E. Walker, "Conservation of complete trimethylation of lysine-43 in the rotor ring of c-subunits of metazoan adenosine triphosphate (ATP) synthases," *Molecular & Cellular Proteomics*, vol. 14, no. 4, pp. 828–840, 2015.

[335] M. J. Suskiewicz, J. L. Sussman, I. Silman, and Y. Shaul, "Context-dependent resistance to proteolysis of intrinsically disordered proteins," *Protein Science*, vol. 20, no. 8, pp. 1285–1297, 2011.

[336] A. R. Mehdipour and G. Hummer, "Cardiolipin puts the seal on ATP synthase," *Proceedings of the National Academy of Sciences*, vol. 113, no. 31, pp. 8568–8570, 2016.

[337] R. Steenbergen, T. S. Nanowski, A. Beigneux, A. Kulinski, S. G. Young, and J. E. Vance, "Disruption of the phosphatidylserine decarboxylase gene in mice causes embryonic lethality and mitochondrial defects," *Journal of Biological Chemistry*, vol. 280, no. 48, pp. 40032–40040, 2005.

[338] G. Tasseva, H. D. Bai, M. Davidescu, A. Haromy, E. Michelakis, and J. E. Vance, "Phosphatidylethanolamine deficiency in mammalian mitochondria impairs oxidative phosphorylation and alters mitochondrial morphology," *Journal of Biological Chemistry*, vol. 288, no. 6, pp. 4158–4173, 2013.

[339] H. Watson, "Biological membranes," *Essays in Biochemistry*, vol. 59, pp. 43–69, 2015.

[340] A. L. Duncan, A. J. Robinson, and J. E. Walker, "Cardiolipin binds selectively but transiently to conserved lysine residues in the rotor of metazoan ATP synthases," *Proceedings of the National Academy of Sciences*, p. 201608396, 2016.

[341] T. H. Haines and N. A. Dencher, "Cardiolipin: a proton trap for oxidative phosphorylation," *FEBS Letters*, vol. 528, no. 1-3, pp. 35–39, 2002.

[342] J. J. García, E. Morales-Ríos, P. Cortés-Hernández, and J. S. Rodríguez-Zavala, "The inhibitor protein (IF_1) promotes dimerization of the mitochondrial F_1F_o-ATP synthase," *Biochemistry*, vol. 45, no. 42, pp. 12695–12703, 2006.

[343] B. Varco-Merth, R. Fromme, M. Wang, and P. Fromme, "Crystallization of the c 14-rotor of the chloroplast ATP synthase reveals that it contains pigments," *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, vol. 1777, no. 7, pp. 605–612, 2008.

[344] A. Wellburn and H. Lichtenthaler, "Formulae and program to determine total carotenoids and chlorophylls a and b of leaf extracts in different solvents," in *Advances in Photosynthesis Research*, pp. 9–12, Springer, 1984.

[345] C. Jiko, K. M. Davies, K. Shinzawa-Itoh, K. Tani, S. Maeda, D. J. Mills, T. Tsukihara, Y. Fujiyoshi, W. Kühlbrandt, and C. Gerle, "Bovine F_1F_o ATP synthase monomers bend the lipid bilayer in 2D membrane crystals," *Elife*, vol. 4, p. e06119, 2015.

[346] S. Maeda, K. Shinzawa-Itoh, K. Mieda, M. Yamamoto, Y. Nakashima, Y. Ogasawara, C. Jiko, K. Tani, A. Miyazawa, C. Gerle, and S. Yoshikawa, "Two-dimensional crystallization of intact F-ATP synthase isolated from bovine heart mitochondria," *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, vol. 69, no. 12, pp. 1368–1370, 2013.

[347] F. Hauer, C. Gerle, N. Fischer, A. Oshima, K. Shinzawa-Itoh, S. Shimada, K. Yokoyama, Y. Fujiyoshi, and H. Stark, "GraDeR: membrane protein complex preparation for single-particle cryo-EM," *Structure*, vol. 23, no. 9, pp. 1769–1775, 2015.

[348] K. Shinzawa-Itoh, J. Seiyama, H. Terada, R. Nakatsubo, K. Naoki, Y. Nakashima, and S. Yoshikawa, "Bovine heart NADH- ubiquinone oxidoreductase contains one molecule of ubiquinone with ten isoprene units as one of the cofactors," *Biochemistry*, vol. 49, no. 3, pp. 487–492, 2009.

[349] M. E. Pullman, H. S. Penefsky, A. Datta, and E. Racker, "Partial resolution of the enzymes catalyzing oxidative phosphorylation i. purification and properties of soluble, dinitrophenol-stimulated adenosine triphosphatase," *Journal of Biological Chemistry*, vol. 235, no. 11, pp. 3322–3329, 1960.

[350] A. Šali and T. L. Blundell, "Comparative protein modelling by satisfaction of spatial restraints," *Journal of Molecular Biology*, vol. 234, no. 3, pp. 779–815, 1993.

[351] S. S. Bird, V. R. Marur, M. J. Sniatynski, H. K. Greenberg, and B. S. Kristal, "Lipidomics profiling by high-resolution LC- MS and high-energy collisional dissociation fragmentation: focus on characterization of mitochondrial cardiolipins and monolysocardiolipins," *Analytical Chemistry*, vol. 83, no. 3, pp. 940–949, 2010.

[352] C. Bechara, A. Nöll, N. Morgner, M. T. Degiacomi, R. Tampé, and C. V. Robinson, "A subset of annular lipids is linked to the flippase activity of an ABC transporter," *Nature Chemistry*, vol. 7, no. 3, pp. 255–262, 2015.