

Limited proteolysis mass spectrometry can provide structural insights into the cellular response to stress

A thesis submitted to the University of Manchester for the degree of
Doctor of Philosophy
in the Faculty of Biology, Medicine and Health

2023

Ellen J Appleton

Division of Cell Matrix Biology and Regenerative Medicine
School of Biological Sciences

Table of Contents

Chapter One: Introduction	13
1.1 Overview	13
1.2 Division of thesis	15
1.3 Protein Homeostasis	16
1.3.1 Determinants of proteostasis	16
1.3.2 The proteostasis network	19
1.4 Consequences of loss of proteostasis	26
1.5 Previous work	28
1.5.1 Human mesenchymal stem cells	28
1.5.2 Loss of proteostasis in early and late passage hMSCs	30
1.6 Measurements of proteostasis	33
1.6.1 Structural proteomics methods	34
1.6.2 LiP-MS	39
1.7 Aims	45
Chapter Two: Materials and Methods	47
2.1 Reagents and Solutions	47
2.2 Equipment	48
2.3 Software	49
2.4 Cell source	49
2.5 Y201s	50
2.6 Cell culture	50
2.7 Cell passaging	50
2.8 Cell freezing	51
2.9 Cell lysis	51
2.9.1 SL-DOC method	51
2.9.2 Dounce homogenization method	51
2.9.3 Triton X-100 method	52
2.10 Protein quantification	52
2.11 Gel electrophoresis	52
2.12 Heat shock of cell lysates	53
2.13 Heat shock of cells in culture	53
2.14 HSP70 inhibition	53
2.15 Mass spectrometry analysis	53
2.15.1 Reduction and alkylation	54

2.15.2 S-trap protein digestion	54
2.15.3 Peptide desalting.....	54
2.15.4 Liquid chromatography mass spectrometry	55
2.16 Progenesis and Mascot data processing.....	55
2.17 MaxQuant data processing	56
Chapter Three: Optimization of the LiP-MS protocol	58
3.1 Optimisation of proteolysis conditions	59
3.1.1 Trial LiP-MS experiment.....	60
3.1.2 Tryptic peptides	62
3.1.3 Nickpred.....	65
3.1.4 Semi-tryptic peptides	67
3.1.5 Peptide mapping.....	71
3.2 Optimisation of lysis conditions	73
3.3 Summary.....	76
Chapter Four: LiP-MS can act as a probe of protein structure in complex samples.....	79
4.1 Tryptic peptides.....	81
4.1.1 Distribution of abundance values.....	83
4.1.2 Significantly changing tryptic peptides	85
4.2 Semi-tryptic peptides.....	87
4.2.1 Distribution of abundance values.....	87
4.2.2 Significantly changing semi-tryptic peptides	88
4.3 Direct comparison of LiP treated samples	92
4.4 Protein examples	93
4.5 Further heat stress: 60 °C experiment	97
4.6 Data processing.....	100
4.7 Summary.....	102
Chapter Five: LiP-MS can provide structural insights into the cellular response to stress.....	105
5.1 Stress-induced changes to protein abundance	108
5.2 Stress-induced changes to protein structure.....	109
5.2.1 LiP-induced alterations to protein signal.....	111
5.2.2 Identification of LiP peptides.....	113
5.2.3 Structural characteristics of LiP peptides	116
5.2.4 GO enrichment analysis of structurally altered proteins.....	125
5.2.5 Post-Translational Modifications	127
5.3 Summary.....	134

Chapter Five: Addendum	137
Chapter Six: Conclusions and Future Work	145
Bibliography	152
Appendix	167

Word Count: 55,079

List of Abbreviations

AD – Alzheimer's disease

ATP - Adenosine triphosphate

BAG - BCL-2-associated athanogene

CHIP - Carboxyl-terminus of HSP70-interacting protein

CMA – Chaperone-mediated autophagy

Cryo-EM - Cryo-electron microscopy

E:S – Enzyme:substrate

EP – Early Passage

HD – Huntington's disease

HDX - Hydrogen-deuterium exchange

hMSC – Human mesenchymal stem cell

HSF1 – Heat shock transcription factor 1

HSP – Heat shock protein

IDP – Intrinsically disordered protein

IDR – Intrinsically disordered region

IF – Immunofluorescence

IM - Ion mobility

INA – Intra-nasal application

LC – Liquid chromatography

LiP – Limited proteolysis

LP – Late Passage

MS – Mass spectrometry

NBD – Nucleotide binding domain

NMR - Nuclear magnetic resonance

PCA – Principal component analysis

PD – Parkinson's disease

PDB – Protein data bank

PK – Proteinase k

PN – Proteostasis network

PTM – Post-translational modification

RT-qPCR - Reverse transcription quantitative polymerase chain reaction

SBD – Substrate binding domain

UPS - Ubiquitin-26S proteasome system

XL - Cross-linking

List of Figures

Figure 1.1: The principal determinants of protein homeostasis.....	17
Figure 1.2: Ubiquitin Proteasome System	24
Figure 1.3: Main autophagic degradation pathways..	25
Figure 1.4: Serial passaging of hMSCs induces a state of replicative senescence.....	28
Figure 1.5: The protein stress response is altered in senescent hMSCs	31
Figure 1.6: Time resolved analysis of HSPA1A transcript and protein levels	32
Figure 1.7: The structural determinants of proteostasis.....	38
Figure 1.8: Visualisation of the LiP-MS workflow.....	41
Figure 1.9: The peptide fingerprint of LiP activity.....	42
Figure 3.1: LiP leads to greater protein fragmentation (SDS PAGE)..	60
Figure 3.2: The effect of LiP concentration on peptide signal (schematic)	61
Figure 3.3: The influence of LiP concentration on tryptic peptide identification	62
Figure 3.4: K means clustering of tryptic peptides LiP-MS data.....	64
Figure 3.5: Nickpred scores by tryptic peptide cluster	67
Figure 3.6: K means clustering of semi-tryptic peptide LiP-MS data.....	68
Figure 3.7: Nickpred scores by semi-tryptic peptide cluster.....	71
Figure 3.8: The influence of LiP concentration on calumenin peptide signal.....	72
Figure 3.9: Triton X-100 does not inhibit the activity of proteinase K	76
Figure 4.1: The number of tryptic and semi-tryptic peptides identified in LiP-MS data.....	81
Figure 4.2: Principal Component Analysis of tryptic peptides	83
Figure 4.3: Tryptic peptide intensity log ₂ (LiP/Control) distribution.	84
Figure 4.4: Significant alterations to tryptic peptide abundance (volcano plots)..	85
Figure 4.5: Semi-tryptic peptide intensity log ₂ (LiP/Control) distribution.	87
Figure 4.6: Significant alterations to semi-tryptic peptide abundance (volcano plots).	88
Figure 4.7: Semi-tryptic peptide signal schematic	89
Figure 4.8: Weighted Venn diagrams of significantly changing semi-tryptic peptides	91
Figure 4.9: Possible interpretation of LiP peptide behaviours	91
Figure 4.10: Direct comparison of LiP-treated samples.....	93
Figure 4.11: Alterations to peptide signal in aldolase A	95
Figure 4.12: Alterations to peptide signal in RACK1	97
Figure 4.13: Analysis of 60 °C LiP-MS experiment	99
Figure 4.14: MaxQuant data processing.....	101
Figure 5.1: LiP-MS workflow for Chapter 5	106
Figure 5.2: Stress-induced protein abundance changes (volcano plot)	108

Figure 5.3: Principal Component Analysis of Protein LFQs	110
Figure 5.4: Number of peptides identified in each sample	111
Figure 5.5: The effect of LiP on protein signal (examples).....	112
Figure 5.6: Alterations to peptide signal in myotrophin	113
Figure 5.7: Significant peptide abundance changes (volcano plots)	114
Figure 5.8: Stress-induced peptide abundance changes.....	115
Figure 5.9: Nickpred scores of LiP peptides	118
Figure 5.10: Secondary structure mapping of LiP peptides.	120
Figure 5.11: Assessment of intrinsically disordered protein regions using MobiDB	123
Figure 5.12: CamSol analysis of LiP proteins.	124
Figure 5.13: GO Biological Processes.....	126
Figure 5.14: GO Molecular Function.	127
Figure 5.15: Barcode plot of PSMD11	129
Figure 5.16: T-complex protein 1 subunit β (CCT2).....	130
Figure 5.17: Alphafold structure of Hsp70 (1B).....	131
Figure 5.18: Alphafold structure of Hsp90 α	132
Figure 5.19: Addendum: number of peptides identified.	138
Figure 5.20: Addendum: coefficient of variance of peptides	139
Figure 5.21: Addendum: normalised peptide intensity	140
Figure 5.22: Addendum: Correlation between normalised peptide intensity across experimental conditions.	141
Figure 5.23: Addendum: number of tryptic and semi-tryptic peptides identified in Chapter 5	142
Figure 5.24: Addendum: contribution of semi-tryptic peptides to the whole protein signal.	142
Figure 5.25: Addendum: Cytoplasmic dynein 1 heavy chain	143

List of Tables

Table 1.1: A summary of major molecular chaperone families and their characteristics.....	20
Table 2.1: Human mesenchymal stem cell donor information.....	49
Table 3.1: Optimisation of lysis conditions for LiP-MS.	74
Table 3.2: Dounce Homogenization (DH) lysis trials	74
Table 3.3: Triton X-100 lysis trials.....	75
Table 4.1: Experimental plan for Chapter 4	80
Table 4.2: LiP peptide table (intermediate and heated).....	94
Table 4.3: LiP peptide table (heated).....	96
Table 5.1: Structural characteristics of LiP peptides.	116
Table 5.2: LiP peptides map to phosphorylation sites.	128
Table 5.3: Complex Portal comparison	133
Table 6.1: Future experimental design.....	1511

Abstract

Proteomics is frequently utilized to identify proteins that are differentially expressed under perturbation or disease conditions and has been instrumental in uncovering the mechanisms of the cellular response to such stresses. However, the effects of perturbation on the proteome are often structural and thus will not be captured by measurements of protein abundance alone. Protein-protein interactions, post-translational modifications and small molecule binding are all events that result in protein conformational change and have important functional consequences. Further, perturbation-induced protein structural changes such as unfolding or aggregation can lead to loss of function or even toxic gain of function.

Limited proteolysis mass spectrometry (LiP-MS) is an emerging structural proteomics technique capable of evaluating changes to protein conformation on an '-omics' scale in complex biological samples. The LiP-MS pipeline employs a non-specific enzyme and restricted digestion time to 'nick' exposed regions of proteins in their native states. Cleavage sites are then identified by mass spectrometry and subsequent bioinformatic analysis.

This thesis first focuses on optimizing and validating LiP-MS for use in primary human cells. The protocol was then utilized to investigate the cellular response to loss of proteostasis, induced through treating cells with heat shock and chaperone inhibitors. This experiment was able to identify hundreds of proteins with altered conformational states. These proteins were significantly enriched for known cellular responses to loss of proteostasis, implying a systematic remodelling of conformations and interactions within the proteostatic machinery in response to stress. Further, specific, peptide level conformational alterations in proteins with well-studied roles in the heat shock response were identified, such as those required for chaperone-client interactions.

Moving forward, LiP-MS has the potential to provide fresh insight into the mechanisms of the cellular response to stress through identifying previously undescribed, functionally critical protein structural changes, the presence and consequences of which can then be further investigated using orthogonal biochemical techniques.

Declaration and copyright statement

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and they have given the University of Manchester certain rights to use such Copyright, including for administrative purposes.

Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see <http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=24420>), in any relevant Thesis restriction declarations deposited in the University Library, the University Library’s regulations (see <http://www.library.manchester.ac.uk/about/regulations/>) and in the University’s policy on Presentation of Theses.

Acknowledgements

I would first like to thank my supervisors Simon and Joe, for their guidance and support.

I am grateful for the support and insight provided by all members past and present of the WTQBB cohort and of the Brennan, Gilmore, Ucar, Swift and Hubbard labs. I am also very appreciative of the help and expertise provided by the University of Manchester BioMS core facility. I would like to thank to Mr. Mobeen Ismail of Trafford General Hospital (TH) and Prof. Tim Broad of Wrightington Hospital (WH) for the provision of human tissue samples for bone marrow-derived mesenchymal stem cell isolation. I would also like to acknowledge the funding body, the Wellcome Trust, who supported this project financially.

Thank you to Anna, for proofreading this thesis and for many excellent business discussions and to Jack, for the support and guidance offered at the outset of this work. Thank you to my family, for their support and constant backing. Finally, thank you to my dear friends for all the laughter, my dear Becky for all the lasagne, and mi vida Javo, for all the love.

Chapter One: Introduction

1.1 Overview

Proteins are a crucial component of the cellular machinery, performing many vital roles including signaling and transport. The concentration, location and structure of proteins are all features that require dynamic regulation in order to meet the numerous functional demands of the cell. A state in which this need is met is referred to as protein homeostasis, or *proteostasis*. Proteostasis can be perturbed by cellular stresses, such as mechanical or thermal stress, leading to protein unfolding. Protein unfolding can result in loss of function and, in particular cases, toxic gain of function through the formation of protein aggregates. However, proteostasis is safeguarded by several mechanisms, known collectively as the proteostasis network [1]. For example, molecular chaperones assist in the *de novo* folding of newly synthesized proteins, as well as resolving misfolding events as part of the stress response [2]. Further, protein degradation pathways enable the removal and recycling of misfolded or excess protein, thus facilitating newly synthesized, native proteins in fulfilling their functions [3, 4].

However, several studies have indicated a correlation between age and decline in proteostasis [1, 5-7]. Age-associated impairment of the proteostasis network, alongside an accumulation of oxidative damage, leads to a cell-wide overload of misfolded or impaired proteins. This results in an increased risk of protein aggregation, which is characteristic of several neurodegenerative pathologies such as Alzheimer's and Parkinson's Disease, for which age is considered the biggest risk factor [8]. Despite the ever-increasing prevalence of such diseases, therapeutic interventions remain overwhelmingly unsuccessful [9]. Therefore, there is a clear and pressing need for better understanding of the fundamental causes and consequences of the loss of proteostasis.

Previous work in the Swift lab has begun to address this need, through assessment of the proteostasis network in cultured primary human mesenchymal stem cells (hMSCs) [10]. There is a specific requirement to better understand proteostasis in hMSCs as they are often used in regenerative medicine. One application of hMSCs is their transplantation to sites of damage, with bone marrow derived hMSCs delivered to the brain in animal models of neurodegenerative diseases [11-13]. Such procedures can cause inflammation, damage to tissue and potentially infection, all of which are known to perturb proteostasis [14]. Furthermore, therapeutic applications often require high numbers of hMSCs, meaning *in vitro* expansion is often required due to the small number attainable *in vivo* [15]. Therefore,

an improved understanding of the effect of cell culture and loss of proteostasis in hMSCs is beneficial towards the development of stem cell therapies [16].

In these previous experiments, loss of proteostasis was induced in donor-matched early and late passage hMSCs by heat shock at 42 °C and the response assessed using transcriptomics, proteomics and immunofluorescence imaging. Late passage hMSCs demonstrated high rates of senescence, which have many phenotypic similarities to aged cells [17, 18]. This analysis revealed strong similarities between the transcriptional stress response in early and late passage hMSCs. However, this was not reflected at the protein level, with late passage cells exhibiting a dampened protein stress response. This indicated that damage to the proteostasis network with senescence may be predominantly post-transcriptional. Incorporation of these findings into a simple mathematical model of four key stress responders led to a prediction of a greater burden of misfolded protein in senescent cells. However, testing of this prediction requires an omics technique capable of assessing protein conformation. Furthermore, a significant lag between transcription and translation was also shown in all cells, suggesting that the role of protein in the early stress response cannot be captured by measures of protein abundance alone. Indeed, the other regulators of proteostasis such as protein relocation and post-translational remodelling of protein structure, such as protein-protein interactions and post-translational modifications, are potentially quicker to enact. Therefore, a structural proteomics technique is needed to gain a fuller understanding.

Limited proteolysis mass spectrometry (LiP-MS) is an emerging structural proteomics technique with the potential to provide novel insight into native protein structure on a cell-wide scale [19]. This is accomplished through utilizing a non-specific enzyme and reduced digestion time to achieve the “limited proteolysis” (LiP) of the native proteome. If these conditions are appropriately optimized, sites of limited proteolysis are dictated by the structure of a protein and comparison of this data over different biological conditions can be used to identify sites of protein conformational change.

The aim of this thesis was to develop the LiP-MS protocol as a tool for providing structural insight into the hMSC stress response. This first required optimization of the protocol for use in hMSCs, with a focus on LiP enzyme concentration, digestion time and the method of cell lysis. It was subsequently established that LiP of hMSCs had been achieved, that this was identifiable through analysis of the LiP-MS peptide signal, and that LiP cleavage sites did appear to be dictated by protein structural features. The capacity of LiP-MS to detect changes to protein structure within complex cell lysates was then established, by perturbing the proteome with heat shock. Analysis of the signal generated by LiP-MS enabled the

identification of differences between the unperturbed and perturbed proteomes, and specific protein unfolding events were discussed. Finally, LiP-MS was applied to stressed and unstressed hMSCs, culminating in the identification of hundreds of proteins with potential stress-induced structural alterations. Moreover, gene ontology enrichment analysis of these structurally altered proteins found they were enriched for many known cellular stress response factors, such as “unfolded protein binding” and “proteasome activating activity”. Thus, when carefully considered, LiP-MS has the capacity to probe both the mechanisms of the proteostasis network and the consequences should these mechanisms fail.

1.2 Division of thesis

This thesis has been divided into six chapters:

- **Chapter 1 – Introduction**

An overview of protein homeostasis and the mechanisms responsible for its safeguarding is provided. The role of loss of protein homeostasis in human ageing and disease is discussed and the importance of a better understanding of the fundamental causes and consequences of this decline stated. Previous work investigating the loss of protein homeostasis in hMSCs is then detailed and how this work would be enhanced by a proteome-wide probe of protein structure highlighted. Following a brief overview of current methods in structural proteomics, limited proteolysis mass spectrometry (LiP-MS) is introduced as the technique that forms the focus of this thesis. The aims of this PhD project are also discussed.

- **Chapter 2 – Materials and Methods**

All reagents, solutions, equipment and software used in this study are disclosed. The methods used to obtain the results described in this thesis are explained in full.

- **Chapter 3 – Results: Optimization of the LiP-MS protocol**

Here optimization of the LiP-MS protocol for use in hMSCs is detailed, with a focus on LiP enzyme concentration, LiP digestion time and the method of cell lysis. A LiP-MS experiment trialing several LiP enzyme concentrations acts as a proof of principle through demonstrating that LiP is identifiable in the MS peptide signal and that LiP cleavage sites correspond with the expected protein structural features.

- **Chapter 4 – Results: LiP-MS can act as a probe of protein structure in complex samples**

In this chapter the capacity of LiP-MS to detect changes to protein structure within complex cell lysates is established. This is tested through perturbing the proteome with heat shock and then investigating the effect of such perturbation on LiP-MS peptide signal. Evidence is provided that an increase in LiP activity correlates with

heat-induced protein unfolding, thus suggesting that LiP-MS can detect protein structural alterations. This is further evidenced through mapping LiP peptides to specific protein structures. Here, the complications arising from MS peptide analysis is discussed, alongside how these complications can be best mitigated.

- **Chapter 5 – Results: LiP-MS can provide structural insights into the cellular response to stress**

In this chapter, the efficacy of LiP-MS as a structural probe of the human cellular stress response is shown. Stress is induced through the application of heat shock and treatment with chaperone inhibitors and LiP-MS performed to identify perturbation-induced structural alterations. Hundreds of proteins with suspected perturbation-induced conformational changes are identified and gene ontology enrichment analysis is used to show that these proteins are enriched for many known cellular stress response pathways. Finally, LiP peptide mapping is used to highlight specific functional alterations in the early stress response.

- **Chapter 6 – Conclusions and future work**

The main findings of this study are evaluated and how the aims of this thesis have been met is discussed. Potential future directions for this work are also detailed.

1.3 Protein Homeostasis

1.3.1 Determinants of proteostasis

Proteins are often referred to as the machinery of the cell, due to the crucial role they play in maintaining cellular and tissue function. They have a multitude of vital jobs including metabolism, signaling, transport and providing structure. Protein homeostasis, or *proteostasis*, is therefore a fundamental component of organismal health. Proteostasis refers to a state in which all proteins are present at the required concentrations, in the required location and in the required conformation at the required time, as shown in Figure 1.1. This is a highly dynamic need.

The 3D structure of a protein is usually central in defining its functionality, hence correct protein folding is imperative for many aspects of cellular homeostasis. Early investigations into the dynamics of *in vitro* protein folding [20] concluded that native protein structure was entirely determined by the specific properties of its amino acid sequence. For example, hydrogen bonding between regions of the polypeptide backbone is responsible for forming common secondary structures such as α -helices and β -sheets. Further, as proteins exist in an aqueous environment, the hydrophobic properties of amino acid side chains lead to changes in tertiary structure, to keep said side chains on the interior of the protein [21].

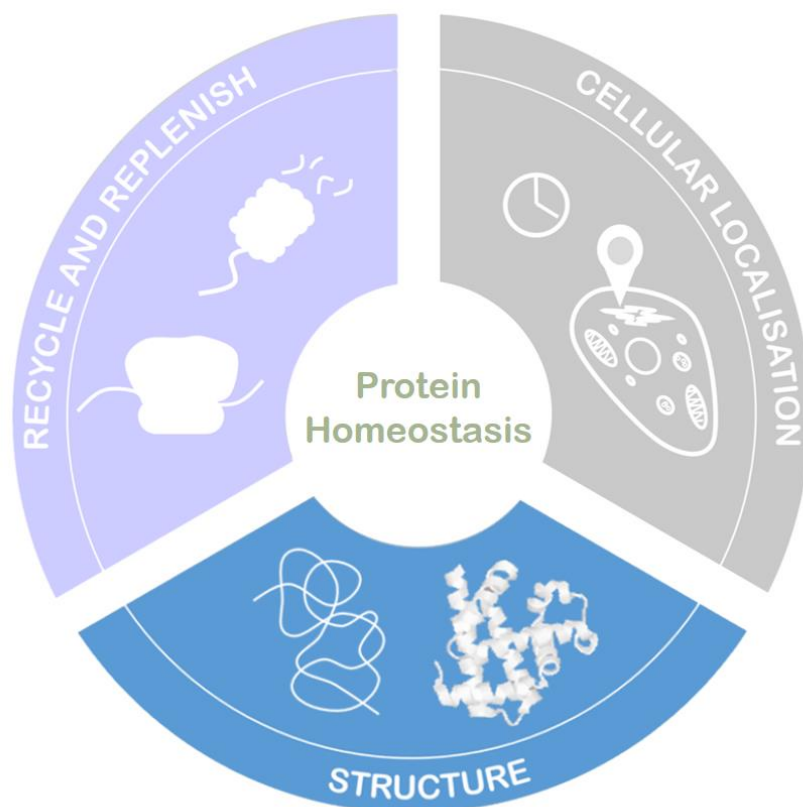


Figure 1.1: The principal determinants of protein homeostasis. Protein homeostasis as a cellular and molecular process can be represented by the three broad categories. Firstly, the folding and structure of a protein is foremost in defining its cognate role. Therefore, the maintenance of proteins in their 3D native conformations is synonymous with maintaining the functionality of the proteome. Next, the capacity to recycle and replenish protein. Influences such as cell cycle stage or the extracellular environment mean the requirements of the cell are dynamic and the availability and abundance of certain proteins must be flexible enough to meet this need. Finally, cellular localization – proteins need to be successfully deployed to their sites of action, a feature which is also strongly influenced by the dynamic cellular environment.

Whilst it is still accepted that small proteins are capable of folding unaided *in vitro*, due to the speed at which they bury hydrophobic side chains, larger proteins, if left unaided, can become stuck in partially folded intermediates [22]. Moreover, the environment in which proteins fold *in vivo* is extremely concentrated, known as macromolecular crowding [23], making it even more difficult for proteins to reach their desired conformation. One way in which this issue can be addressed is by chaperones, which are molecules in the cell responsible for assisting protein folding. Molecular chaperones achieve this in many ways, such as guiding the folding of newly synthesized chains, providing compartments in which proteins can fold away from the concentrated cytosol, and binding to exposed hydrophobic regions during folding to prevent aggregation [24].

A common explanation of protein folding pathways has been the ‘free energy surface’, along with the assumption that final protein conformation is determined by that which is most thermodynamically stable [20]. However, the number of possible conformations is high, resulting in a rugged energy surface and stochasticity in unaided protein folding [25]. These

ideas support the “New View” that molecular chaperones are usually required to assist proteins in traversing the energy landscape and prevent them from settling in folding intermediate states [26]. Further development in understanding protein folding arose from the identification of intrinsically disordered proteins (IDPs). Their presence implies that the structure of the proteome is more dynamic than initially thought [27], and predictions that up to 30% of the proteome is routinely without structure [28].

Furthermore, much of the functional capacity of a protein is acquired post-translationally, largely through various types of molecular interactions. The individual characteristics of each polypeptide chain can result in higher-order protein structural modifications. For example, the association of monomers is highly functionally important, with an estimated 55% of proteins in eukaryotes existing as dimers or higher-order complexes [29]. Much of the vital cellular machinery consists of protein-protein interactions, the cytoskeleton providing one example of many. On the other hand, the capacity of proteins to associate with each other can have deleterious effects if these interactions are unwanted and/or non-native. The oligomerization of certain proteins is associated with neurodegenerative diseases, such as alpha-synuclein and Parkinson’s Disease, where unwanted aggregates and fibrils can form that are considered to be cytotoxic.

Another category of functionally critical protein structural alterations are post-translational modifications (PTMs). For example, one of the most abundant and important PTMs is phosphorylation. Facilitated through protein kinases, a phosphate group is added, usually to serines, threonines or tyrosines, which enables interaction with other molecules, and/or can alter local structure and stability. Phosphorylation is a crucial component in the majority of cellular processes, with phosphorylation events in over two thirds of proteins encoded by the human genome [30]. Protein activity can also be influenced by metabolite binding. Allosteric binding of metabolites is important in the formation and function of many protein complexes [31]. Adenosine triphosphate (ATP) binding, for example, is central to the activity of many molecular chaperones [32], the function of which will be discussed in more detail below.

An important feature of the aforementioned post-translational protein structural alterations and additions is their transience. Phosphorylation is reversible and protein complexes are often capable of formation and dissociation on short time scales. The rapid and reversible nature of metabolite-protein binding has, historically, made identifying interactions challenging [33]. However, it is exactly this transience that enables the dynamic needs of the cell to be met, and proteostasis to be maintained.

Protein subcellular localization is also a key factor in protein function and so proteostasis. Different organelles provide specialized environments both chemically and in terms of

interaction partners. Some proteins are unique to certain cellular compartments, but many change location to regulate different activities [34]. For example, many transcription factors move between the nucleus and cytoplasm based on the requirement for gene transcription and therefore DNA binding [35].

1.3.2 The proteostasis network

Proteostasis can be perturbed by cellular stress. For example, mechanical, thermal, or oxidative stressors can result in protein unfolding and hence loss of function, toxic gain of function, and protein aggregation. However, several cellular safety mechanisms are triggered in response to such stressors in order to reduce protein unfolding and restore proteostasis. These mechanisms are known collectively as the Proteostasis Network (PN) and comprise of some 2000 component proteins [1].

An important factor in the cellular response to stress is the global halting of translation initiation, in tandem with the increased translation of PN components [36]. Protein synthesis is an energetically expensive process and so reduction in general translation frees up energy for the PN machinery and reduces the overall folding burden from newly synthesized proteins that are not essential in the stress response.

Disruption to protein synthesis during stress is regulated by eukaryotic translation initiation factor 2 α (eIF2 α), through phosphorylation at serine 51 [37]. When bound to GTP, eIF2 α transports initiator transfer RNA (tRNA) to ribosomes. Together, initiator tRNA and active eIF2 α is known as the tertiary complex (TC), which is also implicated in AUG start codon recognition. Phosphorylation of eIF2 α at serine 51 prevents GDP/GTP exchange, thus reducing active TCs and, consequently, limiting translation global initiation [38], whilst still allowing for the translation of specific, stress response mRNAs such as activating transcription factor ATF4 [39]. The transcription factor most associated with cellular response to heat stress is HSF1 which will be discussed further [40].

Phosphorylation of eIF2 α in response to stress is regulated by protein kinases and the precise kinase responsible is dependent on the type of stress. For example, heme-regulated inhibitor kinase (HRI) is activated by heat shock and oxidative stress, general control non-repressible-2 (GCN2) is the kinase activated in response to UV irradiation and amino acid starvation, protein kinase RNA (PKR) responds to double-stranded RNAs and PKR-like ER kinase (PERK) activates upon ER stress or protein unfolding [36].

Furthermore, the reduction in translation initiation is further facilitated by the formation of phase separated condensates such as processing bodies (P-bodies) and stress granules (SG). mRNA that has ceased translating can be stored in these condensates and either

released upon the appeasement of stress or de-capped and degraded [41]. Stress granules are more typically associated with storing mRNA and RNA-binding or translation-initiation proteins, whereas P-bodies as the name suggests contain the mRNA degradation machinery in addition to translationally repressed mRNAs [42]. However, there is evidence of interaction between the two condensates, with many shared proteins and mRNAs [43]. This is an active area of current research which continues to uncover novel ways in which cells seek to regulate function via post-transcriptional control of RNA at the translational level, by sequestration and binding of RNAs by proteins into granules of various different types. In addition, as noted above, selected RNAs may also be protected from this general suppression of translation, so genes actively involved in a stress response such as those in the PN continue to be expressed and proteins synthesized. This can be via binding by RNA-binding proteins, which are over-represented in many granule types and indeed translation does appear to continue in some of these condensates [43, 44].

Molecular chaperones

Molecular chaperones represent a crucial component of the PN, with greater than 300 chaperones and co-chaperones thought to be responsible for both assisting in the folding of newly synthesized proteins and refolding proteins that have become conformationally compromised in response to stress [45]. These molecular chaperones, collectively referred to as the *chaperome*, assist in the folding of over 60% of all cellular protein [2]. As noted above, molecular chaperones are often necessary to guide certain proteins across the folding energy landscape, to prevent settling in local energy minima. Chaperones are also responsible for preventing proteins settling into highly stable aggregates, often the global energy minima of intermolecular interactions [26]. In this way chaperones are often classified as “foldases”, “holdases”, or “disaggregases”, based on whether they primarily assist in folding protein, preventing aggregation, or reversing aggregation, respectively. The chaperome is also often subdivided by ATP-dependence and cellular location. Information for the main chaperone families is provided below and listed in Table 1.1.

Table 1.1: A summary of major molecular chaperone families and their characteristics. Molecular chaperones and co-chaperones can be broadly distributed into 5 main chaperone families – HSP70, HSP90, the chaperonins, small HSPs and HSP100. Listed here is the cellular location in which this family is active as well as the main functions they have been observed performing. Foldases are responsible for assisting in the folding of clients proteins. Holdases prevent protein aggregation by

binding to exposed hydrophobic regions on their clients. Disaggregases are capable of reversing or removing protein aggregation.

Chaperone Family	Location	Function
HSP70	Cytosol Nucleus ER Mitochondria	Foldase Disaggregase
HSP90	Cytosol Nucleus ER Mitochondria	Foldase
Chaperonins	Cytosol Mitochondria	Foldase
Small HSPs	Cytosol Nucleus	Holdase
HSP100	Cytosol Nucleus	Disaggregase

HSP70

Heat shock protein 70 (HSP70) proteins are highly conserved, ubiquitous, and associated with several mechanisms of proteostasis such as protein import, assembly, refolding and disaggregation [46]. The structure of HSP70 is characterized by a C-terminal substrate-binding domain (SBD) capable of recognizing and binding extended hydrophobic protein regions that would normally be folded away from solvent. This process is mediated by ATP binding at the N-terminal nucleotide-binding domain. Delivery of client proteins to the ATP-bound, open conformation of HSP70 is facilitated by co-chaperones from the HSP40 family. This process initiates hydrolysis of the bound ATP, resulting in a closed HSP70-substrate complex. The substrate is released following NEF-triggered ADP/ATP exchange. In this way the HSP70 family prevents exposed hydrophobic residues from forming undesirable inter-molecular interactions and facilitates the folding of proteins to their native structure [47].

HSP90

HSP90 chaperones are also highly conserved and ubiquitously expressed, with many hundreds of client proteins. In mammalian cells, members of the HSP90 family have different localizations; HSP90 α and HSP90 β are specific to the cytosol, 94 kDa glucose-regulated protein (GRP94) the ER, and tumor necrosis factor receptor-associated protein 1 (TRAP-1) the mitochondria [48, 49]. The general structure of HSP90 chaperones consists of a N-terminal domain responsible for ATP binding, a middle domain involved in ATP hydrolysis and substrate binding, and a C-terminal domain associated with dimerization which is

essential for HSP90 function. In eukaryotes, the N-terminal and middle domains are connected by a charged linker region. Client protein folding by HSP90 is a complex process involving interactions with several other chaperones and co-chaperones. For example, clients bound by the HSP70/HSP40 machinery can be additionally bound by HSP90 when facilitated by co-chaperones, such as Hop (STI1).

Chaperonins

The term chaperonins refers to a family of 60 kDa heat shock proteins and can be divided into two main classes. HSP60 is a type 1 chaperonin specific to the mitochondria in eukaryotes. It is active when held in a complex with its co-chaperone, HSP10. Two large rings, each consisting of seven HSP60 molecules are held back to back to form a barrel-like structure. HSP10 molecules then bind to the open ends of this “barrel”, enclosing client proteins inside the structure and so enabling their folding in a reduced macromolecular concentration [50].

Conversely, type 2 chaperonins are localized to the cytoplasm of eukaryotic cells and do not require co-chaperones. For example, the CCT/TriC complex consists of eight subunits that associate to form a football-like structure with an in-built “lid” in the apical domain, capable of closing to encapsulate substrates following ATP binding [51]. Major CCT/TriC clients include the cytoskeletal proteins actin and tubulin, with some evidence of transient reactions with approximately 10% of newly synthesized proteins [52]. The transference of client proteins to CCT/TriC is facilitated by upstream chaperones such as the HSP70 family and the prefoldins [53].

Small HSPs

The small heat shock proteins (smHSP) describe a class of ATP-independent chaperones typically 12-43 kDa in size. There are currently ten identified in mammalian cells, each sharing an α -crystallin domain between more versatile N/C-terminal domains. These chaperones typically recognize and bind to exposed hydrophobic residues of misfolded proteins, thus preventing aggregation. These client proteins are subsequently refolded or degraded by other families of chaperones or co-chaperones [54]. The smHSPs are often referred to as holdases as they are not thought to actively refold or disaggregate their clients, but instead to stabilize their conformation until the foldase, disaggregase or degradation machinery becomes available [55].

HSP100

The HSP100 family are molecular chaperones capable of disassociating protein aggregates and are thus referred to as disaggregases. Working in conjunction with HSP70, substrates

are threaded through the central pore of the HSP100 hexamer, a process mediated by ATP binding and hydrolysis, leading to unfolding [56]. HSP100 has been found in bacteria, plants and yeast, but not in humans. Disaggregase activity in humans is thought to be primarily controlled by interactions between HSP70, HSP40 and HSP110 [57], however the mechanism of action is still the source of much debate [58].

Chaperone activity, both constitutively and in response to stress, is regulated both transcriptionally and post-translationally. At the transcriptional level, transcripts encoding chaperone proteins are upregulated in response to stress, in order to match the increased workload due to stress-induced protein unfolding. This change is effected by heat shock transcription factors, the most well-studied of which is heat shock transcription factor 1 (HSF1) which is often referred to as the master regulator of the chaperome [59]. The mechanism by which HSF1 is activated is somewhat debated, but a favoured explanation is the chaperone titration model [60, 61]. In homeostasis, HSF1 is bound to an abundant molecular chaperone, HSP70. HSP70 is itself a crucial component of the stress response and, under stress conditions, is titrated away from HSF1 to aid in protein refolding. Free HSF1 then trimerises and is translocated to DNA to induce transcription of various chaperones, including HSP70. Upon elimination of misfolded protein, HSP70 is then free to rebind HSF1 and thus a negative feedback loop is formed.

Protein degradation

An alternative way in which the threat posed by misfolded proteins can be mediated is through degradation, either via the proteasomal or autophagic protein degradation pathways.

The Ubiquitin-26S Proteasome System (UPS) is capable of degrading proteins in an ATP-dependent manner. Misfolded proteins are marked for degradation with a protein ligand ubiquitin, which causes them to be recognized by the 26S proteasome where they are digested into peptides within the proteolytic core. The marking of misfolded proteins with ubiquitin is mediated by three associated enzymes; E1 and E2 are responsible for activating and conjugating ubiquitin, and E3 is the ubiquitin-protein ligase responsible for identifying specific substrates and facilitating the binding of activated ubiquitin via a substrate lysine residue [3], forming an isopeptide bond, as shown by the schematic in Figure 1.2. An example of a ubiquitin E3 ligase is CHIP (carboxyl-terminus of HSP70-interacting protein), which is also a molecular chaperone. CHIP, in combination with BAG1, is thought to interact with HSP70 and mark protein clients that are not re-foldable for degradation via this system. CHIP is also capable of ubiquitinating HSP70 itself, but demonstrates a preference for its

clients. CHIP may therefore act as a regulator of HSP70 activity, by instigating the degradation of excess HSP70 upon the appeasement of stress [62].

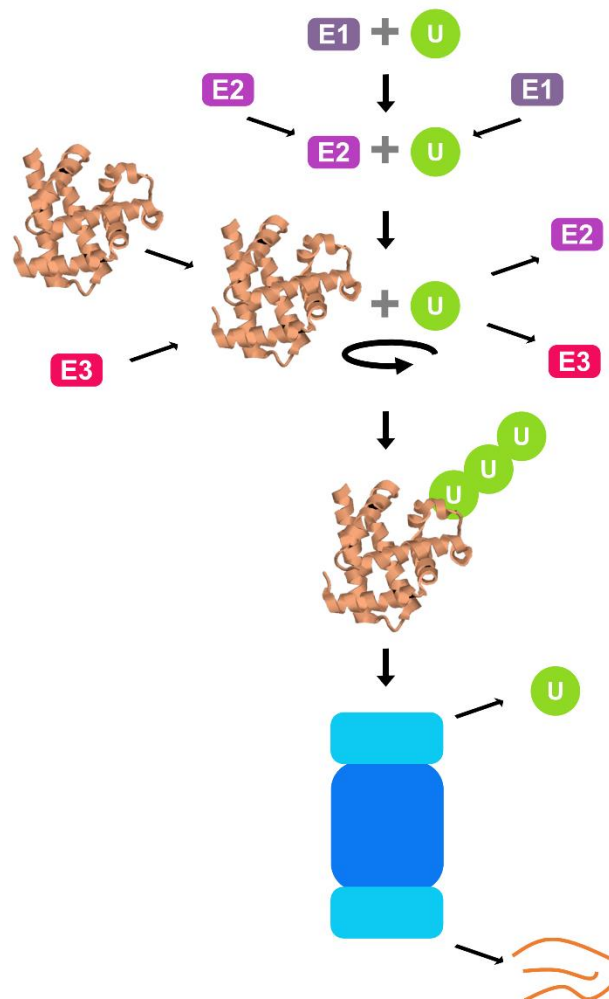


Figure 1.2: Ubiquitin (U) is activated by enzyme E1 through creating a thio-ester bond in an ATP dependent manner. Activated, E1 bound ubiquitin is then transferred to a ubiquitin conjugating enzyme (E2). E2 facilitates the binding of the C-terminal of ubiquitin to the substrate protein, initially via a lysine residue. E3 is the ubiquitin-protein ligase crucial for catalyzing this reaction, and for further binding of ubiquitin to itself, creating a chain rooted in the substrate. E3 is responsible for the substrate specificity of the UPS. Ubiquitinated substrates are then recognized by the 19S-20S-19S (26S) proteasome where the substrate is unfolded and digested into peptides and ubiquitin is recycled.

Alternatively, proteins can be removed via the autophagic protein degradation pathway that relies on the delivery of proteins to lysosomes – membrane-enclosed organelles with acidic interiors. Lysosomes contain a variety of hydrolytic enzymes, including proteases and nucleases, that are able to digest macromolecules. The digested substance is then transferred from the lysosome to the cytosol, facilitated by transport proteins within the organelle's membrane. The three main mechanisms of autophagy are macroautophagy, microautophagy and chaperone-mediated autophagy, distinguishable by the manner in

which proteins are delivered to the lysosome and summarized briefly by the schematic in Figure 1.3.

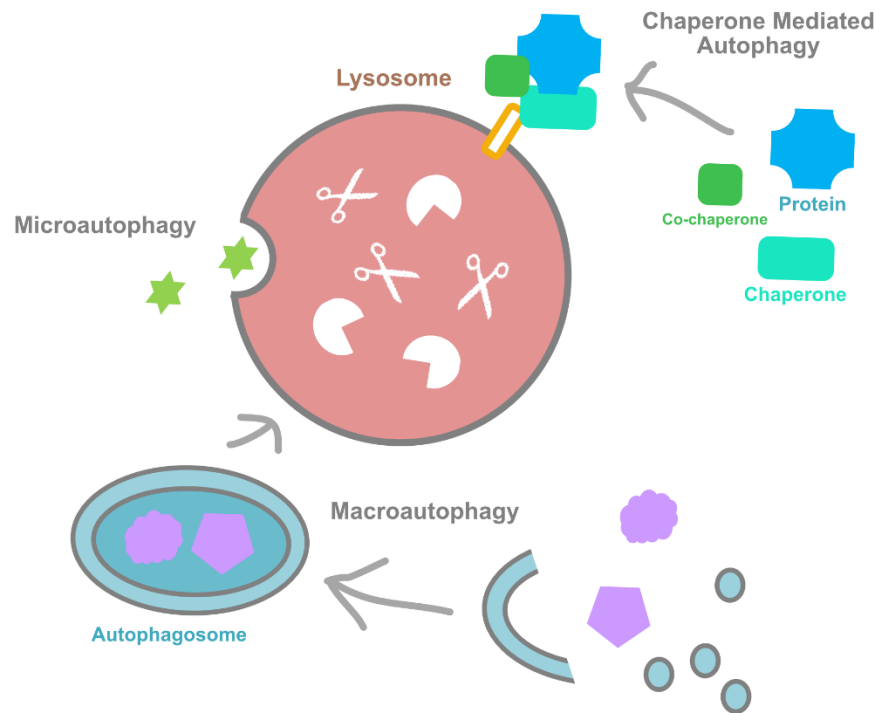


Figure 1.3: Main autophagic degradation pathways. During macroautophagy, vesicles fuse together to enclose macromolecules in a double-membraned compartment called an autophagosome. The autophagosome then fuses with the lysosome, where its contents are digested. Microautophagy is characterised by direct engulfment of cytosolic material by the lysosome, via invagination of the lysosomal membrane. Chaperone mediated autophagy is a selective form of autophagy and involves molecular chaperones recognising specific protein motifs to identify substrates. The chaperone-substrate complex is transported to a receptor on the lysosomal membrane (LAMP2A), where the substrate is unfolded and delivered into the lysosome, with the help of co-chaperones.

Macroautophagy pertains to the transport of damaged proteins via double membraned, intracellular compartments called autophagosomes, formed by the fusion of vesicles when signaled by protein kinases. Unlike the UPS, macroautophagy is capable of removing large structures, such as protein aggregates and organelles. Autophagosomes then fuse with lysosomes, where their contents are digested and recycled. In contrast, microautophagy does not require autophagosomes and instead involves the engulfing of cytosolic material by the lysosome itself, via invagination of the lysosomal membrane. Finally, chaperone-mediated autophagy refers to the transport of protein to the lysosome via chaperones. Molecular chaperone heat shock cognate protein 71 kDa (HSC70) recognises and binds to a particular pentapeptide motif on the substrate protein, as part of a chaperone complex with co-chaperones. This complex then interacts with a lysosomal membrane receptor, lysosomal-associated membrane protein 2A (LAMP2A), where the substrate unfolds and is delivered into the lysosome [4, 63].

1.4 Consequences of loss of proteostasis

Critically, a breakdown or dysfunction in proteostasis can lead to cellular damage, and several studies have revealed a correlation between age and decline in proteostasis [1, 5-7]. With age, factors such as decreased proteasomal and chaperone activity, alongside an accumulation of oxidative damage, lead to a system-wide overload of misfolded or impaired proteins. Not only can compromised protein structure lead to a loss of function, but exposed hydrophobic residues increase the risk of proteins forming unwanted attachments, leading to potentially toxic aggregates. Protein aggregation is characteristic of several neurodegenerative diseases, such as Alzheimer's (AD), Parkinson's (PD), and Huntington's Disease (HD) [64, 65], for all of which age is considered the most significant risk factor [8]. Understanding and tackling these diseases is more important than ever as the ageing population has increased their prevalence. For example, there are currently thought to be approximately 44 million cases of AD worldwide, with the potential to reach over 100 million by 2050 [9]. Meanwhile, drug development for this disease is progressing poorly, with 99.6% of clinical trials failing [9]. The causes of this dismal success rate are the source of much debate [66, 67], but what is abundantly clear is the need to better understand the fundamental causes and consequences of the loss of proteostasis.

Many investigations into the decline of proteostasis with age have implicated an impaired molecular chaperone response [68-70]. For example, production of HSP70, the most abundant chaperone of the heat shock response, is attenuated in aged rats compared to their young counterparts [71, 72]. Similarly, studies have shown an improved HSP70 response in human centenarians when compared to other aged humans [73]. In addition to this, perturbations of HSF1, the transcription factor often considered the master regulator of the heat shock response, in *Caenorhabditis elegans* (*C.elegans*) have provided further evidence for the link between the heat shock response and healthy ageing. Namely, decreasing active HSF1 was shown to accelerate ageing phenotypes, whereas overexpression of HSF1 led to increased lifespan [74].

More generally, studies of human brain ageing revealed that 32% of the chaperome is repressed with age [74]. This result was again echoed in *C.elegans* by the discovery of many functionally critical chaperones with an age-associated decrease in expression [45].

Given this clear connection between decreased chaperone activity and decline in proteostasis, many clinical studies of protein folding diseases have explored molecular chaperone augmentation as a possible treatment. Therapeutic interventions studied so far include the development of "pharmacological chaperones", or the use of RNA to alter ageing signaling pathways [75]. For example, mutating the molecular chaperone domain Bri2

BRICHOS in *ex vivo* mouse brain tissue was shown to alleviate amyloid formation [76]. Furthermore, decreasing insulin growth factor-1-like signaling in *C.elegans* led to reduced aggregation in an Alzheimer's disease model [77].

Despite this, many questions remain. For example, is the inadequacy of the molecular chaperone network in age a consequence of the decreased expression of key chaperones, or due to the sequestering of available chaperones by an increasingly damaged proteome [78]? Is there an equal burden on all molecular chaperone families, or is this placed more heavily on a few "key" chaperones?

There is also some evidence of an age-associated attenuation of the protein degradation machinery. Reduced or impaired proteasome activity has been linked to ageing, such as the decrease in active 26S proteasomes with age in *Drosophila melanogaster* [79]. This has also been shown in human fibroblasts, with reduced proteasome activity shown to be a feature of senescence [80]. Moreover, fibroblasts subjected to protease inhibitors began to display senescent phenotypes [81]. Finally, a correlation between increased proteasomal activity and longevity has been shown in studies of the exceptionally long-lived naked mole rate [82, 83] as well as the fibroblasts of healthy centenarians [84].

Impaired autophagy has been implicated as a cause of proteostatic damage in several tissues [85, 86]. *In vivo* evidence implicates a decrease in the abundance of proteins critical to successful autophagy, with a reduction in active LAMP2A shown to be detrimental to chaperone-mediated autophagy in rats [87]. This is supported by evidence that overexpression of critical autophagic proteins can ameliorate ageing. Increasing autophagy protein 5 (Atg5) levels in mice, a protein crucial for the formation of autophagosomes, led to both extended lifespan and reversal of ageing phenotypes [88]. This is reflected in human studies, as a main regulator of autophagy was found to be overexpressed in healthy centenarians [89], implicating autophagy as a factor in human longevity.

As with molecular chaperones, the link between the decline in protein degradation pathways and ageing has been exploited in clinical studies of age-associated disease. This is demonstrated by the reduction of aggregates in *in vitro* models of AD [90] and HD [91] as a result of the therapeutic enhancement of the proteasomal machinery.

However, much remains unclear regarding the exact relationship between the mechanisms of protein degradation and decline in proteostasis with age. For instance, AD associated aggregates have been shown to cause proteasomal damage [92], whereas proteasomal damage has been shown to cause HD aggregates [93].

1.5 Previous work

Previous work in the Swift lab has begun to investigate the proteostasis network in primary human mesenchymal stem cells (hMSCs), with a particular focus on how the response to loss of proteostasis may be altered in senescent cells.

1.5.1 Human mesenchymal stem cells

In this thesis and previous work, hMSCs harvested from bone marrow have been used as an experimental system. Primary human cells have the advantage of biological relevance but are also very well suited to studying the effect of external perturbations such as mechanical stress, heat shock, or drug treatment. Furthermore, serial passaging of hMSCs *in vitro* induces a state of replicative senescence, arrested division, highly relevant to ageing since aged tissues have a much higher incidence of this cell type.

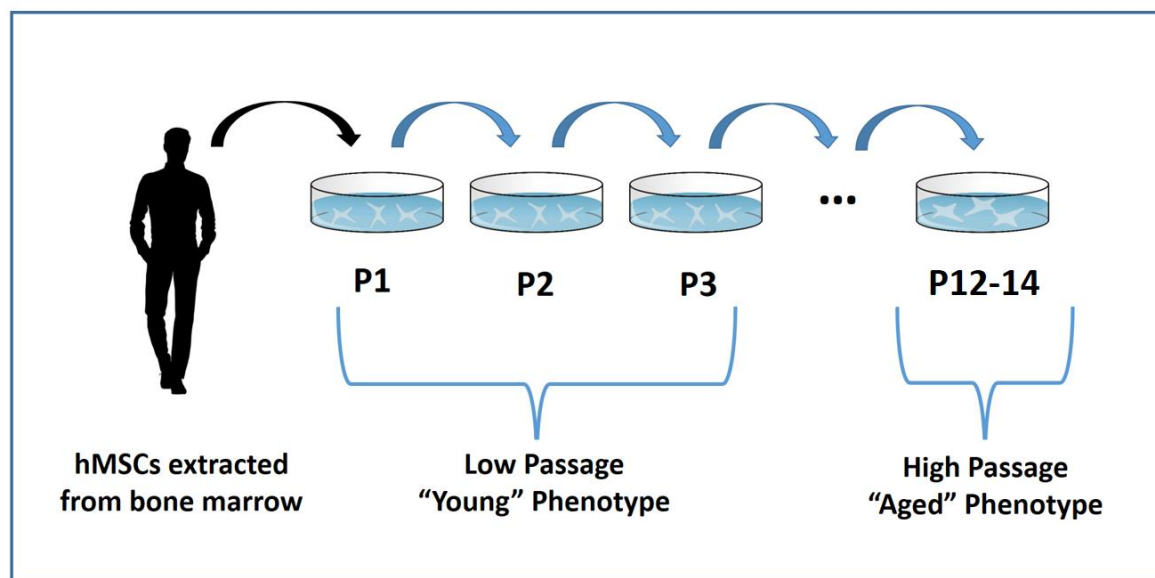


Figure 1.4: Human mesenchymal stem cells from bone marrow donors are placed in cell culture. The cells are then serially passaged until they reach senescence. The high passage, predominantly senescent cells then represent the aged phenotype, and the low passage cells the young phenotype.

Replicative senescence describes an irreversible interruption of the cell cycle in response to telomere attrition. This is known as the Hayflick limit, and reflects the continuous shortening of telomeres with each DNA replication, until replication is no longer possible. Once the Hayflick limit has been reached, a DNA damage response is triggered, which in turn triggers senescence. This is thought to be an anti-cancer mechanism, as it prevents the propagation of damaged cells throughout tissues, as senescent cells are subsequently recognised and cleared by the immune system.

Senescent cells are further characterised by many morphological and molecular changes. With senescence, hMSCs become larger and rounder, have decreased lamin B expression

and increased β -galactosidase activity. They are further characterised by the senescence associated secretory phenotype (SASP), a pro-inflammatory response that facilitates the recruitment of immune cells [94]. Functionally, senescent cells have been shown to have decreased proteasomal [80, 95] and mitochondrial [96-98] activity, as well as an increase in oxidative protein damage [5].

Senescent cells are disproportionally represented in aged tissues [17, 18], perhaps due to insufficient clearing as a result of the age-associated deterioration of the immune response [99]. Therefore, understanding senescence is part of understanding ageing, as senescence itself is considered to be one of the nine key hallmarks of ageing [99]. Moreover, senescent cells have many phenotypic similarities to aged cells. Many of the characteristics discussed above such as reduced protein degradation capacity, increased protein and mitochondrial impairment and accumulation of DNA damage are all also classically associated with cellular ageing.

Replicative senescence can be induced in hMSCs via repeated passaging *in vitro*. As shown in Figure 1.4, cells are isolated from human bone marrow donors and placed in cell culture. The low passage cells then represent the “young” phenotype and the high passage, senescent cells the “aged” phenotype. These can then be compared to identify age-associated discrepancies in proteostasis. A further advantage of this method is the capacity to compare donor-matched cells.

As with any model system, cell culture is not without issue, the most common criticism being its inability to replicate the *in vivo* extracellular environment. Altered mechanical cues, proliferative capacity and lack of inter-cellular interactions undoubtedly have adverse effects on cell behavior [100]. However, *in vitro* studies can act as a starting point and are well suited to studying the effect of external perturbations. Furthermore, the applications, both potential and realized, of cultured hMSCs in regenerative medicine mean gaining a more detailed understanding of the effects of cell culture and stress on this cell type could be fundamental in advancing clinical studies, since stem cells would be the primary source of material.

Indeed, stem cell based therapies are a rapidly expanding field, capable of treating a wealth of human diseases and injuries [101]. Therapeutic applications of hMSCs uncovered so far range from cartilage tissue engineering [102-104] to tackling age-associated pathologies such as neurodegenerative diseases. Neurodegenerative diseases are often characterized by loss of neurons, such as the reduction of dopaminergic neurons in the substantia nigra of PD patients. Stem cell transplantations are currently being developed to alleviate this loss. For example, in 2011, Danielyan et al., successfully delivered bone marrow-derived MSCs to

several areas of the rat brain, via intranasal application (INA). Further, this technique was shown to lead to improved motility in a rat model of PD [12]. The same group later successfully delivered MSCs intranasally to the brain in transgenic PD and AD mice models [11]; further demonstrating that the use of this technique led to an increased survival rate in an HD mouse model [13]. Generally, over the past ten years much progress has been made in the INA of stem cells and is reviewed here [105].

Alternatively, other studies have sought to utilize the protective capacity of IN-delivered hMSCs. For example, this method was used to prevent neurological complications arising from radiotherapy treatment for brain cancer. In this case, human MSCs were cultured to passage 4-7 and then delivered intranasally to radiation-damaged mouse brains, resulting in improved neurological function and increased brain injury repair [106].

Furthermore, the hypo immunogenic properties of MSCs mean they are increasingly being used in cell-based immunomodulatory therapies, particularly to increase the success of organ transplantation [107]. For instance, treatment with donor-derived MSCs was found to aid in kidney transplant recovery [108].

A systematic review of MSC clinical trials found that over 350 trials were conducted between 2004 and 2018 that involved the use of bone marrow-derived MSCs. The review suggested that the dose of cells (number administered) was an important contributor to the efficacy of the trials, with doses of 70 million or lower proving to be less effective [15]. Due to this requirement for high numbers of MSCs to maximize therapeutic potential, and the small numbers attainable *in vivo*, expansion of MSCs *in vitro* is often required. This creates a specific need for an improved understanding of the effect of cell culture and repeated *in vitro* passaging of hMSCs [16]. For example, there is some evidence that differentiation potential is biased towards certain lineages after repeated passaging, a detail that could be critical in the success of clinical applications [109]. What's more, MSC transplantation carries the risk of inflammation, damage to tissue, and potentially infection, all of which are triggers of the heat shock response [14]. Further, fevers are a common consequence of MSC transplantation, which has the potential to induce heat stress in the transplanted cells [110].

Therefore, an improved understanding of the effects of cell culture, senescence, and heat shock in bone marrow derived-hMSCs is important in improving the efficacy of stem cell based therapies.

1.5.2 Loss of proteostasis in early and late passage hMSCs

In previous work from our lab, loss of proteostasis was induced in donor-matched early and late passage (EP, LP) hMSCs via 2 hours of heat shock at 42 °C. The response was then

characterized using a combination of transcriptomics, proteomics and immunofluorescence imaging [10]. The transcriptomic response was similar in EP and LP cells. There was a substantial amount of overlap between transcripts significantly changing with stress in each condition and among those commonly upregulated were many known stress responders such as the HSP70 chaperone family. Further, the product moment correlation coefficient was calculated for the log2 fold change of each transcript with stress and demonstrated good correlation between EP and LP cells ($R^2 = 0.4633$). On the other hand, the proteomics data showed little similarity between the stress response in EP and LP cells. There was very little overlap between the proteins changing with stress, and the product moment correlation coefficient in this case found no correlation between stress-induced change in protein abundance in the EP cells compared to the LP ($R^2 = 0.0049$). Moreover, comparison of the protein level and transcript level EP stress response also revealed little correlation ($R^2 = 0.0048$), with the same phenomena observed but to an even greater degree in the LP cells ($R^2 = 0.0002$). All comparisons are shown in Figure 1.5.

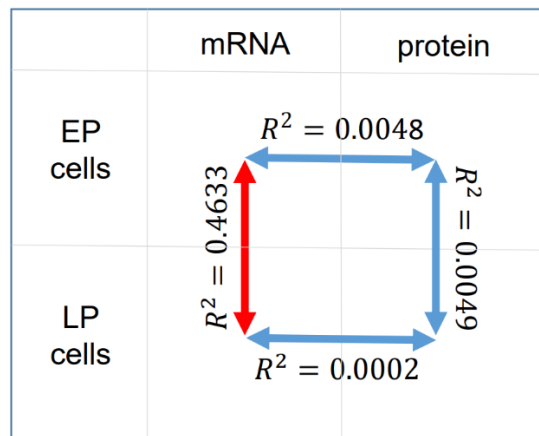


Figure 1.5: Correlation of RNA and protein log2 fold changes for stressed vs unstressed cells for early passage and late passage cells from [10]. R^2 shows the (Pearson) product-moment correlation coefficient, with the arrows indicating the comparison being made. For example, the top centre figure shows the correlation between the proteomic and transcriptomic stress response in early passage cells and the bottom centre this same analysis but for late passage cells. The data on the left shows the correlation between the transcriptomic response in early vs late passage cells (highlighted in red to demonstrate evidence of correlation), and the right the same for the proteomic response.

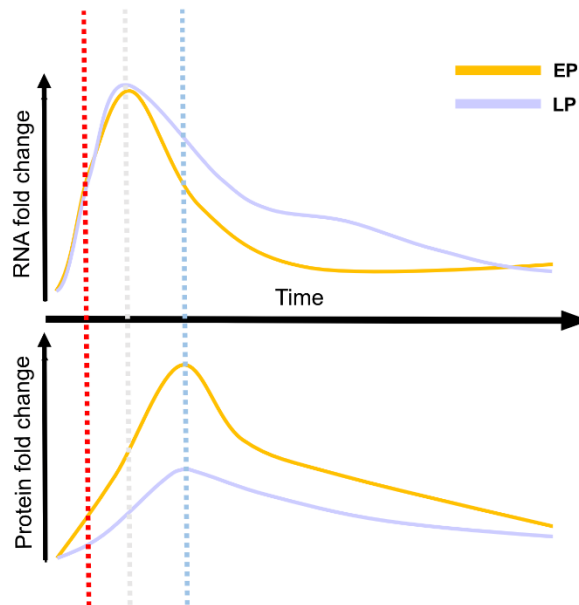


Figure 1.6: Temporal comparison of the transcriptomic and proteomic stress response of HSPA1A [10]. The RNA response (top) was found to be similar in EP (orange) and LP (lilac) cells in terms of magnitude and timing, with the peak response reached at approximately one hour (grey dashed line) post heat shock (red dashed line). Conversely, the protein response was dampened in LP cells, and the peak response was reached later, at approximately 4 hours post heat shock (blue dashed line).

The behaviour of specific proteins of interest was then investigated further, with a particular focus on molecular chaperones. For example, the cellular concentration of HSPA1A protein (stress-inducible HSP70) was measured before, during and up to 24 hours after heat shock using immunofluorescence (IF) imaging, for both EP and LP cells (summarised in schematic form, Figure 1.6). A similar analysis of HSPA1A RNA was also performed using RT-qPCR (summarised, Figure 1.6). Aligning with the previous data, the initial response to heat stress at the RNA level is similar for EP and LP cells, both in terms of magnitude and timing, with peak response reached at approximately one hour post heat shock. One difference highlighted was the LP cells seemingly take a lot longer to recover to pre-stress levels of HSPA1A RNA. At the protein level, the HSPA1A stress response does not reach its peak until 4 hours post heat shock, in both instances. This indicates a lag between the transcriptional response to stress and the time it takes for this to become a protein abundance response to stress. This has been observed previously. A temporally resolved study of heat shock in *Saccharomyces cerevisiae* (*S.cerevisiae*) demonstrated that the stress-induced change to transcripts peaked between 30 and 60 minutes after the application of heat stress, whereas the majority of proteins affected by stress were most differentially abundant after 3 hours [111]. This lag may provide an explanation as to the lack of correlation between transcriptomic and proteomic data as mentioned above, as the omics data was sampled immediately post heat shock. Furthermore, once again the initial protein response to heat shock was dissimilar in early and late passage cells both in terms of magnitude and timing as the response is both delayed and dampened in LP cells. Therefore,

whilst the transcriptional stress response appears to be preserved in senescent cells, the translational stress response appears to be lessened.

To predict the consequences of this attenuation, a system of ordinary/delay differential equations was built. This model focused on the behaviour of four key stress responders: the transcription factor HSF1, the molecular chaperone HSPA1A, the E3 ubiquitin ligase CHIP, and the abundance of misfolded protein. The parameters and bounds of the model were informed by the IF imaging data and optimised against the EP response. The model predicted a greater burden of misfolded proteins in senescent cells. However, the veracity of this prediction was only subject to limited testing within the scope of the study.

Overall, this previous work suggested that damage to the stress response in senescent cells is predominantly post-transcriptional. However, it is also clear that, due to the lag between transcription and translation, the role of protein in the early stress response cannot be captured by measures of protein abundance alone. Indeed, perhaps the other regulators of proteostasis such as protein relocation and post-translational protein remodelling – protein-protein interactions, PTMs, protein-small molecule interactions – are quicker to enact and so form the crux of the early stress response. The techniques used so far are not capable of providing insight into these features. Instead, a structural proteomics approach is required to complement and enhance these findings. Further, the effects of a damaged or depleted proteostasis network are predominantly structural. The burden of misfolded protein and even the subsequent aggregation are factors that are critical to understand and can also be measured using structural proteomics.

Therefore, omics scale structural studies have the capacity to both probe both the mechanisms of proteostasis network and the consequences should these mechanisms fail.

1.6 Measurements of proteostasis

As discussed above, proteostasis refers to a state in which all proteins are present at the required concentrations, in the required location and in the required conformation at the required time, as summarized in Figure 1.1. To gain a complete understanding of the factors governing cellular proteostasis, and its loss, one must be able to accurately measure each of these features at the cell-wide level.

For example, cumulative improvements in instrument sensitivity have meant that MS is capable of determining the abundance of many thousands of proteins directly from complex cell lysates. Moreover, truly outstanding advances in sample preparation methods and ultra-high sensitivity MS have enabled proteome analysis at the single cell level [112, 113].

Techniques to gain a systems level understanding of the dynamics of subcellular localization of proteins have also seen development. For example, organelle fractionation combined with MS enabled the absolute quantification and spatial resolution of thousands of proteins in HeLa cells. This technique also proved capable of capturing protein translocation events following perturbation [34]. Moreover, the development of an extensive and unique catalogue of antibodies as part of the Human Protein Atlas project has enabled mapping of a significant number of human proteins. Cell Atlas is now able to provide spatial information for over 60% of the human proteome. High-resolution immunofluorescence images have facilitated the mapping of thousands of proteins to specific subcellular sites and are further capable of identifying changes to protein localization across cells [114].

However, the focus of this thesis is how one can gain insight into the structural aspects of proteostasis. A brief history of protein structural studies and their development is provided below and limited proteolysis mass spectrometry is introduced.

1.6.1 Structural proteomics methods

In 1958 the first protein structure was unveiled, owing to the development of X-ray crystallography [115]. Since then, X-ray crystallography has gone from strength to strength, with technical improvements made in both efficiency and resolution, resulting in the thousands of high-resolution protein structures now available on the Protein Data Bank (PDB). Despite being near unparalleled in the level of detail provided, X-ray crystallography still has many caveats as a probe of protein structure. For example, protein complexes, membrane proteins, and flexible protein regions remain difficult to crystallize. Further, this technique lacks the capacity to capture the dynamics of protein structure and requires them to be fully folded. Finally, large amounts of purified protein are needed to attain structural information.

A later developed alternative to X-ray crystallography is Nuclear Magnetic Resonance (NMR) [116]. This method is similar to X-ray crystallography in its capacity to provide high-resolution protein structural information, but has the advantage of being performed in solution, meaning flexible regions of proteins can be measured. Despite this, NMR struggles to determine the structure of large proteins and also requires large amounts of purified protein.

Most recently, cryo-electron microscopy (cryo-EM) is seeing a rise in popularity as a technique to determine protein structure [117]. Technological advances have greatly improved the sensitivity of this method and it has proved particularly adept at resolving the structure of membrane proteins. Perhaps most importantly, developments in cryo-electron

tomography have facilitated the study of large macromolecular assemblies directly *in situ* [118].

MS-based structural proteomics methods lack the atomic resolution of X-ray crystallography, NMR and cryo-EM, but some do facilitate a systems-level insight into protein structural dynamics [119, 120]. A small selection of structural proteomics techniques: hydrogen-deuterium exchange MS (HDX-MS), cross-linking MS (XL-MS), and ion mobility MS (IM-MS) are discussed briefly below, all of which have undergone significant technical improvements in the past 10 years [121].

Hydrogen-Deuterium Exchange

When proteins are in solution, hydrogens present on exposed regions of the protein are in constant exchange with the hydrogens present in the surrounding solution. HDX-MS exploits this principle [122] by replacing the hydrogens in solution with deuterium – an isotope of hydrogen distinguishable by its additional atomic mass unit. This means that amide hydrogens on the backbone of each amino acid (except proline) are capable of exchanging with deuterium. Deuterated peptides are then identifiable in MS data, due to the shift in mass. The accessibility and dynamics of protein regions are thought to correlate with exchange speed, with flexible regions such as loops having high exchange rates compared to the substantially slower exchange of buried regions. Therefore, by adjusting the time proteins spend in the deuterium buffer prior to MS analysis, structural information can be acquired [123].

To date, HDX-MS has been used many times to provide much needed insight into the conformational dynamics of various proteins and protein complexes [124-128], as well as protein-protein interactions [129]. In addition, it has been utilized as an effective tool for probing protein folding dynamics [130, 131], the response of structure to pH change [132] and analysis of intrinsically disordered proteins [133]. However, while HDX-MS is consistently applied to purified proteins or protein complexes its efficacy as a proteome-wide probe of structure is in doubt.

Ion Mobility

Ion mobility mass spectrometry (IM-MS) is another example of a structural proteomics technique in which recent technical advancements have led to a surge in popularity [134]. In short, IM-MS uses the differential rates of travel of protein ions through a gas-filled tube to glean information about protein structure. In drift-time ion mobility mass spectrometry (DTIMS), the oldest form of IM-MS, protein ions enter a drift tube, usually filled with helium, in which a uniform electric field is applied. Ions of distinct structures travel through the drift

tube at different rates, as extended ions interact with the gas more and hence travel slower [135]. From this, the ‘collision cross sectional area’ Ω (CCS), defined below, can be calculated and used to inform on the protein structure, with ions identified by mass spectrometry analysis.

$$\Omega = \frac{3ze}{16N} \left(\frac{2\pi}{\mu k_B T} \right)^{\frac{1}{2}} \frac{1}{K_0}$$

K_0 = Standard mobility, z = Ion charge state, e = Elementary charge, N = Density of the drift gas, μ = Reduced mass of the ion–neutral drift gas pair, k_B = Boltzmann constant, T = Gas temperature.

The biggest question of IM-MS is whether aspects of native protein structure are maintained in the gas phase. Early investigation revealed that the addition of water vapor to the gas phase encouraged native conformation [136]. It has since been demonstrated that aspects of native protein structure – from secondary to quaternary – can be maintained in the gas phase and hence can represent a “near native” structure [137-139].

Such developments have seen IM-MS successfully assess protein structural dynamics, including the identification of folding intermediates in aggregation prone proteins [140] and the structure of oligomers formed during the dynamic assembly of amyloid fibrils [141, 142]. Other benefits of this method include the small amounts of sample needed, the ability to isolate ions from complex samples, and the capacity to differentiate between very similar ions. Furthermore, IM-MS was recently used to investigate the qualities of recombinant proteins overexpressed in crude cell lysates [143].

Despite this, unfortunately, the link between CCS and structure is not wholly determinative. In a recent comparison between CCS values determined with IM-MS and those determined from known crystal structures of the same protein, IM-MS values were lower than their crystal structure derived counterparts in every single example. This issue was exacerbated in larger proteins and those with a lack of structure in the centre [144].

Chemical Cross-Linking

Chemical cross-linkers most often consist of a pair of molecules with reactive groups covalently joined together by a spacer group. The reactive molecules can have various properties, and the spacer group has many possible lengths and degrees of flexibility. These qualities can be refined and exploited to obtain information regarding protein structure and dynamics when cross-linkers are applied to protein samples. For example, when cross-

linkers bind to certain protein residues, identifiable by subsequent MS, the spacer group length and flexibility can provide information as to which regions of the same protein interact with each other, or perhaps protein-protein interactions within complexes.

A major issue in the field of chemical cross-linking mass spectrometry (XL-MS) is the post-MS identification of cross-linked residues. This is due to significant expansion in search space caused by the binding of the residues. However, the incorporation of affinity tags, the introduction of photo-cleavable cross linkers and the creation of XL-MS tailored data analysis pipelines have meant huge progress has been made [145].

Thus far XL-MS has been shown to be an effective method of probing the various levels of protein structure. For example, the technique has been recently used to model the quaternary structure of a large protein complex [146], to investigate change in the tertiary structure of individual proteins [147] and to identify secondary structures [148]. In the latter, however, some issues with the XL-MS identification of β -sheets were highlighted.

Until 2008 XL-MS was only shown capable of probing the structure of single proteins, or purified samples of small protein complexes. This was due to both the computational complexity of considering all possible peptide-peptide combinations and the predominance of non-cross linked peptides in whole cell lysates. Rinner *et al.* [149] overcame these issues by combining isotopically labelled cross-linkers with an in-house developed specialist search engine (xQuest) to identify cross-linked peptides in a whole *Escherichia coli* (*E.coli*) lysate. Despite much success, this method proved capable of identifying only the most abundant cross-linked peptides. Improvements in scoring mechanisms and labelling techniques led to an increase in cross-linked peptide identifications in *E.coli* as well as some in the more complex *C.elegans* [150]. The most recent developments resulting in the identification of >2000 cross-linked peptides from a whole HeLa cell lysate using the software XlinkX [151].

Finally, whilst most XL-MS studies have been carried out *in vitro*, the introduction of membrane-permeable cross-linkers now make it possible to utilize XL-MS to study *in vivo* protein-protein interactions. For example, a membrane permeable, enrichable, and MS cleavable cross-linker was shown able to probe protein-protein interactions directly in mammalian cells [152].

Proteostasis and structural proteomics

In order for any structural proteomics technique to act as an appropriate tool for the study of proteostasis, it will ideally meet the following criteria:

1. Provide the level of spatial resolution required to probe each of the structural determinants of proteostasis, as summarized in Figure 1.7. These features are all

examples of post-translation protein remodeling that either form part of, or trigger, the PN. For example, mechanical, thermal, or oxidative stressors can result in protein unfolding and aggregation. Molecular chaperones form a crucial component of the PN response to such stressors, and often require ATP binding or complex formation with co-chaperones in order to become functionally active [50, 153].

2. Able to assess the structure of proteins in as near- native state as possible. This ensures the results are physiologically relevant.
3. Can supply proteome-wide structural information. This is required to gain a systems-level insight into the cellular response to perturbation.

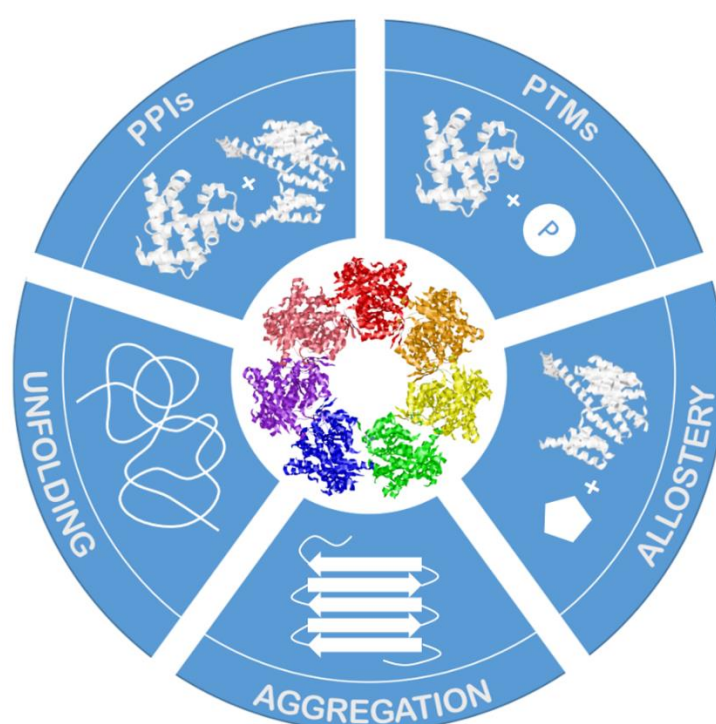


Figure 1.7: The structural determinants of proteostasis. An overview of the different types of structural alterations associated with the maintenance or loss of proteostasis. Protein unfolding and misfolding can occur more frequently in response to cellular stress. In certain cases, protein un/misfolding can result in aggregation, in which exposed hydrophobic residues bind together to create large insoluble structures. Increases in such structures are a result of loss of proteostasis, which would otherwise be able to clear them. On the other hand, protein-protein interactions (PPIs) are a crucial component of the PN, such as the association of chaperones with co-chaperones and their client proteins. PTMs are also heavily involved in proteostasis. For example, phosphorylation of eIF2 α is vital for the inhibition of translation in response to stress. Finally, allostery plays an important role in regulating the protein components of the PN. For example, many chaperone-client interactions are reliant on ATP.

The proteomics methods discussed here meet these criteria to varying degrees. For instance, there is much evidence showing HDX-MS to be an in-depth probe of the various levels of protein structure and their dynamics, but it is currently unable to provide omics level folding information. IM-MS has been successfully utilized to assess the dynamics of aggregate formation, but, at the outset of this project, has not yet been shown capable of

proteome-wide investigation. In contrast, not only is XL-MS able to extract structural information from human whole cell lysates, but membrane-permeable cross linkers facilitate *in vivo* protein studies, and thus the criteria for physiological relevance is met. However, these larger scale experiments have focused on identifying protein-protein interactions and so whether XL-MS can identify local changes to protein structure is uncertain.

Limited proteolysis mass spectrometry (LiP-MS) is another recent development in the field of structural proteomics and has the potential to meet the above criteria for the study of proteostasis. This is the focus of this thesis and is discussed in more detail below.

1.6.2 LiP-MS

Limited proteolysis mass spectrometry (LiP-MS) is a structural proteomics technique with the potential to provide novel insight into native protein structure on a global scale. This is accomplished through utilizing a non-specific enzyme and reduced digestion time to achieve the “limited proteolysis” (LiP) of the proteome. Sites of limited proteolysis should be dictated by the structure of a protein and comparison of this data over different biological conditions can be used to identify sites of protein conformational change.

The understanding and interpretation of LiP data requires a robust definition of LiP cleavage sites. An early study provided evidence for the importance of segmental mobility [154]. In this work, the protease subtilisin was used to probe the structure of thermolysin and the properties of the resulting LiP cleavage sites were investigated. A significant correlation was found between cleavage sites and B values – crystallographic temperature factors indicative of the dynamic properties of the polypeptide chain, by describing the mean squared displacement of each atom. From this it was concluded that flexibility is a significant factor in determining cleavage sites for LiP. Several further examples of the requirement of flexibility for LiP cleavage sites of globular proteins have since been published [155].

The capacity for local unfolding has also proved a significant requirement for LiP cleavage [156-158]. Whilst accepting the correlation between flexibility, accessibility and cleavage sites, these studies stressed the importance of the substrates competence in hosting the protease without having to undergo serious conformational change. When considered in the context of secondary structure, α -helices were found to be capable of such local unfolding but β -strands were largely not.

In contrast to this, some argued for the importance of surface accessibility over regional flexibility as the main determinant of LiP cleavage [159]. Spherical probes were used to uncover enzyme-accessible surface regions of thermolysin, trypsinogen, ribonuclease A and subtilisin, with the radii of each probe based on structural data for each protein. The results

generally demonstrated a strong correlation between B-values, surface accessibility and cleavage sites, aside from a cleavage of trypsinogen that was shown to be surface accessible but with a low B-value. However, whilst perhaps demonstrating that surface accessibility is indeed an important factor in determining LiP cleavage sites, this study lacked scope to make general statements regarding the relative importance of different factors. A later, more comprehensive study took advantage of the increase in available, high-resolution 3D protein structures, and much larger databases of proteolytic events to compare the relative significance of different structural features in determining proteolysis sites [160]. Analysis of over 300 documented cleavage sites revealed a hierarchy of structural attributes, namely exposure>flexibility>local interactions. The same analysis of the influence of secondary structure demonstrated a ranking of loops> α -helices> β -strands. It was further noted that little correlation was found between primary sequence and LiP sites, as expected. However, when these findings were used to predict cleavage sites on known protein structures the false prediction rate was extremely high, around 50%. This could be somewhat explained by potential cleavage sites that have not yet been documented.

However, despite some disagreement as to the importance of the different features, LiP can still be used as a powerful tool for assessing protein structural changes. This is because, when comparing two conditions, a difference in LiP activity will infer conformational change, as each of the suggested determinants of LiP cleavage sites are themselves inherently structural characteristics. For instance, if a region of Protein A is resistant to LiP cleavage in Condition 1, but not in Condition 2, whether due to a change in flexibility, surface exposure, or local unfolding capacity, this still implies the region has undergone some structural alterations.

The LiP-MS workflow is visualized in Figure 1.8. First, cells from experimental condition 1 (left) and experimental condition 2 (right) are lysed in a manner that best preserves the native protein structure. The LiP step is then performed as quickly as possible, using an enzyme:substrate (E:S) ratio and digestion time that has been carefully tuned to ensure only accessible, flexible protein regions are subject to cleavage.

Non-specific enzyme activity is then quenched. Following this, the protein samples are digested to completion with a highly specific enzyme such as trypsin, and the resultant peptides are identified via MS [19].

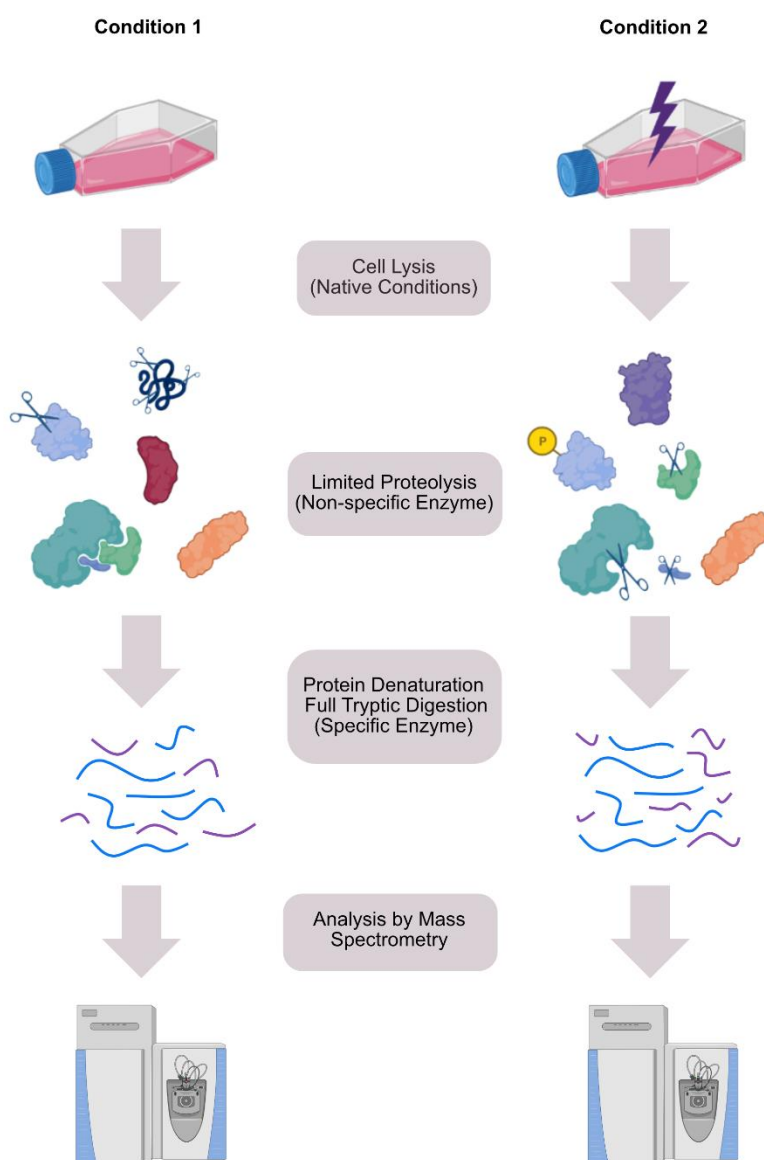


Figure 1.8: Visualisation of the LiP-MS workflow. Cells representing two different experimental conditions of interest are lysed in a manner which best preserves native protein structure. A non-specific enzyme (often proteinase-k) is then applied to the lysates for a limited digestion time. Enzymatic activity is quenched and the proteome then fully digested into peptides with a highly specific enzyme such as trypsin. These peptides are then identified by mass spectrometry. Analysis of the “types” of peptides identified are then used to pinpoint LiP cleavage sites and how LiP accessibility, and so structure, may be changing across the chosen conditions. Created with BioRender.com.

LiP cleavage sites, and hence regions of protein conformational change across samples, can then be identified by looking at the “type” of peptides present in the MS data. Trypsin consistently cuts at the C terminus of the amino acids arginine and lysine, meaning peptides created due to the activity of trypsin (tryptic peptides) are identifiable, as they adhere to

these rules. However, the activity of our LiP enzyme will result in additional, non-sequence-specific cuts along the protein, and hence the presence of semi-tryptic peptides.

Therefore, LiP activity is identifiable in the MS data by a decrease in tryptic peptide signal and a corresponding increase in semi-tryptic peptide signal, as visualised in Figure 1.9. By comparing LiP activity over different experimental conditions, protein regions that have undergone structural alteration can be identified. Peptides mapping to such regions of differential LiP activity over conditions are often referred to as “LiP peptides”.

It is important to note that the retrieval of such data relies upon the stability of the proteins of interest. If a protein is unstable, a single enzyme cut could lead to unravelling of the entire protein structure [156]. In this case, the proteolysis will no longer be limited, and the protein potentially digested to completion. Hence, LiP cleavage sites will no longer correlate with physiological protein structural change.

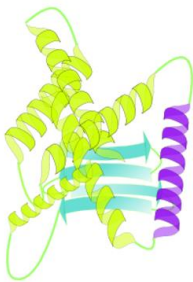
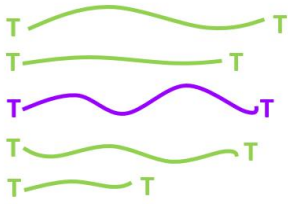

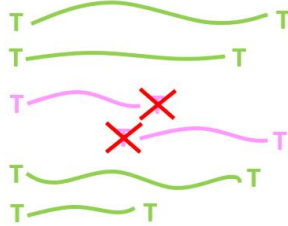
	Structure	Peptide Signal
Condition 1		
Condition 2		

Figure 1.9: The peptide fingerprint of LiP activity. In Condition 1, the protein region highlighted in purple is not cleaved by the LiP enzyme and so the peptides mapping to this area will be tryptic. However, in Condition 2, this protein region has become unfolded (highlighted in pink) and is now subject to LiP cleavage. Therefore in the MS data this will be observed as a loss of tryptic peptide signal and an increase in semi-tryptic peptide signal mapping to this region. Therefore, through analysis of the type of peptides identified in the MS data, sites of differential LiP activity, and so structural change, can be found.

The earliest examples of LiP-MS focused on the investigation of individual proteins or protein complexes. Fontana (1997) first demonstrated the capacity of LiP (and multiple proteases) as a probe of protein secondary structure [161]. This was shown in the context of horse myoglobin, by comparing the structure of holomyoglobin and its heme-dissociated counterpart apomyoglobin. Previous investigations into the structure of myoglobin through X-

ray crystallography, circular dichroism and NMR spectroscopy had suggested that holomyoglobin and apomyoglobin are structurally similar, globular, largely α -helical secondary structure, apart from a specific site, named on holomyoglobin as α -helix F. LiP was performed on both proteins and the results analyzed using gel electrophoresis. Even at short reaction times, apomyoglobin was observed to be split into 2 fragments, whereas under the same experimental conditions holomyoglobin remained intact. This was further investigated using MS and the precise cleavage site mapped to helix F. These results imply helix F is unfolded in apomyoglobin and verify LiP as capable of providing insight into secondary structure. Further work from the Picotti lab [162] used a combination of amino acid replacement and LiP-MS to show that proline 88 is the residue responsible for this local unfolding.

Advances in the sensitivity of MS then led to LiP becoming capable of identifying protein structural changes when said protein was spiked into complex biological samples. For example, the above result was replicated with the apo/holomyoglobin spiked into a yeast proteome extract [163]. Further, LiP-MS was also successfully employed to identify higher order structural changes such as protein aggregates, specifically the conformational changes of α -synuclein [163]. The protein was spiked into a yeast proteome extract in both its unfolded (M- α -syn) and folded/fibrillous (F- α -syn) state and then analyzed by LiP-MS. Peptides mapping to the β -sheet rich amyloid core of the protein were identified in their fully tryptic form in the F- α -syn sample, whereas semi-tryptic peptides mapping to this region were only identified in the M- α -syn sample. In this way, LiP-MS was shown to be capable of identifying protein conformational change at the quaternary level and hence the formation of protein aggregates.

Further evidence of the effectiveness of employing LiP to uncover higher order protein structural alterations include the identification of linker regions and domains of *E.coli* transcription factor NtrC [164], as well as analysis of the alternative conformations of prothrombin [165]. However, the identification of protein aggregates by LiP-MS has thus far only been shown to be possible when the exact conformation of the aggregated form of the protein is known. Given the complexity of LiP-MS data, which will be discussed further throughout the results chapters of this thesis, the identification of unknown aggregates is challenging.

The latest advances in LiP-MS mean the technique now has the capacity to analyze whole cell lysates. Therefore, LiP-MS can act as a proteome-wide probe of protein structure. For example, the method was used to investigate the thermostability of several proteomes, including yeast and *E.coli*. Cell lysates were treated to increasing temperatures and the

proteome analyzed by LiP-MS. Greater LiP activity, identifiable by the LiP-MS peptide fingerprint of reduced tryptic peptide signal and increased semi-tryptic peptide signal, was considered to infer protein unfolding. Through quantification of this fingerprint across the different temperature treatments, a “melting point” was defined for the identified proteins [166]. The study then utilized network analysis to gain systems-level insight into the effect of heat on the proteome. For instance, the analysis revealed that loss of cellular function due to heat in *E.coli* is because of the unfolding of a small subset of proteins that are functionally crucial to bioprocesses. Many correlations were also observed through the investigation of the properties of identified unfolded proteins. These included the thermostability of proteins negatively correlating with their length, and positively correlating with their abundance.

This study also compared the thermostability analysis of *E.coli* lysates to the same analysis but of purified protein samples. The data sets demonstrated little similarity ($R^2 = 0.1322$), thus indicating that the native environment is crucial for thermostability. This illustrates the importance of proteome-wide studies for investigating effects on proteostasis.

Moreover, the capacity of LiP-MS to identify post-translational alterations to protein structure is not limited to protein unfolding. Indeed, the latest studies have demonstrated the efficacy of LiP-MS in providing insight into many functionally critical processes. For example, in the most comprehensive LiP-MS study to date, the technique was used to probe the proteostatic response to thermal and oxidative stress in yeast [167]. Analysis of the protein abundance data revealed little change, whilst many LiP peptides were identified, mapping to hundreds of proteins with potential stress-induced structural alterations. Functional enrichment analysis was performed for these proteins in each stress condition and many of the known stress responses were highlighted. For example, heat shock in yeast is known to initiate protein unfolding, as well as a molecular chaperone response and associated terms such as “protein refolding” were enriched. Further, in osmotic stress, glycerol production increases, and terms such as “glycerol metabolic process”, “NADPH regeneration”, and “NADPH metabolism” were also enriched.

Moreover, LiP peptides (those with differential enzyme accessibility pre and post stress, identifiable by the LiP “fingerprint” discussed above) demonstrated a strong correlation with sites of functional importance. For example, in the case of Gdp1, a MAP kinase target, three LiP peptides were identified, one mapping to the proteins active site, one in proximity to a known phosphosite and the final mapped to the C-terminal domain, a region of substrate-induced conformational change. LiP peptides also mapped to known phosphosites, identified through comparison with a parallel phosphoproteomics experiment. Thus, LiP-MS was capable of detecting a number of stress-induced phosphorylation events.

The structurally altered proteins in heat-stressed yeast were significantly enriched for molecular chaperones. One such chaperone was the protein disaggregase ATPase Hsp104. In this case, the LiP peptides identified mapped to the chaperone's ATP binding site or substrate channel. The capacity of Hsp104 to bind substrates is thought to be ATP mediated, and therefore the differential enzyme-accessibility of these sites with heat treatment seems to indicate a stress-induced activation of Hsp104, as has been reported previously [168].

In addition, LiP-MS has also recently been used to assess protein-metabolite interactions in *E.coli* [169]. In this study, the protein binding partners for 20 metabolites of interest were identified through comparing LiP activity in proteome extracts in which the endogenous metabolites were removed, to those in which 2 physiologically relevant concentrations of the metabolite had been added. In this way, decreased LiP activity inferred metabolite binding. These LiP peptides mapped to known protein-metabolite interaction sites, identified novel interactions and also indicated metabolite-induced higher order structural changes such as protein complex formation or dissociation. The identification of known interaction sites was verified through calculation of the minimum distance between LiP peptides and metabolic binding sites as recorded by the Protein Data Bank Ligand Expo. A small selection of novel interactors were verified using *in vitro* enzyme assays and metabolite-induced protein quaternary structure alterations were validated using size exclusion chromatography MS.

Overall, previous studies have shown that LiP-MS can act as a probe of native/near-native protein structure at the omics scale and can provide the level of spatial resolution required to probe each of the structural determinants of proteostasis stated in Figure 1.7. However, LiP-MS has thus far mainly been used as a systems-level probe of cellular processes in yeast and *E.coli*. To my knowledge, this approach has not yet been used in primary human cells.

1.7 Aims

The primary aims of this thesis can be summarized as follows:

- **To establish the efficacy of LiP-MS for use with hMSCs.** Although demonstrated in other cellular systems, this technique has not previously been used with hMSCs. Therefore, the LiP-MS protocol will be optimized for use in this system, with a focus on optimal LiP enzyme concentration, LiP digestion time and the method of cell lysis. The influence of LiP on MS peptide signal will be used to verify that LiP has been achieved. Further, the correlation between sites of potential LiP cleavage and the known structural determinants of LiP will be tested.
- **To use LiP-MS to identify protein-structural alterations on a proteome-wide scale.** Using conditions determined from the first aim, the denaturing effect of heat

on the proteome will be tested as a perturbation to evaluate if LiP-MS is able to detect induced regions of protein unfolding, by virtue of alterations to the detectable peptide signal. Further, the robustness of LiP-MS data in hMSCs will be assessed and quality control procedures implemented.

- **To determine the capacity of LiP-MS to provide structural insights into the cellular response to stress.** LiP-MS will be performed on hMSCs subject to heat shock. In this instance, heat shock is applied to intact hMSCs, rather than lysates as was the case in aim two. This will enable the identification of both heat-induced unfolding events, and the activation of functional stress response pathways, through analysis of structural LiP-MS data. The proteins and protein regions identified will then be further investigated, both with regards to the enrichment or depletion of signaling pathways associated with the PN, and potential loss of function structural alterations, such as protein folding or aggregation. Functional structural changes will then be compared to the current literature regarding the stress response. Loss of function structural changes will be compared to the known structural determinants of LiP sites, and to protein stability data available in the literature.

Chapter Two: Materials and Methods

2.1 Reagents and Solutions

- 2-phenylethynylsulfonamide (Sigma Aldrich, #P0122)
- Acetonitrile (Thermo Fisher Scientific, #10629112)
- Ammonium bicarbonate (Sigma-Aldrich, #09830)
- Bicinchoninic acid (BCA) assay kit (Thermo Fisher Scientific, #23227)
- BioRad 4-20% acrylamide pre-cast gradient gels (BioRad, #456-1094)
- Bromophenol blue (Sigma-Aldrich, #B5525)
- Copper(II) sulphate (Sigma-Aldrich, #12852)
- Dimethyl sulphoxide (Sigma-Aldrich, #D2650)
- Dithiothreitol (Roche, #10197777001)
- Dulbecco's Modified Eagle Medium (DMEM), low glucose, pyruvate (Gibco, #31885-023)
- Dulbecco's phosphate buffered saline (Sigma-Aldrich, #D8537)
- Foetal bovine serum (Labtech)
- Glycerol (Thermo Fisher Scientific, G/0650/08)
- Glycine (Thermo Fisher Scientific, G/0800/60)
- Hepes (Sigma-Aldrich, #H4034)
- Industrial methylated spirit (IMS) (Thermo Fisher Scientific, #M/4450/17)
- Instant Blue (Expedean)
- Iodoacetamide (Sigma-Aldrich, #I1149)
- Liquid nitrogen (BOC Ltd)
- Magnesium chloride (Sigma Aldrich, #M3634)
- OLIGO R3 beads (Thermo Fisher Scientific, #1-1339-03)
- Penicillin-streptomycin (Sigma-Aldrich, #P0781)
- Phosphatase inhibitor cocktail (Sigma-Aldrich, #P0044)
- Protease inhibitor cocktail (Sigma-Aldrich, #P8340)
- Proteinase K (Promega, #V3021)
- Proteomics grade trypsin (Promega)
- Sodium chloride (Thermo Fisher Scientific, #S/3160/65)
- Sodium deoxycholate (Sigma-Aldrich, #D6750)
- Sodium dodecyl sulphate (Sigma-Aldrich, #75746)
- Spectra Broad Range protein marker (Thermo Scientific, #26634)

- Tris-HCL (Thermo Fisher Scientific, #10316893)
- Triton X-100 (Sigma Aldrich, #X100-500ML)
- Trypsin-ethylenediaminetetraacetic acid (EDTA) solution (Sigma-Aldrich, #T3924)
- Ultrapure water (Millipore)

2.2 Equipment

- 0.5, 1.5 and 2 mL tubes (Starlab Ltd.)
- 1.5 mL Protein LoBind tubes (Eppendorf, #022431081)
- 1.6 mm diameter steel beads (Next Advance)
- 26 gauge syringe needle (BD Biosciences, #300300)
- 96-well filtration membrane plate (Corning, #3504)
- Block heater (Techne)
- Bullet Blender Tissue Homogeniser (Next Advance)
- Centrifuge, refrigerated, with 15 mL, 50 mL, and plate attachments (Eppendorf, #5810R)
- Class II Biological Safety cabinet (Thermo Fisher Scientific)
- CO2 incubator (New Brunswick)
- Cryovials (STARLAB, #E3370-6122)
- Dounce homogeniser (Wheaton, #257538)
- Dry incubator (Genlab)
- Freezer, -80°C (New Brunswick)
- Hemocytometer (NanoEntek)
- Liquid chromatography system (Dionex Corporation)
- Liquid nitrogen storage system (Thermo Fisher Scientific)
- Microscope (Olympus, CKX31 inverted)
- MS sample vials (Thermo Fisher Scientific)
- Nitrile gloves (STARLAB)
- pH meter (Hanna Instruments)
- Plate reader (BioTek, #ELX800)
- Q Exactive Orbitrap Mass Spectrometer (Thermo Fisher Scientific)
- S-trap columns (ProtiFi)
- Tissue culture-treated culture dishes (Corning, #430641U)
- Vacuum centrifuge (Heto Cooling System)
- Vortex mixer (Thermo Fisher Scientific)
- Water bath (Grant Instruments)

2.3 Software

- Affinity Designer (Serif, version 1.10.5.1342)
- Mascot Daemon (Matrix Science UK)
- MATLAB (MathWorks, R2017a)
- MaxQuant (version 1.6.7)
- Nickpred [170]
- Prism (GraphPad, version 9)
- Progenesis QI (Nonlinear Dynamics, version 4.1)
- Protti (version 0.3.1, [171])
- Python (version 3.7.9)
- RStudio (version 4.2.1)
- CamSol (version 2.2, [172])
- MobiDB (version 5.0, [173-175])
- Complex Portal (accessed April 2019, [176])
- PhosphoSitePlus (version 6.7.1.1, [177]).

2.4 Cell source

Primary human bone marrow derived mesenchymal stem cells were retrieved from knee and hip replacement surgeries using established protocols, with the informed written consent of all donors [178]. The experiments detailed in this thesis adhered to the guidelines and regulations dictated by the World Medical Association Declaration of Helsinki and the UK Human Tissue Authority. Approval was granted by the NHS Health Research Authority National Research Ethics Service (approval number 10/H1013/27). The gender and age of each donor used is detailed in Table 2.1, alongside information regarding the source of the hMSCs.

Table 2.1: Human mesenchymal stem cell donor information. Here the sex, age and extraction site of hMSCs used in this thesis are listed, alongside the anonymised donor ID.

Donor ID	Gender	Age	Source
WH143	M	58	Hip
TH269	M	70	Knee
TH194	M	65	Knee
TH309	F	69	Knee
TH305	M	67	Knee
WH318	M	34	Hip

2.5 Y201s

The cells used in Chapter 4 of this thesis were an immortalised, clonal line of human mesenchymal stem cells referred to as Y201s. Generated by James et al [179], immortalisation was achieved through overexpression of telomerase reverse transcriptase (TERT). The Y201 line they created demonstrated exponential growth but with no evidence of tumorigenicity and a cell surface marker profile consistent with MSCs. Further, Y201s were still amenable to chemical induction of various lineages and have been shown to maintain the mechano-responsiveness of primary hMSCs [180]. All aspects of cell culture, lysis and analysis of Y201s was exactly the same as for hMSCs, as detailed in the below sections.

2.6 Cell culture

Human MSCs were submersed in low-glucose Dulbecco's Modified Eagle Medium (DMEM) with pyruvate (Gibco, #31885-023) with additional 10% (v/v) fetal bovine serum (FBS, Labtech.com) and 1% (v/v) penicillin/streptomycin cocktail (PS, Sigma-Aldrich, #P0781) and cultured in tissue culture-treated polystyrene (TCTP, Corning). MSCs were incubated (New Brunswick) at 37 °C and 5% CO₂. Copper sulphate solution (Sigma-Aldrich, #12852) was placed in the incubators in order to prevent microbial contamination. Media changes and passaging were performed in a Class II Biological Safety cabinet (Thermo Fisher Scientific) using nitrile gloves (STARLAB). The cabinet and all equipment were sterilised with 70% (v/v) IMS in MilliQ ultrapure water (Millipore).

Expansion in culture was carried out within T75, T150 or T225 vented flasks (Corning, #430641U), containing 10 mL, 15 mL or 20 mL media respectively. This media was removed and replaced every 2-3 days and at this time cultures were checked for contaminants using a brightfield microscope (Olympus). Cultures were further tested for microplasma infection using the services of Eurofins Genomics.

2.7 Cell passaging

Media was removed from cells and flasks were washed with 5 mL Dulbecco's phosphate buffered saline (DPBS, Sigma-Aldrich, #D8537) to remove any excess. Cells were then detached using 5 mL 1x porcine trypsin-EDTA (Sigma Aldrich, #T3924), incubated at 37 °C for 5 minutes. Detachment was verified using a brightfield CKX31 inverted microscope (Olympus). A further 5 mL media was added to inactivate the trypsin and the combined volume centrifuged (Eppendorf, #5810R) at 400 RCF for 5 minutes. The supernatant was then removed to leave the cell pellet. The pellet was then washed through the addition of 5 mL DPBS and vortexing (Thermo Fisher Scientific) followed by centrifugation at 400 RCF for

5 minutes to re-isolate the cell pellet. If the intention was to expand the cells, this pellet was then typically split 1:3 and placed into fresh flasks. For storage and lysis, see below.

2.8 Cell freezing

Following on from the steps above, the number of cells present was quantified using disposable hemocytometers (NanoEntek). 500,000 cells were suspended in 1 mL media containing 10% (v/v) dimethyl sulphoxide (DMSO, Sigma-Aldrich, #D2650) and the solution placed in a cryovial (Starlab Ltd., #E3370-6122). If only short term storage was required (<3 months) cryovials were placed at -80 °C (New Brunswick). If long term storage was required, the cells were placed in liquid nitrogen (BOC Ltd.) at -185 °C in a Cryolus Storage System (Thermo Fisher Scientific).

Cells were defrosted in a water bath (Grant Instruments) at 37 °C and seeded in the T75 flask with 10 mL media. This was replaced the following day in order to remove DMSO.

2.9 Cell lysis

Following on from “Cell passaging”.

2.9.1 SL-DOC method

The cell pellet was resuspended in 1 mL DPBS and the solution transferred to 1.5 mL LoBind tubes (Eppendorf, #022431081), followed by centrifugation at 400 RCF for 5 minutes. The DPBS was then removed and replaced with 30 µL SL-DOC lysis buffer, consisting of 25 mM ammonium bicarbonate (AB) (Sigma-Aldrich, #09830) in ultrapure water containing 1.1% (w/w) sodium dodecyl sulphate (SDS) (Sigma-Aldrich, #L3771), 0.3% (w/w) sodium deoxycholate (Sigma-Aldrich, #D6750), 0.1% (w/w) protease inhibitor cocktail (Sigma-Aldrich, #P8340) and 0.1% (w/w) phosphatase inhibitor cocktail (Sigma-Aldrich, #P0044). Six 1.6 mm diameter steel beads (Next Advance) were added to the cell pellet and buffer and the tubes placed in a Bullet Blender Tissue Homogeniser (Next Advance) for 2 minutes, maximum speed at 4 °C in order to complete the cell lysis. The lysate was then centrifuged at 10,000 RPM for 5 minutes and the supernatant extracted and placed in a fresh 1.5 mL LoBind Eppendorf.

2.9.2 Dounce homogenization method

The cell pellet was resuspended in 1 mL DPBS and the solution transferred to 1.5 mL LoBind tubes followed by centrifugation at 400 RCF for 5 minutes. The DPBS was then removed and replaced with 100 µL hypotonic buffer containing 10 mM Hepes (Sigma-Aldrich, #H4034), 10 mM sodium chloride (Thermo Fisher Scientific, #S/3160/65), 1.5 mM magnesium chloride (Sigma Aldrich, #M3634) in ultrapure water at pH 7.6. The cell pellet and buffer were then transferred to a Dounce Homogenizer (Wheaton, #257538) on ice and

treated to 20 strokes homogenization. The lysate was then passed twice through a 26 gauge syringe needle (BD Biosciences, #300300) to shear the DNA, followed by centrifugation at 14000 RCF for 5 minutes. The supernatant was then extracted and placed in a fresh 1.5 mL LoBind Eppendorf.

2.9.3 Triton X-100 method

The cell pellet was resuspended in 1 mL DPBS and the solution transferred to 1.5 mL LoBind tubes followed by centrifugation at 400 RCF for 5 minutes. The DPBS was then removed and replaced with 30 μ L Triton Buffer consisting of 1% (v/v) Triton X-100 (Sigma Aldrich, #93443), 0.1% protease inhibitors in DPBS. The cell pellet was incubated in this buffer at 25 °C for 1 minute followed by centrifugation at 20000 RCF for 5 minutes. The supernatant was then extracted and placed in a fresh 1.5 mL LoBind Eppendorf.

2.10 Protein quantification

Quantification of protein yield from cell lysates was achieved using a Bicinchoninic acid (BCA) assay kit (Thermo Fisher Scientific, #23227). This is a copper-based colorimetric assay based on the reduction of copper with protein in an alkaline medium, followed by reaction with BCA. The BCA/copper complex creates a purple colour that is absorbed at 562 nm and linearly correlates with protein concentration. The protocol was carried out according to the manufacturer's instructions and absorbance was measured using a plate reader (BioTek - #ELX800).

2.11 Gel electrophoresis

Lysates were split and, if necessary, diluted with ammonium bicarbonate so each sample contained the same amount of protein (values used ranged from 30-50 μ g, as determined by BCA assay) suspended in 40 μ L. Proteins were then fully denatured through the addition of 10 μ L loading buffer (50% glycerol (Thermo Fisher Scientific, G/0650/08), 10% SDS, 0.1% bromophenol blue (Sigma-Aldrich, #B5525), 500 mM dithiothreitol (DTT) (Roche, #10197777001) and 250 mM Tris-HCL pH 6.8 (Thermo Fisher Scientific, #10316893)), followed by boiling on a heat block at 95 °C for 5 minutes. Samples were then added to BioRad 4-20% acrylamide pre-cast gradient gels (BioRad, #456-1094), with 10 μ L Spectra Broad Range protein marker (Thermo Scientific, #26634) added to the first lane. The gel was then placed in a gasket and the tank filled halfway with SDS running buffer (250 mM Tris, 1.92 M Glycine (Thermo Fisher Scientific, G/0800/60), 1% SDS). This was plugged into a power bank and run with constant current, 35 mA for approximately 1 hour. After removal, protein bands were visualised through staining with Instant Blue (Expedean).

2.12 Heat shock of cell lysates

The cells used in this experiment were Y201s, lysed using the SL-DOC bead beating method. For each sample, the appropriate volume of lysate containing 50 µg protein was isolated and placed in a LoBind Eppendorf. For Condition 1 (Normal), three samples were immediately subject to limited proteolysis. The limited proteolysis step here consisted of adding 20 µL ammonium bicarbonate containing Proteinase K (Promega, #V3021) at an enzyme:substrate ratio (w:w) of 1:2000 to the lysate for a reaction time of 2 minutes, whilst on a heat block at 25 °C. This reaction was then quenched through heating the sample to 95 °C on a heat block for 5 minutes. As the control for Condition 1 (Normal), three samples were only subjected to the heating step - 20 µL ammonium bicarbonate was added to each sample, followed by heating at 95 °C for 5 minutes. For Condition 2 (Intermediate), six samples were placed on a heat block and incubated at 37 °C for 30 minutes, followed by cooling through placing on a heat block at 25 °C for 90 seconds. Three samples were then subject to limited proteolysis and three only to the heat step, exactly as described in Condition 1. For Condition 3 (Heated), six samples were placed on a heat block and incubated at 42 °C for 30 minutes, followed by cooling through placing on a heat block at 25 °C for 90 seconds. Three samples were then subject to limited proteolysis and three only to the heat step, exactly as described in Condition 1.

2.13 Heat shock of cells in culture

For the experiment discussed in Chapter 5, the cells used were bone marrow derived human mesenchymal stem cells, donor ID TH305, passage number 5. For the No Perturbation condition, flasks containing hMSCs were kept in the 37 °C incubator in which they had been cultured. For the Perturbation condition, flasks containing hMSCs were quickly transferred to a 42 °C, 5% CO₂ incubator where they were heated for 2 hours. All flasks were then removed from the incubators and cells were detached (Cell Passaging) and lysed using the Triton X-100 lysis method detailed above (Cell Lysis).

2.14 HSP70 inhibition

Media was removed from cells and replaced with 10 mL fresh media containing 0.01% DMSO and 10 µM 2-phenylethanesulfonamide (Sigma Aldrich, #P0122) [181]. Flasks were then incubated at 37 °C for 30 minutes, before proceeding to heat shock. In the control condition, media was also removed and replaced with 10 mL fresh media containing only 0.01% DMSO.

2.15 Mass spectrometry analysis

Mass spectrometry sample preparation began with ~50 µg protein in 25 µL buffer (see above).

2.15.1 Reduction and alkylation

For each sample, 25 μ L SDS buffer was added (10% SDS, 100 mM TEAB at pH 7.5) in order to reach the required SDS concentration for the later S-trap step to be effective. Enough DTT was then added to increase the total sample concentration to 5 mM, followed by incubation at 60 °C on a covered heat block for 10 minutes. DTT is a reducing agent that disrupts disulphide bonds between cysteine residues. Iodoacetamide (Sigma-Aldrich, #I1149) was then added to the samples to achieve a final concentration of 15 mM. The solution was then vortexed and incubated in darkness for 30 minutes at room temperature. Iodoacetamide is an alkylating agent and reacts with the cysteine residues, exposed by the DTT step, in a manner which prevents them from reforming disulphide bonds. Therefore, the reduction and alkylation of protein samples facilitates the subsequent digestion step through giving trypsin greater access to the protein. Iodoacetamide activity was quenched by adding the same amount of DTT as in the first step, before moving on to digestion of the samples.

2.15.2 S-trap protein digestion

Samples were acidified through addition of 12% phosphoric acid at a 1:10 dilution. Subsequently, proteins were made capable of binding to the S-trap columns (ProtiFi) through addition of 6x S-trap binding buffer (100 mM TEAB at pH 7.1, 90% Methanol (LC-MS grade)). This solution was loaded onto S-trap columns (atop a 2 mL tube) 150 μ L at a time, followed by centrifugation at 4000 RCF for 1 minute. It is expected that protein is now bound to the S-trap column, so flow through is discarded. The bound protein was then subject to 4x wash with S-trap binding buffer (add 150 μ L, centrifuge 4000 RCF for 1 minute, discard flow through). Proteins were then digested through the addition of 0.8 μ g/ μ L trypsin solution (proteomics grade trypsin, Promega) and incubation at 47 °C for 1 hour. S-trap columns were moved to a fresh collection tube and 65 μ L digestion buffer (50 mM TEAB, pH 8.5) added, followed by centrifugation at 4000 RCF for 2 minutes, in order to elute the peptides. Peptides were eluted further through addition of 65 μ L 0.1% formic acid in water, centrifugation at 4000 RCF for 2 minutes, followed by addition of 30 μ L 0.1% formic acid, 30% acetonitrile (ACN, Thermo Fisher Scientific, #10629112) in water and centrifugation at 4000 RCF for 2 minutes. Peptides should now be removed from the S-trap column.

2.15.3 Peptide desalting

100 μ L of OLIGO R3 beads (Thermo Fisher Scientific, #1-1339-03) was added to a 96-well filtration membrane plate (Corning, #3504). The beads were washed through resuspension in 100 μ L 0.1% formic acid, 30% ACN in LC-MS grade water and centrifugation at 200 RCF for 1 minute atop a collection plate. Run through to this collection plate was discarded. Beads were further washed with 2x addition of 0.1% formic acid in LC-MS grade water, centrifugation at 200 RCF for 1 minute and disposal of the run through. Next, the samples

(peptides) were added to each well of the membrane plate and incubated on a plate shaker at 800 RPM for 2 minutes followed by centrifugation at 200 RCF for 1 minute. Peptides should now be attached to the R3 beads. Peptides were then washed twice with 200 μ L 0.1% FA, plate shaker at 800 RPM for 2 minutes followed by centrifugation at 200 RCF for 1 minute. The collection plate was then changed for a fresh one and the peptides eluted from the beads through 2x addition 50 μ L 0.1% FA in 30% ACN to the membrane plate wells, plate shaker at 800 RPM for 2 minutes and centrifugation at 200 RCF for 1 minute. Eluted peptides were then transferred to MS sample vials (Thermo Fisher Scientific) and dried down in a vacuum centrifuge (Heto Cooling System).

2.15.4 Liquid chromatography mass spectrometry

Digested samples were analysed by LC-MS/MS using an UltiMate 3000 Rapid Separation LC (RSLC, Dionex Corporation) coupled to a QE HF (Thermo Fisher Scientific) mass spectrometer. Mobile phase A was 0.1% formic acid in water and mobile phase B was 0.1% formic acid in acetonitrile and the column used was a 75 mm x 250 μ m i.d. 1.7 mM CSH C18, analytical column (Waters).

A 1 μ L aliquot of the sample was transferred to a 5 μ L loop and loaded on to the column at a flow of 300 nl/min for 5 minutes at 5% B. The loop was then taken out of line and the flow was reduced from 300 nl/min to 200 nl/min in 0.5 minute. Peptides were separated using a gradient that went from 5% to 18% B in 63.5 minutes, then from 18% to 27% B in 8 minutes and finally from 27% B to 60% B in 1 minute. The column is washed at 60% B for 3 minutes before re-equilibration to 5% B in 1 minute. At 85 minutes the flow is increased to 300 nl/min until the end of the run at 90 minutes.

Mass spectrometry data was acquired in a data directed manner for 90 minutes in positive mode. Peptides were selected for fragmentation automatically by data dependant analysis on a basis of the top 12 peptides with m/z between 300 to 1750 Th and a charge state of 2, 3 or 4 with a dynamic exclusion set at 15 sec. The MS Resolution was set at 120,000 with an AGC target of 3e6 and a maximum fill time set at 20 ms. The MS2 Resolution was set to 30,000, with an AGC target of 2e5, a maximum fill time of 45 ms, isolation window of 1.3 Th and a collision energy of 28.

2.16 Progenesis and Mascot data processing

MS output files (.RAW) were imported into Progenesis QI (Nonlinear Dynamics, version 4.1). Ion intensity maps were first visually inspected to identify any potential issues, according to Nonlinear Dynamics guidelines. The files were then aligned using the software's automatic alignment function, using an alignment score of 70% as a minimum acceptance threshold. Progenesis QI was also used for peptide filtering and quantification. Relative quantification

using non-conflicting peptides was chosen and the results filtered to exclude peptides with less than 2 isotopes, charge greater than 4 or rank greater than 3.

The results were then exported to Mascot Daemon (Matrix Science UK) for identification against the SWISS-Prot and TREMBL human databases. These were filtered to exclude unannotated protein sequences (The UniProt Consortium, 2019) and include phosphorylation of serine, tyrosine, threonine, histidine or aspartate, oxidization of methionine, hydroxylation of asparagine, aspartic acid, proline or lysine and alkylation of cysteine. The digestion specificity parameter was set to either 'trypsin' or 'semi-trypsin' depending on the search. In both instances, up to two missed cleavages were permitted. Automatic decoy database search was selected and the default PSM FDR of 0.05 was used. The resultant Mascot files were then imported into Progenesis QI before the final peptide and protein identifications were exported.

2.17 MaxQuant data processing

MS output files (.RAW) were imported into MaxQuant (version 1.6.7.0). Samples were separated by assigning them to different fractions, for example, replicates of LiP treated samples were assigned to one fraction, whilst the replicates from the control samples were assigned to another. Given the difference between LiP and control samples, this step was taken to reduce peptide misidentifications in the match between runs step. In the group specific parameters section, the search parameters used were the default settings – fixed modifications were carbamidomethyl cysteine, variable modifications were the oxidation of methionine and the acetylation of protein N-terminus and the maximum number of modifications per protein was set to five. The digestion parameters were set to Trypsin, with a maximum of 2 missed cleavages allowed. Digestion mode was set to “Specific” or “Semi-specific”, depending on the type of peptides requiring consideration.

In the global parameters section, peptides were identified against the Uniprot human proteome (uniprot-proteome_UP000005640.fasta) and MaxQuant's default contaminants. The default minimum peptide length was 7 and maximum peptide mass 4600 Da. Unique peptides only were used for protein quantification. MaxQuant's reversed decoy database was used and default identification settings. These are PSM FDR 0.01, protein FDR 0.01 and minimum 1 peptide per protein. Match between runs was selected.

2.18 Protti data processing

In Chapter 5, the LiP-MS data was processed using a new R package “Protti” (version 0.3.1) [171]. Peptides and proteins were normalized using the inbuilt “normalize()” function with method = “median”. This function outputs a normalized intensity for each feature, calculated

through taking away the median intensity for that sample from the original intensity and adding the global median intensity for all samples. After normalization, the data is filtered to only include proteins with a coefficient of variance of less than 25% in at least one experimental condition. This was achieved using the protti function “filter_cv()”. For example, this filtration step for the MaxQuant protein groups dataset filtered out 1215 groups out of a total 3272, so 63% of the data remained. Next, missingness was determined using the “assign_missingness()” function. Here features were labelled missing at random (MAR) if at least 70% of total replicates were present in the conditions being compared, and missing not at random (MNAR) if the feature was present in 100% of replicates for one condition and less than 20% in the second condition. Protti then provides the option to impute MAR values using a variety of methods. However, for the data shown in Chapter 5, values were not imputed. Differentially abundant peptides/proteins were then identified using the “calculate_diff_abundance()” function. The parameters chosen were a Welch t-test and the Benjamini Hochberg multiple testing correction. Peptides/proteins were considered significantly changing if the reported adjusted p-value was below 0.05.

The gene ontology enrichment analysis described in Chapter 5 was also performed using Protti, with the function “calculate_go_enrichment()”. This function compares the gene ontology terms associated with proteins tagged as significant with gene ontology terms associated with all detected proteins. Significant enrichment or depletion of terms is determined using a two-sided Fisher's exact test.

Chapter Three: Optimization of the LiP-MS protocol

A comprehensive understanding of the proteostasis network underpins a complete understanding of human health, in particular due to the increasing prevalence of neurodegenerative diseases and the progressive loss of proteostasis with ageing. The maintenance of proteostasis is governed by dynamic changes to protein abundance, protein localization and protein structure. To gain a systems-level understanding of proteostasis and its breakdown, each of these features must be characterized at the omics scale. Mass spectrometry has been used for many years to assess alterations to protein abundance, with increasing levels of accuracy. Advances in sample preparation methods and ultra-high sensitivity mass spectrometry now even enable proteome analysis at the single cell level [112, 113]. More recently, the field of spatial proteomics has made many leaps forward. Organelle fractionation followed by MS is capable of analyzing the subcellular location of the proteome and identifying dynamic changes to protein localization [34, 182]. Finally, structural proteomics techniques have also seen significant advancement in recent years. The aim of this thesis was to develop one such technique – LiP-MS – for application to primary human cells.

Limited proteolysis (LiP) refers to using an enzyme to probe the structure of a protein. The chosen enzyme should be **non-specific** and applied to the **natively folded protein** for a **reduced digestion time**. If these conditions are optimized, the enzymatic cleavage will only occur at surface accessible, flexible regions of the protein. In this way, LiP cleavage sites are dictated by the structure of a protein and can be compared across different biological conditions in order to identify protein conformational changes. Crucially, many functionally critical alterations to protein structure are minor, meaning the LiP-MS protocol must be carefully tuned to achieve the level of spatial resolution required to maximize biological insight. This chapter details the optimization of LiP-MS for use in hMSCs, with a discussion of the refinement of the LiP enzyme concentration, LiP digestion time and the method of cell lysis.

Optimising enzyme concentration is important for maximising the generation of identifiable peptides. Too high a concentration or too long a reaction time would result in excessive digestion of the proteome, leading to loss of information as cleavage sites would no longer be dictated by protein structure. Further, the identification of LiP cleavage sites is reliant upon such cleavages being limited to one per fully tryptic peptide. Multiple cleavages will

result in non-tryptic peptides, which can be identified but at hugely increased computational expense, and significantly reduced confidence. Multiple cleavages per tryptic peptide are also likely to produce much shorter peptides, which make MS analysis more difficult. Likewise, if proteolysis is too limited, either due to low enzyme concentration or reduced reaction time, even the most accessible, flexible regions of protein structure may not be cleaved sufficiently to produce a detectable concentration of peptides.

Importantly, the 3D conformations present across the human proteome are extremely diverse and so there is no “one size fits all” enzyme concentration. For some unstable proteins even the most limited proteolysis may result in complete digestion, whereas certain globular or aggregated structures may prove impervious to any cleavage. When identifying structural alterations over different biological conditions, different types of structural change may require different levels of LiP. For example, substantial protein unfolding may be easy to identify with a low enzyme concentration, but subtler changes such as accessibility of binding sites may be obscured. For these reasons, great care must be taken in tuning the conditions of LiP to either maximise the output of protein structural information or to prioritise the type of structural change being investigated.

Another crucial aspect of the LiP-MS protocol in need of optimization for this model system is the method of cell lysis. The manner of lysis must be carefully controlled in order to ensure proteins retain as close to native structure as possible. This entails mitigating against any potential denaturants such as temperature, pH or harsh detergents whilst still maintaining the level of protein solubility required for MS analysis.

3.1 Optimisation of proteolysis conditions

The first aim of this project was to investigate which enzyme substrate ratio (E:S) and reaction time are most suited to the LiP of hMSC lysates, as well as confirming that LiP is indeed possible in such a system. The non-specific enzyme chosen was proteinase K (PK).

Preliminary analysis was carried out using gel electrophoresis. SDS-PAGE is based on the principle that smaller proteins will travel faster down the gel front, and it can therefore act as a crude measure of enzymatic activity, as more cleavage sites will result in smaller protein fragments. Consequently, LiP lysates will on average travel further down the gel than a control sample of intact proteins. If proteolysis is too limited, the LiP and control samples will be indistinguishable. Conversely, excessive proteolysis may result in molecular weights low enough to diffuse out of the gel. This system was used to test a wide variety of E:S ratios and reaction times in order to identify the conditions associated with these extremes, and then to hone in on more suitable settings.

An example of such gels is provided in Figure 3.1. The LiP treated hMSC lysates (Fig 3.1(c)-(f)) appear more fragmented than the control (no LiP) sample (Fig 3.1(b)), indicating that some level of proteolysis has indeed occurred. Further, the highest enzyme concentration (Fig 3.1(f)) displays the most fragmentation, as expected. However, SDS page can only provide a crude idea of the effect of limited proteolysis on the proteome.

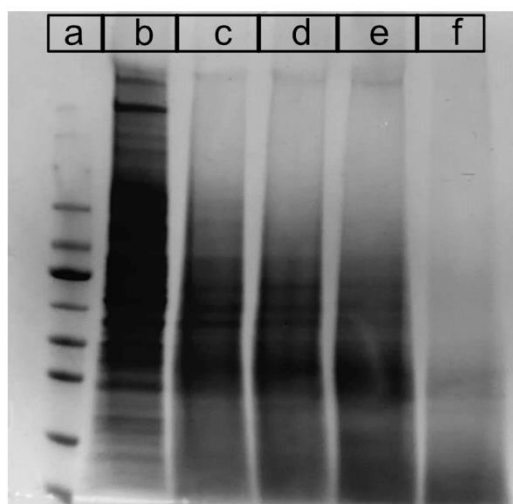


Figure 3.1: SDS page of different levels of LiP. (a) Spectra Broad Range Protein Ladder, (b) hMSC lysate, no LiP enzyme. (c) hMSC lysate, incubated with PK at a ratio of 1:2000 E:S (w:w) for 2 minutes, (d) the same as (c), but with E:S of 1:1000, (e) E:S 1:500, (f) E:S 1:100.

3.1.1 Trial LiP-MS experiment

To learn more, a shortlist of candidate E:S ratios (w:w) 1:100, 1:500, 1:1000 and 1:2000 were then investigated by MS. The four LiP conditions were tested on hMSC lysates, with a reaction time of 2 minutes and each were subsequently digested to completion with trypsin and the peptides identified by LC-MS/MS. The motivation for this experiment was twofold. Firstly, it enabled the identification of the optimum E:S ratio for LiP by utilizing the peptide level information provided by MS and secondly, to assess the clarity of whole cell lysate LiP-MS data. For example, on the omics scale one can determine whether the relative number of semi-tryptic peptides identified meets expectation, whereas evaluation of specific LiP peptides can elucidate protein level effects. For example, do LiP peptides indeed map to flexible, surface accessible regions of protein structure?

The level of proteolysis achieved was assessed through the quantification of LiP peptides, identified using the following criteria:

1. Decrease in fully tryptic peptides – the number of fully tryptic peptides present in each sample should negatively correlate with enzyme concentration. Increased

enzymatic activity should result in more cleaved tryptic peptides and hence a decreased presence of fully tryptic peptides (Figure 3.2).

2. Increased semi-tryptic peptides – the number of semi-tryptic peptides present in each sample should positively correlate with enzyme concentration. However, this may only hold true up to a certain E:S ratio, where protease concentration exceeds a certain tipping point where sufficient cleavage of the substrate protein occurs that precipitates complete unfolding and/or denaturation, leading to complete elimination of some detectable semi-tryptic peptides. Hence, semi-tryptic identification may also decrease (Figure 3.2). This would be expected to vary on a protein by protein basis.

In the local sense, once LiP peptides have been identified, these can then be mapped onto known protein structures. This information can then be used to investigate the characteristics of LiP cleavage sites.

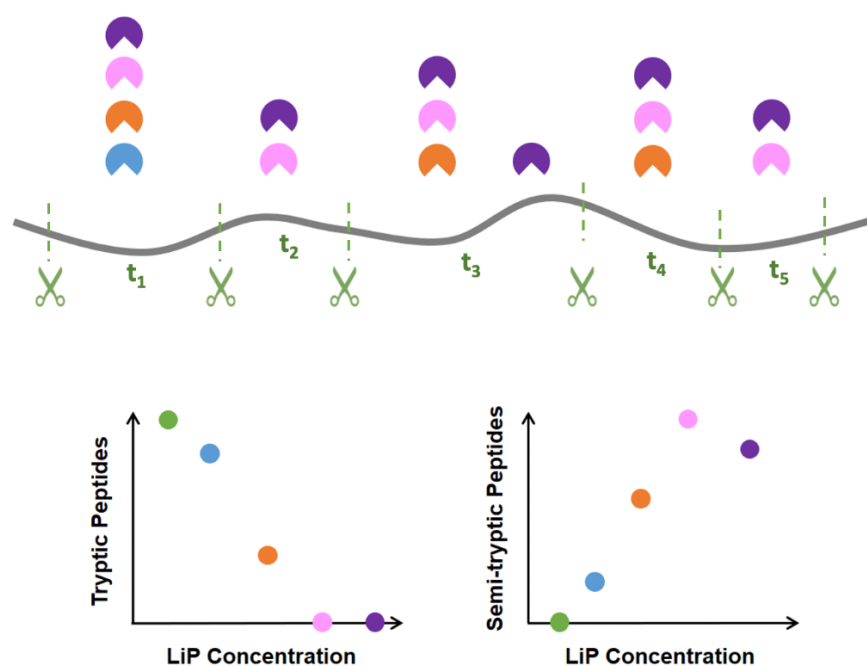


Figure 3.2: The effect of PK (LiP) concentration on the type and number of peptides identified. Theoretically, trypsin (visualised here by green scissors) cleaves protein (grey line) only at the C-terminus of amino acids arginine and lysine, in a reliable manner. Therefore, the peptides produced will all be tryptic and not semi-tryptic (green dot). The inclusion of PK (pac-man symbol) to the digestion protocol will result in additional cleavages and so reduce the abundance and number of tryptic peptides (e.g. t_1, t_2, t_3 , etc), with a corresponding increase in semi-tryptic peptides. The size of this effect is dependent on the PK concentration used. Increasing PK concentration (blue, orange, pink, purple represent lowest to highest concentration respectively) is expected to increase the number of additional cleavages, and so negatively correlate with tryptic peptide identifications, and positively correlate with semi-tryptic peptide identifications. However, after a certain threshold, PK may cleave multiple times per tryptic peptide, resulting in non-tryptic peptides, as shown by the purple marker. This may reduce the number of semi-tryptic peptides that are identifiable, as they are more likely to be shorter in length, thus making MS analysis more difficult and effectively removing the parent protein from the pool of detectable species. However, it is important to note that the extent to which the scenario depicted here is true is principally dependent on the structure of the protein in question.

3.1.2 Tryptic peptides

As a trial run and the first experiment conducted in this work, for each enzyme condition only one sample was submitted for MS analysis. Unfortunately, there were some issues with the control condition (no LiP enzyme). For this reason, when assessing tryptic peptides, this sample was excluded from the analysis and instead the data was normalised to the most comparable enzyme condition – the 1:2000 E:S ratio.

To begin, the number of tryptic peptides identified in each condition was quantified. The results, displayed in Figure 3.3(a), show a generally negative correlation between enzyme concentration and the number of tryptic peptides identified, which decreases with enzyme concentration. This is in line with expectation, as increased PK activity should result in additional non-tryptic cleavages and so a loss in tryptic peptides. This was further investigated in terms of protein coverage, normalised to that obtained at the lowest enzyme concentration. Analysis of protein coverage can also provide a readout of LiP activity as the tryptic peptide signal associated with the protein may be diminished in the LiP treated sample compared to the control. For each protein present in the 1:2000 sample, the percentage of that protein's sequence identified in the MS data (via tryptic peptides) was calculated. The top 20 proteins were selected (with the highest sequence coverage at 1:2000) and for each the corresponding coverage was calculated for the other samples. This was then normalized to the 1:2000 value, representing the baseline where the least signal loss due to additional PK cleavage should be present. This is plotted in Figure 3.3(b) and again demonstrates a general decrease in sequence coverage with increasing enzyme concentration, though with little difference between the 1:2000 and 1:1000 samples.

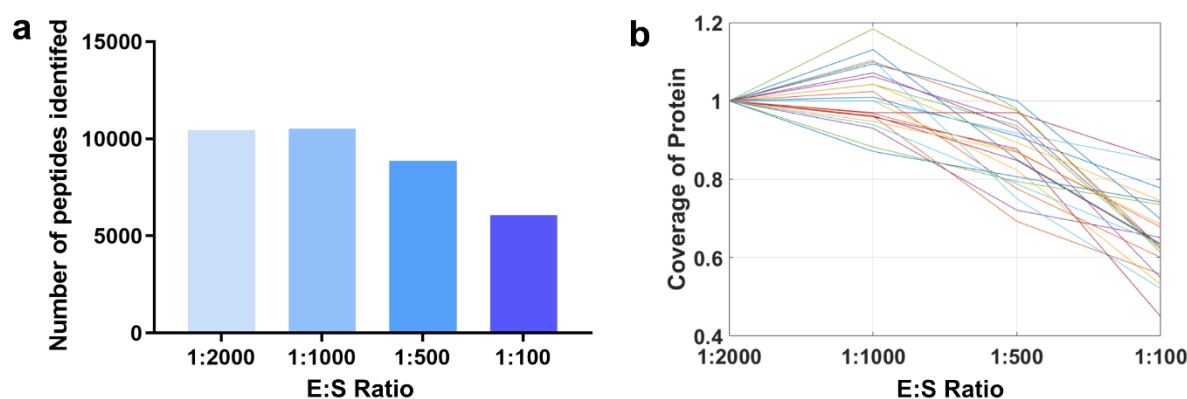


Figure 3.3: (a) The number of tryptic peptides identified in each experimental condition was quantified. This demonstrates a generally negative correlation between tryptic peptides and enzyme concentration. (b) The percentage of each protein's sequence identified in the MS data was quantified for each enzyme condition and this value was normalised to the 1:2000 value. Plotted here are the results for the top 20 proteins with the highest sequence coverage (as determined by the 1:2000 sample). This demonstrates a reduction in sequence coverage for E:S ratios 1:500 and 1:100.

Next, the influence of enzyme concentration on the MS signal of identified tryptic peptides was investigated. Again, peptide signal intensities were normalised to that of the 1:2000 PK digestion, generating normalised profiles across the enzyme concentrations investigated, which were subsequently clustered using K means clustering. This is a technique that enables the identification and quantification of patterns of behaviour in large data sets, here being how the reported intensity of a peptide changes with enzyme concentration. Briefly, the K means algorithm initially selects N random data points (“centres”) and then assigns the rest of the dataset to its closest centre. These initial centres are then altered iteratively until the distance between the data and cluster centre is minimised. The number of clusters N was determined using the Elbow method, a technique to minimise variance within clusters whilst also minimising the number of clusters.

The results of such clustering are shown in Figure 3.4. The predominant behaviour extracted by this analysis is that of little change in normalised peptide abundance over conditions, as shown in Clusters 1 and 5 (Fig 3.4(a)) which together represent 79% of the dataset (Fig 3.4(b)). These peptides likely map to regions of protein that are always either accessible or inaccessible to the enzyme in this range of concentrations, and whose susceptibility to proteolytic attack remains stable. It is intuitive that this behaviour is the most common, as the most native protein structure in unperturbed cells is expected to be globular and although the enzyme concentration varies the solution conditions do not [183]. Hydrophobic peptides mapping to the core of such proteins are expected to be completely shielded from enzymatic access, whereas, on the protein’s surface, linker regions between secondary structures, or indeed between domains, may be cleaved more easily even at low enzyme concentrations such as 1:2000. If such regions are cleavable at 1:2000, they will also be cleaved at the higher enzyme concentrations. Furthermore, other types of protein structure follow the same rules. For examples, intrinsically disordered protein domains or regions are highly enzyme accessible and thus prone to cleavage even at low enzyme concentrations. Conversely, tightly packed globular structures may prove impervious to all the concentrations tested.

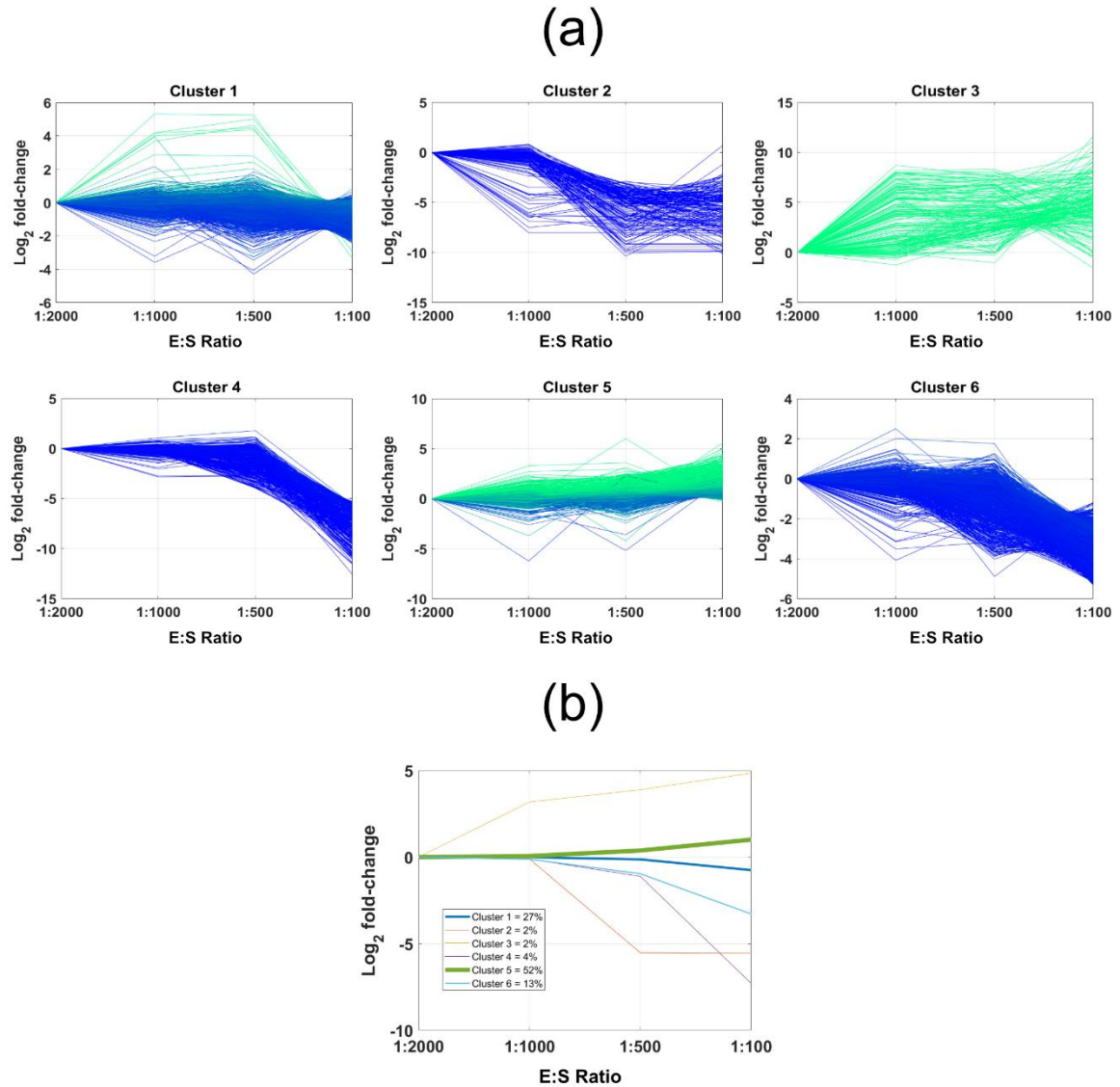


Figure 3.4: (a) *K* means clustering of tryptic peptide abundance data in 6 clusters, as determined by the Elbow method. Each line represents a single peptide. The y axis displays the log₂ fold change of this peptides abundance relative to the 1:2000 sample. The colours are indicative of the ranked mean change, with the highest mean (most positive) in green to lowest (most negative) in blue. (b) Each line depicts the mean of each of the 6 clusters. The thickness of this line is determined by the percentage of the population represented by that cluster. E.g, 27% of tryptic peptides identified fall into Cluster 1.

The second most common behaviour is a negative correlation between the abundance of tryptic peptides and enzyme concentration, as is displayed in Clusters 2, 4 and 6, altogether representing 19% of peptides present. This is the central dogma of LiP as a structural probe – regions of LiP cleavage will be identifiable by a loss in tryptic peptide signal, and the extent of this loss is expected to correlate with LiP enzyme concentration. Therefore, the peptides represented in these clusters to are predicted to constitute accessible protein regions and hence sites of LiP.

Contrary to this, Cluster 3 contains peptides with the opposite behaviour, but is poorly represented, accounting for only 2% of the data. This can be considered to result from the

underlying noise in the MS data contributing to apparent signal 'gain'. This might occur when peptides are inherently low abundance and close to limits of detection, and noise contributes to an apparent gain in signal, or potentially they are misidentified in one run or another. On the other hand, this tryptic signal gain could be indicative of protein insolubility. Particularly proteolytic resistant proteins may require aggressive or additional cleavage steps, as provided by the LiP enzyme, in order to enable subsequent trypsin access which means significant signal is only recorded in the higher enzyme concentrations. This is an idea that is explored further in Section 5.

3.1.3 Nickpred

The potential LiP cleavage sites identified here can be investigated further using Nickpred – a program designed to predict the most likely sites of LiP cleavage in proteins with known or predicted structure [170]. Nickpred is able to calculate many of the known determinants of limited proteolytic susceptibility, as discussed in Section 1.6.2, and combines these measurements into a set of prediction scores.

The parameters included in the model are:

- 1) Solvent Accessibility – Solvent accessible protein regions have been shown to correlate with sites of LiP, as they are readily bound and cleaved by an endoproteinase and do not require significant remodeling in order to host the proteolytic enzyme [159]. The solvent accessibility of each protein residue is determined using a 1.4Å probe and summing the atomic accessible regions, and can be expressed as a relative accessibility with respect to a Ala-X-Ala peptide in extended conformation [184].
- 2) Protrusion Index – For globular proteins, LiP is considered to be least likely in regions mapping to the protein core and most likely at regions protruding from the central protein body. The protrusion index of each protein residue is determined using binning of protein atoms into successive ellipsoidal shells, starting with an equimomental ellipsoid about the proteins centre of mass [185].
- 3) Temperature Factors – The requirement for flexibility at LiP cleavage sites has been demonstrated by many authors [154, 155, 157]. Although there are no common and universal empirical measures of this property, flexibility can be estimated through assessment of B values – crystallographic temperature factors indicative of the dynamic properties of the polypeptide chain, through describing the mean squared displacement of each atom. A flexibility score is assigned to each protein residue by summing and normalizing atomic B-values if they are available.

- 4) Secondary Structure - The capacity for local unfolding has also proved a significant requirement for LiP cleavage sites, as the substrate should be able to adapt to the LiP enzyme binding site without the need for costly conformational change [156-158]. In the context of secondary structure, α -helices were found to be capable of such local unfolding but β -strands were not, at least not without substantial unfolding of the protein domain in question. Protein residues are scored based on whether they form part of a turn, helix or strand, and this score is then summed and normalised [186].
- 5) Hydrogen Bonding – The capacity to locally unfold the substrate to adapt to the protease is also influenced by the number of local intramolecular interactions that would otherwise need to be broken in the substrate. In this sense, protein regions with many hydrogen bonds are unlikely LiP cleavage sites. A score is assigned to each protein residue based on the number of hydrogen bonds formed between a short window centred on the current residue and all other regions of the structure. Regions which are heavily hydrogen bonded to the rest of the structure should be harder to locally unfold and cleave.
- 6) Ooi Numbers – Due to the requirement for flexibility and local unfolding, LiP cleavage sites less frequently map to crowded protein regions. This can be reflected in Ooi number scores, which are assigned to each protein residue through calculating the number of α -carbon centres within a certain radius [187]. This parameter is a simple measure of the local packing density, with the logic that tightly packed regions are similarly difficult to deform for proteolytic cleavage.

These 6 parameters are each weighted and smoothing windows applied to take into account the influence of neighbouring amino acids. In this way, the parameters are combined to produce a score reflecting the predicted susceptibility of each residue of the protein under consideration to be cleaved by a given protease, given the protein is in a native or near-native state.

For each protein identified in this trial LiP-MS experiment, the corresponding AlphaFold prediction PDB file [188, 189] was downloaded and run through Nickpred. The scores for each protein residue were then averaged to give a single readout for each tryptic peptide. Nickpred scores for the entire tryptic dataset were then compared to those assigned to the clusters in which LiP cleavage is mostly likely to be occurring (Figure 3.4, Clusters 2,4,6). This comparison is shown in Figure 3.5. Whilst there was no significant difference between all peptides and Clusters 2 and 6, Cluster 4 was found to have significantly higher prediction scores. Therefore, in this instance, the LiP sites identified in the MS dataset correlate with the known structural determinants of limited proteolysis.

However, there are caveats to this approach which must be appreciated. Firstly, the PDB files used were based on AlphaFold protein structure predictions. Whilst the accuracy of many of AlphaFold's structures is outstanding, due to the high number of protein searches a model confidence cutoff was not able to be imposed. Therefore, low confidence protein regions were not excluded from the analysis. Also, AlphaFold structures do not necessarily properly account for the cognate higher order structure of the protein, for example if the protein is normally part of a complex or multimer [190], and so the solvent accessibility of some protein regions may be overstated. Indeed, intra-domain and flexible linkers regions are potentially the least confidently predicted, but also likely to be regions of high susceptibility to attacking proteases [156]. Furthermore, the capacity of Nickpred is reduced for proteins with high levels of disorder, which may be precisely where one expects to find some of the identified LiP cleavage sites. Finally, analyzing tryptic peptides does not provide the precise location of LiP cleavage, as PK could theoretically have cleaved anywhere along the length of the tryptic peptide. Therefore, averaging Nickpred scores over peptides may potentially dilute the prediction power, particularly over longer peptides.

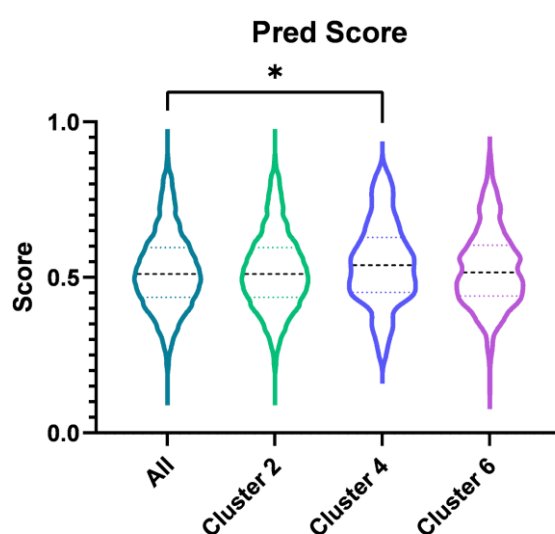


Figure 3.5: Box and whisker plots showing the distribution of Nickpred scores for each peptide identified in the trial LiP-MS experiment. Here, 1 is the highest likelihood of LiP cleavage, and 0 the lowest. Cluster 4 demonstrated a significantly higher prediction score set ($p < 0.05$) compared to that of the remaining LiP-MS dataset (ALL), as determined using a Mann-Whitney test.

However, overall these results are encouraging and largely align with expectation regarding the tryptic peptide signature of limited proteolysis. However, a suitable E:S ratio for hMSCs has not yet been determined. To do so, semi-tryptic peptides were investigated.

3.1.4 Semi-tryptic peptides

Semi-tryptic peptides, where only one end is generated by tryptic cleavage, were identified in the MS data by changing the theoretical digest conditions in the search engine, Mascot, to

semi-specific. This means peptides will be included in the search output if either the N terminal or C terminus is tryptic. Whilst this increases computational time significantly, it is not unfeasible.

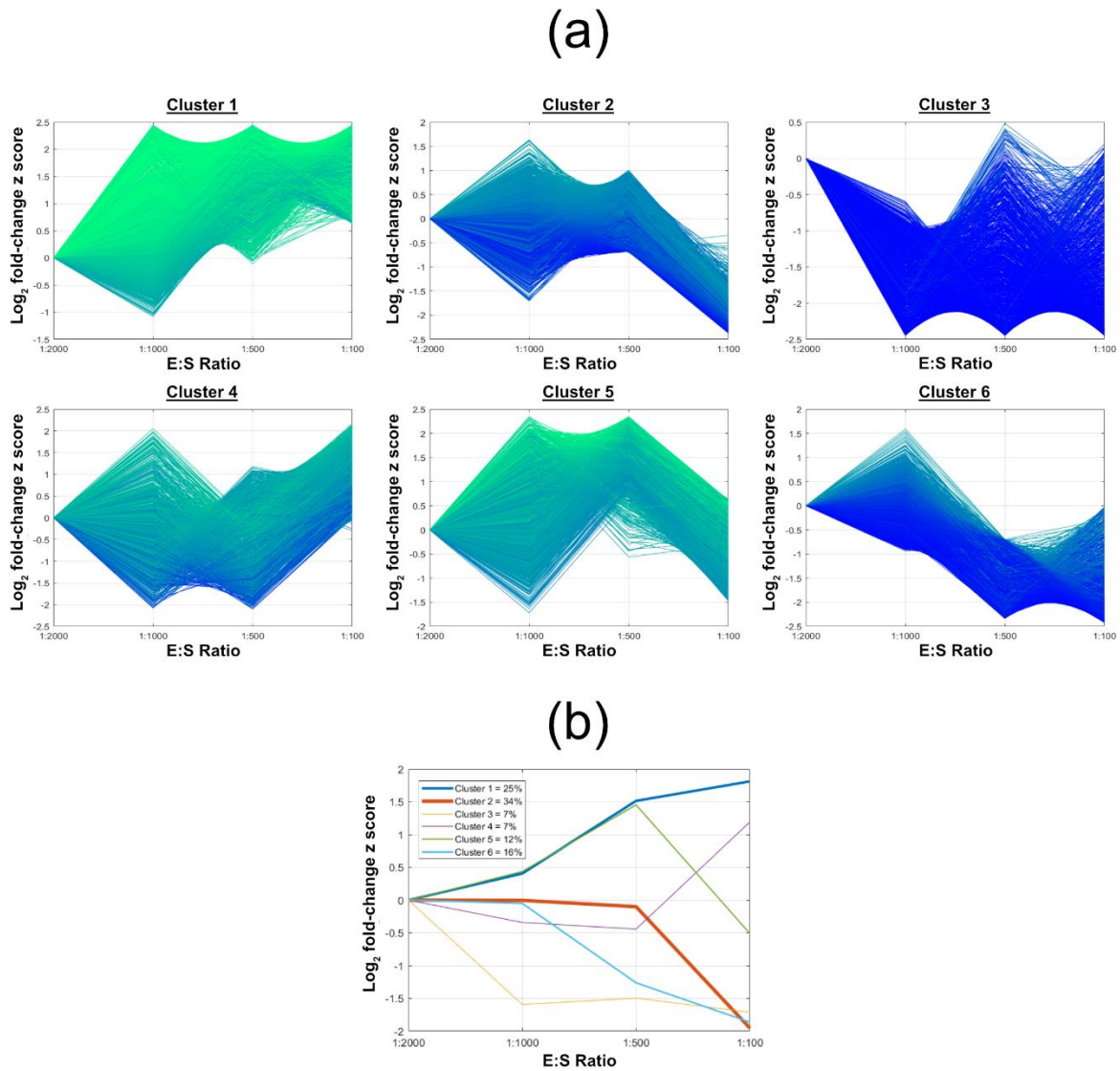


Figure 3.6: (a) K means clustering of semi-tryptic peptide abundance data in 6 clusters, as determined by the Elbow method. Each line represents a single peptide. The y axis displays the log2 fold change of this peptides abundance relative to the 1:2000 sample, converted into a z-score which scales the standard deviation to 1. The colours are indicative of the ranked mean change, with the highest mean (most positive) in green to lowest (most negative) in blue. (b) Each line depicts the mean of each of the 6 clusters. The thickness of this line is determined by the percentage of the population represented by that cluster. E.g, 25% of semi-tryptic peptides identified fall into Cluster 1.

As in the fully tryptic case, K means clustering analysis was used to assess the abundance of semi-tryptic peptides in each condition and identify patterns of behaviour (Figure 3.6). Here, the log2 fold change of semi-tryptic peptide abundance relative to the 1:2000 sample was converted into a z-score for each peptide, scaling the standard deviation to one. This

preserves the dynamics of peptide signal across the different enzyme conditions, but reduces the amplitude, thus enabling similar behaviours to be better clustered together.

In line with expectation, one of the most prevalent behaviours identified was a positive correlation between enzyme concentration and semi-tryptic peptide abundance, as shown by Cluster 1, which represents 25% of the whole semi-tryptic peptide dataset. Cluster 5 represents a further 12% of the dataset and displays a similar positive correlation between enzyme concentration and semi-tryptic peptide abundance until an E:S ratio of 1:500, after which there is a marked decrease. This is potentially indicative of the presence of multiple cleavages of single tryptic peptides. In this scenario, enzymatic activity is still increasing with concentration, but the number of additional cleavages per single tryptic peptide reaches a tipping point after 1:500, after which peptide fragments become non-tryptic and potentially too small for identification by MS. This idea is demonstrated by the schematic in Figure 3.2.

Another common feature of the clusters is a lack of change in semi-tryptic peptide abundance with enzyme concentration until an E:S ratio of 1:500, after which there is again a marked decrease. This is shown by Cluster 2 and represents 34% of the dataset. In a similar manner to the tryptic clustering (Figure 3.4) these peptides may map to protein regions that are easily enzyme-accessible, and so consistently cleaved at most enzyme concentrations. Once again, an E:S ratio of 1:100 may represent a tipping point, at which excessive cleavage results in the loss of identifiable semi-tryptic peptides. Therefore, this is similar to the behaviour of Cluster 5, but in a more stepwise manner in which 1:2000-1:500 produce the same result, and this is altered at 1:100.

Cluster 4 also shows little difference in semi-tryptic peptide abundance until an E:S ratio of 1:500. Conversely, this is followed by a marked increase at 1:100. The opposite of the previous, these peptides may map to protein regions that are generally inaccessible to the enzyme, and so substantial cleavage is only possible at the highest enzyme concentrations. Again, this appears to be in a stepwise rather than continuous manner. Further, this behaviour is less representative, at just 7% of the dataset.

The primary aim of this analysis was to determine which E:S ratio is most suited to LiP-MS studies of hMSC lysates. Overall, the K-means clustering of the semi-tryptic dataset revealed little difference between the 1:2000 and 1:1000 sample, with the exception of Cluster 3, which seems to suggest that for a small number of peptides (7%) 1:2000 may generate the most semi-tryptic peptides. Furthermore, in 78% of the dataset, an E:S of 1:500 either increased semi-tryptic peptide abundance, or demonstrated little change. Perhaps most importantly, the clustering also provided evidence that, for a number of proteins, an E:S

of 1:100 may produce excessive cleavages, resulting in a loss of identifiable semi-tryptic peptides and so a reduction in the level of structural data attainable by LiP-MS.

Taken together with the results from the tryptic peptide analyses, this suggests that in order to achieve appropriately limited proteolysis, the maximum recommended E:S ratio for hMSCs is 1:500. Naturally, whatever enzyme concentration is selected will be a form of compromise, since some proteins will be almost entirely digested to undetectable peptides, whilst other more stable ones will be entirely resistant. However, at this concentration of PK there is evidence of both a significant decrease in tryptic peptide abundance and a concordant increase in semi-tryptic abundance.

It can be additionally noted that tryptic peptide behaviour over LiP conditions was more clearly defined, here shown through less variance in the clustering. The semi-tryptic peptide data clustered poorly, but the effect of enzyme concentration effect was extractable upon conversion to a z-score. The variability and caveats of semi-tryptic peptides are investigated further in subsequent chapters.

Nickpred of semi-tryptic peptides

As in the tryptic case, the semi-tryptic dataset was run through Nickpred, to provide insight into the structural characteristics of the cleavage sites. While many of the caveats of this method (as discussed above) remain, when using semi-tryptic peptides one can identify the exact cleavage sites and so here the Nickpred scores were not averaged along the peptide, but instead corresponded precisely to the residue at the non-tryptic terminus of the peptide. Figure 3.7 displays such Nickpred scores for the dataset, where protein structural information was available. The scores are grouped according to cluster membership, as determined by the K means clustering of Figure 3.6.

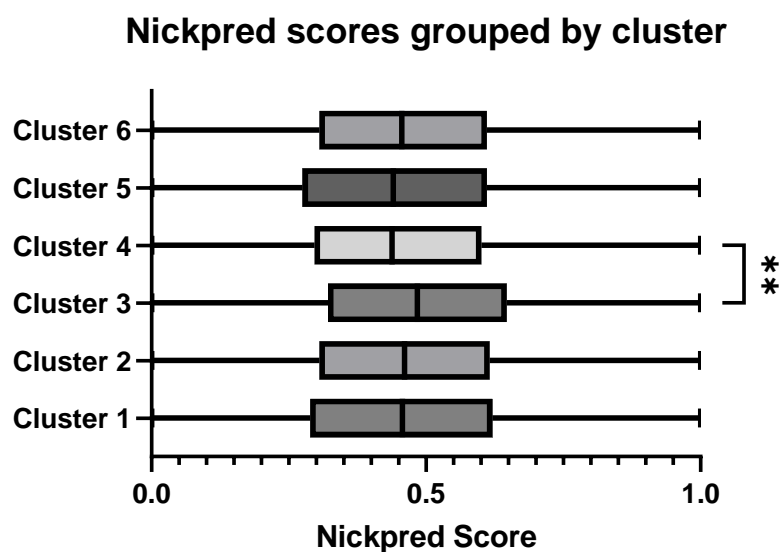


Figure 3.7: Box and whisker plots showing the distribution of Nickpred weighted norm scores for each cleavage site identified by the semi-tryptic peptides in the trial LiP-MS experiment. Here, a score of 1 represents the highest likelihood of LiP cleavage, and 0 the lowest. Cluster 3 demonstrated a significantly higher weighted norm score set (** = $p < 0.01$) compared to Cluster 4, as determined using a Mann-Whitney test.

The biggest difference was between Clusters 3 and 4, with Cluster 3 found to have significantly higher Nickpred scores than Cluster 4.

Cluster 4 was characterised by a consistent semi-tryptic peptide abundance across samples, followed by an increase at E:S 1:100. It was hypothesized that this peptide footprint could be characteristic of less-accessible protein regions, that require higher enzyme concentrations for cleavage to occur. Conversely, Cluster 3 was characterised by a decrease in semi-tryptic peptide abundance following E:S 1:2000. Here, the lowest enzyme concentration produced maximum cleavage and so potentially represent protein regions that are highly enzyme-accessible.

Nickpred utilises several structural indicators (Section 3.1.2) to determine the LiP-susceptibility of each residue for a given protein. Therefore, the significant increase in Nickpred scores between Clusters 4 and 3 indicates that Cluster 3 represents more accessible protein regions, as predicted by the semi-tryptic peptide footprint. Furthermore, Cluster 3 has the highest median Nickpred score of the dataset, which is again supportive of the hypothesis that this cluster captures particularly accessible peptides. Thus, this provides some evidence that semi-tryptic peptide behaviour is linked to protein structural features.

3.1.5 Peptide mapping

Once LiP peptides have been identified, these can then be mapped onto known protein structures. This allows investigation the LiP peptide fingerprint in more detail, illustrated with an example case here.

Calumenin (UniProt ID: O43852) is a calcium binding protein, thought to be localised to the endoplasmic reticulum (ER) [191]. Figure 3.8 shows the intensity of each peptide identified in calumenin, displayed along the protein's sequence from N- to C-terminus. For each MS run (1:2000, 1:1000, 1:500, 1:100) the total peptide signal, from both tryptic and semi-tryptic peptides, was calculated across the whole protein. Each peptide was then normalised to this total, thus becoming a percentage of how much this peptide contributed to the overall detected protein signal in that run. The contribution of tryptic and semi-tryptic peptides to total protein signal, and how this changes with enzyme concentration, can then be assessed.

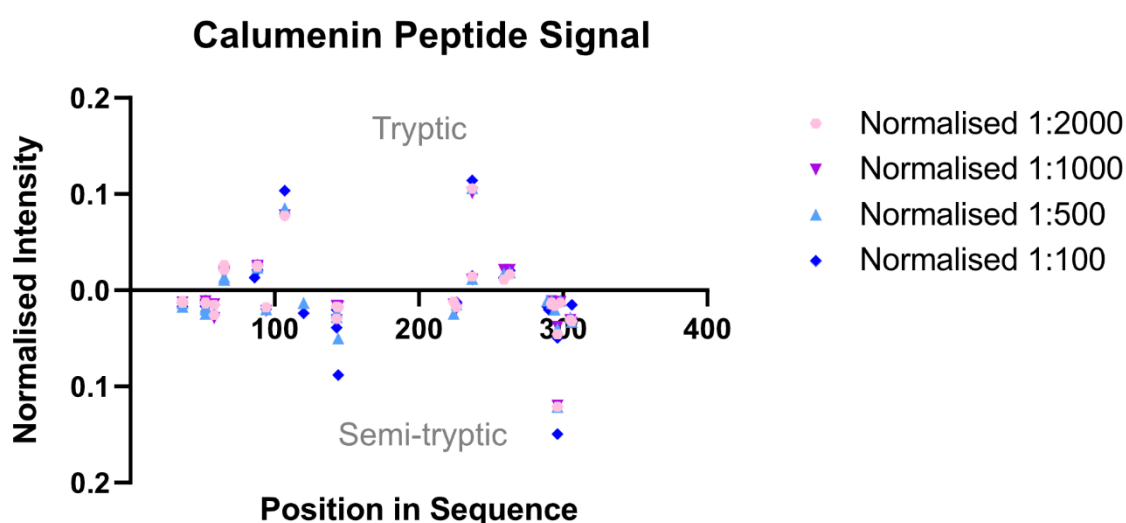


Figure 3.8: The normalised intensity of each peptide mapping to calumenin (UniProt ID: O43852). The residue number for the protein sequence is shown along the x-axis, with the position of each peptide within this sequence shown (median position). Each peptide intensity is normalised by dividing by the sum of all peptide intensities for that sample. Tryptic peptides are shown above the x-axis and semi-tryptic peptides are shown below the x-axis.

This analysis revealed that the semi-tryptic peptide signal (shown in Figure 3.8 as the peptides below the x-axis) was increased at the higher enzyme concentrations. For this protein, semi-tryptic peptides accounted for 49% and 48% of the total signal in the 1:2000 and 1:1000 samples respectively. This increased to 55% and 58% in the 1:500 and 1:100 samples respectively. Again, this conforms to expectation as detailed above.

Although this data is undoubtedly interesting and some insight can be gained, without more robust statistical filtering it is difficult to separate semi-tryptic peptides that are resultant from PK cleavage from the underlying noise of the MS data. Semi-tryptic peptide identifications are less robust than their tryptic counterparts, and although the global trends in semi-tryptic peptide properties match expectation (they are a higher percentage of global signal in higher PK concentrations) the clustering and protein level trends might be more susceptible to false positives and noise.

In the following chapters, semi-tryptic peptides are further investigated in experiments including technical replicates and controls against which to normalise the LiP data.

3.2 Optimisation of lysis conditions

To ensure that the information provided by LiP-MS is as biologically relevant as possible, great care must be taken regarding the manner in which cells are lysed. The initial trial experiments described above were conducted using a SL-DOC lysis buffer, with additional mechanical lysis using bead-beating (Section 2.9.1). This is an effective way of solubilising protein and is amenable to MS analysis. However, the sodium dodecyl sulphate (or sodium laurate (SL)) used in the buffer is a chemical denaturant and the friction generated during the bead beating step could cause a build-up of heat in the lysate (although this step is performed at 4 °C), a further denaturing factor. For these reasons, gentler methods of lysis, that also maintain a good protein yield, were investigated. This is discussed below and summarised in Table 3.1.

Table 3.1: Optimisation of lysis conditions for LiP-MS. The pros and cons of three cell lysis methods are stated: SL-DOC (Section 2.9.1), Dounce Homogenizer (Section 2.9.2) and Triton X-100 (Section 2.9.3). Features of concern include temperature, incubation time, chemical denaturants, mechanical disruption and protein yield.

Lysis Method	Pros	Cons
SL-DOC	<ul style="list-style-type: none"> ❖ Excellent protein yield ❖ Protocol performed quickly and at low temperatures 	<ul style="list-style-type: none"> ❖ Contains chemical denaturants ❖ Potentially mechanically disruptive
Dounce Homogenizer	<ul style="list-style-type: none"> ❖ Does not contain chemical denaturants. ❖ Mechanically gentle ❖ Protocol performed quickly and at low temperatures 	<ul style="list-style-type: none"> ❖ Protein yield is extremely poor
Triton X-100	<ul style="list-style-type: none"> ❖ Reasonable protein yield ❖ Detergent used is mild ❖ No mechanical disruption ❖ Protocol performed quickly and at low temperatures 	<ul style="list-style-type: none"> ❖ Lysis of intracellular compartments may be limited

The first alternative trialled was Dounce Homogenization [19]. This involved resuspension of the cell pellet in PBS, 20 strokes homogenization whilst on ice, followed by passage through a 26 gauge syringe and centrifugation (Section 2.9.2). The benefits of this method are that it can be performed quickly and at 4 °C, the buffer the cells are suspended in contains no denaturing chemicals and is of physiological pH, and the mechanical method of lysis is gentle and will not generate excessive heat. However, despite many alterations to the protocol, such as trialling different centrifugation times and speed, increasing or decreasing the suspension volume, and the introduction of a hypotonic buffer to aid lysis, the protein yield remained extremely poor (summarised in Table 3.2). Therefore, despite its many advantages, this method is not amenable to MS experiments.

Table 3.2: Dounce Homogenization (DH) lysis attempts. The DH protocol was altered in terms of the buffer and centrifugation time used, and the resultant protein yield quantified using BCA assay (Section 2.10). Here the BCA readout is shown to 2

significant figures. For each trial approximately 1 million hMSCs were lysed. Further technical detail can be found in Section 2.9.2. For reference, the same number of cells were subject to SL-DOC lysis (Section 2.9.1) and the results shown in red here.

Buffer	Centrifugation	BCA
500µL PBS	1000g	-
30µL Hypotonic	1000g	-
100µL Hypotonic	14000g	600µg/ml
SL-DOC method		10,000µg/ml

Next, chemical lysis with Triton X-100 buffer was trialed. This involved resuspension of the cell pellet in PBS containing 1% Triton X-100, incubation for 1 minute at 25 °C, followed by centrifugation (Section 2.9.3). Triton X-100 is a gentle detergent (non-ionic surfactant), capable of lysis through permeating the cell membrane. It should not be denaturing to proteins, and is often chosen for native protein studies [192]. Further, lysis can be achieved without any additional mechanical method. The benefits of this method are that it can be performed quickly and with no prolonged exposure to high temperatures, the buffer should not be denaturing to protein, is of physiological pH and there is no mechanical stress placed on the cells. Moreover, protein quantification of the lysates revealed an acceptable protein yield (summarised in Table 3.3). Therefore, this method appears amenable to MS experiments.

Table 3.3: Triton X-100 lysis attempts. The Triton X-100 protocol was altered in terms of the incubation time and temperature used, and the resultant protein yield quantified using BCA assay (Section 2.10). Here the BCA readout is shown to 2 significant figures. For each trial approximately 500,000 hMSCs were lysed. Further technical detail can be found in Section 2.9.3. For reference, the same number of cells were subject to SL-DOC lysis (Section 2.9.1) and the results shown in red here.

Temperature	Time	BCA
On Ice	30 minutes	1000µg/ml
25 °C	30 minutes	3000µg/ml
25 °C	5 minutes	2300µg/ml
25 °C	1 minute	2200µg/ml
SL-DOC method		3600µg/ml

Concerns with this method were that the lysis would not be extended to intracellular compartments, and that the presence of Triton X-100 may inhibit the activity of PK. However, MS analysis of the protein yielded by this lysis method, shown in Section 5, revealed many organelle enclosed proteins. Further, it was demonstrated that PK remains active in this buffer, and has an activity level comparable to that in SL-DOC, as shown in Figure 3.8.

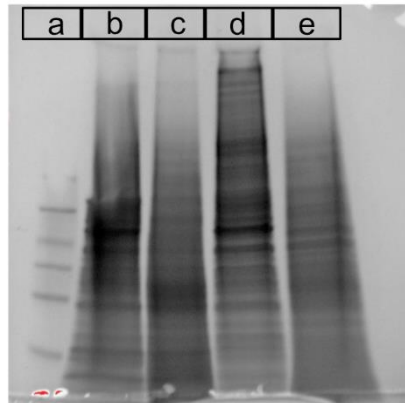


Figure 3.9: SDS page demonstrating LiP of SL-DOC and Triton X-100 lysed samples. (a) Spectra Broad Range Protein Ladder, (b) hMSC SL-DOC lysate, no LiP enzyme. (c) hMSC SL-DOC lysate, incubated with PK at a ratio of 1:2000 E:S (w:w) for 2 minutes, (d) hMSC Triton X-100 lysate, no LiP enzyme, (e) hMSC Triton X-100 lysate, incubated with PK at a ratio of 1:2000 E:S (w:w) for 2 minutes.

In all buffers chosen, a protease inhibitor cocktail was added to prevent endogenous protease activity during lysis. This cocktail did not affect the activity of PK, as demonstrated by the gel in Figure 3.9 and MS data shown in Chapters 4 and 5.

The aim of these trials was to find the lysis protocol most suitable for LiP-MS. Lysis-induced protein denaturation will potentially obscure perturbation-induced protein structural changes and thus limit the amount of useful structural information one can glean from LiP-MS experiments. For this reason, lysis methods containing chemical denaturants, heat, excessive time at non-physiological molecular concentrations, non-physiological pH and harsh mechanical disruption must be avoided. However, it is also important that the method chosen is effective at solubilising protein to the level required for a MS experiment. Here, gentler methods of lysis were found to be ineffective at extracting protein from hMSCs and thus, a middle ground was sought. This middle ground is Triton X-100.

3.3 Summary

In this chapter the optimization of the LiP-MS protocol for use in hMSCs was discussed, with a focus on LiP enzyme concentration, LiP digestion time and the method of cell lysis. An experiment trialling several LiP-MS enzyme concentrations was performed in order to establish bounds for effective LiP and to assess the clarity of whole cell lysate LiP-MS data. The effect of LiP on tryptic peptide signal was investigated using a K means clustering approach to highlight behavioural subsets. The most common behaviour observed was a modest change to tryptic peptide signal with increasing LiP treatment. This implies that appropriately “limited” proteolysis has been achieved in this system. Consequently, the protein structures present are in the main stable and present a consistent resistance to limited proteolytic attack. As a corollary, the majority of tryptic peptides identified in the MS data are not expected to map to surface accessible, flexible protein regions and so should

be unaffected by the LiP enzyme. The second most common behaviour identified in the tryptic dataset was a negative correlation between enzyme concentration and tryptic peptide signal, which is precisely the expected MS fingerprint of LiP cleavage.

It was then tested whether these potential LiP cleavage sites correlated with the known structural determinants of LiP by running the entire tryptic peptide dataset through Nickpred, a program that assigns scores based on these structural features in order to predict which protein regions will be most susceptible to LiP. This revealed that a subset of the indicated sites of LiP activity did indeed have a significantly higher LiP-susceptibility score than the dataset as a whole.

K means clustering analysis was also performed on the semi-tryptic peptide dataset and revealed complementary patterns of behaviour to the tryptic case. In line with expectation, a prevalent behaviour highlighted was a positive correlation between enzyme concentration and semi-tryptic peptide signal. Together with the tryptic analysis, this analysis also proved insightful into which E:S ratio is most suited to LiP-MS of hMSC lysates. For example, the majority of the semi-tryptic peptides identified did not demonstrate a substantial difference in signal between the 1:2000 and 1:1000 samples, and the 1:500 sample either increased semi-tryptic peptide signal, or similarly showed little change. Crucially, another prevalent behaviour highlighted was a decrease in signal at E:S 1:100. This potentially indicates that this enzyme concentration can produce excessive cleavages, thus resulting in a loss of identifiable semi-tryptic peptides and so a reduction in structural data. Taken together, this suggests that in order to achieve appropriately limited proteolysis, the maximum recommended E:S ratio for hMSCs is 1:500.

The semi-tryptic dataset was also analysed in Nickpred. To some extent, the different peptide behaviours highlighted in the clustering analysis were also reflected in the associated Nickpred scores, thus providing some evidence that semi-tryptic peptide signal is linked to protein structural features. It was noted however that this data appears noisier.

The contribution of tryptic and semi-tryptic peptides to total protein signal, and how this changes with enzyme concentration, was also considered at the individual protein level. In the example considered, the contribution of semi-tryptic peptides to protein signal increased with enzyme concentration, as expected.

However, it is important to note that as only one replicate per condition was used in this MS experiment, one must not over-interpret this data. This experiment was intended as a proof of principle, "trial" LiP-MS to establish that this technique is capable of producing data. The aim was not to draw biological inference, but instead to use different LiP concentrations to determine the extent to which LiP-MS is achievable in hMSCs and to generally optimise the

LiP-MS protocol for this system. Despite this, this experiment did facilitate the identification of suitable LiP conditions, as demonstrated by the clustering analysis detailed above.

This initial insight must be followed up by more robust statistical filtering in order to separate semi-tryptic peptides that are resultant from PK cleavage from the underlying noise of the MS data. Semi-tryptic peptide identifications are seemingly less robust than their tryptic counterparts, and although the global trends in semi-tryptic peptide properties match expectation (they are a higher percentage of global signal in higher PK concentrations) the clustering and protein level trends might be more susceptible to false positives and noise. In the following chapters, semi-tryptic peptides are investigated further in experiments including technical replicates and controls against which to normalise the LiP data.

Finally, cell lysis-induced perturbations to native protein structure will reduce the capacity of LiP-MS to act as a protein structural probe. To reduce this risk, various lysis methods were trialled and optimised. Overall, a Triton X-100 based method was chosen as striking a good balance between maximising protein yield for MS and lack of denaturing variables.

In conclusion, the work detailed in this chapter indicates that LiP-MS is indeed possible in hMSC lysates as the MS signal from tryptic and semi-tryptic peptides largely conforms to expectation. Further, the LiP protocol, in particular the LiP reaction and method of cell lysis was optimised and these recommendations enacted going forward.

Chapter Four: LiP-MS can act as a probe of protein structure in complex samples

In the previous chapter a basic proof of concept for LiP-MS was established and the protocol optimized for application to hMSCs. The protocol was shown to be able to extract sufficient quantities of protein for MS analysis, whilst also limiting the inclusion of potentially denaturing variables in the cell lysis procedure. Importantly, LiP-induced shifts in peptide signal in hMSC lysates were detectable via MS, with attendant LiP peptide fingerprints that largely conformed to the central dogma of LiP-MS: protease accessible regions of protein structure are additionally cleaved by the broad specificity LiP protease and such cleavage sites are identifiable in LiP-MS data, indirectly through a decrease in tryptic peptide signal and directly through a corresponding increase in semi-tryptic peptide signal.

To extend this principle further, one must examine the ability of the technique to detect **changes** to protein structure, which will be the aim of this chapter. Historically, LiP has been used to gain information about static, single protein structures, such as identifying disordered regions in proteins of unknown structure, or the site of domain linker regions [155, 161, 162, 193, 194]. However, the cell is a highly dynamic environment and so maintenance of proteostasis is both contingent upon, and can be corrupted by, dynamic alterations to protein structure. For example, as discussed in Section 1.5, much of the early stress response in hMSCs may be governed by changes to protein conformation, such as complex formation or small molecule interactions. Therefore, if LiP-MS is to act as an effective, global probe of protein structure it must be capable of identifying specific regions of perturbation-induced protein conformational change, in this case in hMSC lysates as the subject of this study.

This was tested by utilizing the denaturing effect of heat on the proteome. Temperature changes of just a few degrees can have a substantial impact on protein stability [195], with proteins becoming quickly unfolded [166]. Hence, a heat shock will be expected to perturb the integrity of the proteome by inducing additional unfolding. There is a particular need for the systems-level identification and assessment of protein unfolding and protein aggregation events, as such structural alterations are characteristic of several neurodegenerative diseases [64, 65]. Further, previous work in the Swift lab has predicted an increased burden of misfolded protein in senescent hMSCs [10].

The aim is to establish that LiP-MS is able to detect regions of heat-induced protein unfolding, by virtue of a comparative signal loss from tryptic peptides or signal gain from semi-tryptic peptides. Further, to assess the degree of proteome coverage achieved by LiP-MS of hMSCs, whilst also investigating potential quality control procedures and testing the data's robustness.

The experimental design is summarized in Table 4.1 and is similar to an experiment conducted previously by the Picotti lab, in which they aimed to uncover the determinants of protein thermostability by performing LiP-MS on protein lysates heated to several temperatures [166]. This experiment was conducted using Y201 cells, an immortalized hMSC line which has been shown to maintain much of the primary hMSC stress response (Section 2.5) [180].

In the first condition, all previously discussed steps (Section 3.2) were taken to preserve native protein structure, with LiP performed immediately following cell lysis. In the second condition, the cell lysate was incubated at 37 °C for 30 minutes before the LiP enzyme was applied. Finally, in the third condition, the cell lysate was incubated at 42 °C for 30 minutes before application of the LiP enzyme. In each case, the conditions under which the LiP digestions occur are identical. Heated lysates are rapidly cooled to 25 °C before application of PK, which is applied at an E:S of 1:2000 for 2 minutes, followed by quenching with boiling. For each of these conditions 6 samples were analyzed by mass spectrometry, half of which were subjected to LiP treatment and half were digested only with trypsin (referred to as "Control").

Table 4.1: Experimental Plan. For each of the 3 heat conditions, 3 replicates were subjected to LiP treatment (E:S 1:2000 for 2 minutes) and 3 replicates were digested with trypsin only, thus acting as a control.

	<u>Condition 1</u> No Heat "Normal"	<u>Condition 2</u> 37 °C "Intermediate"	<u>Condition 3</u> 42 °C "Heated"
Control	x3	x3	x3
LiP	x3	x3	x3

Exposure of the proteome to a temperature of 42 °C is sufficient to induce heat shock [196] and so the expectation is to observe increased protein unfolding events in Condition 3. Increased protein unfolding should result in increased LiP enzyme activity, and so an increase in the abundance/identification of semi-tryptic peptides and a corresponding

decrease in the number and intensity of attendant tryptic peptides, as discussed in Section 3.1. Condition 2 serves as an intermediate, with the prolonged incubation time consistent with Condition 3, but in this case performed at a physiological temperature. Consequently, the experimental design supports an additional intermediate step and should highlight peptides displaying a consistent trend from control, through 37 °C to 42 °C . For ease of understanding, Conditions 1 to 3 will henceforth be referred to as “normal”, “intermediate” and “heated”, with “normal” used simply to refer that without additional perturbation.

This experiment will examine the proficiency of LiP-MS as a global probe of protein conformation in human cell lysates through assessment of the peptide fingerprint associated with each of the experimental conditions.

4.1 Tryptic peptides

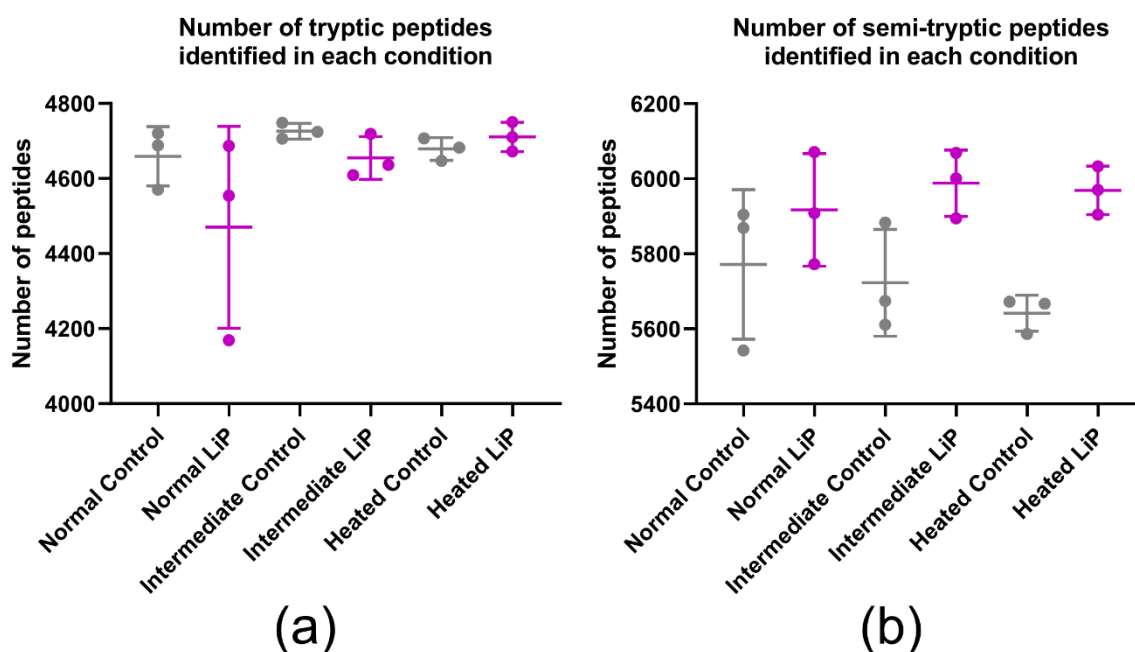


Figure 4.1: Number of peptides identified in each run, defined as having a non zero reported intensity value. Internal controls for each condition are shown in grey. Samples to which the LiP enzyme is applied are shown in pink. Error bars display the mean and standard deviation of the 3 replicates.

As an initial assessment of the MS data, the number of peptides identified (non-zero reported intensity) in each run was calculated (Figure 4.1). As discussed, the expected peptide footprint of global LiP cleavage is reduction in tryptic peptide abundance and increase in semi-tryptic peptide abundance comparative to the control case. In addition, for some highly enzyme-accessible protein regions or marginally stable proteins, if this area has been consistently digested the associated tryptic peptide(/s) may not be identifiable (/present) in the LiP sample at all. In an extreme case, the protein itself may be destabilised

by the perturbation and/or limited cleavage so that it is digested to an undetectable level. There is some potential evidence of these phenomena in Figure 4.1(a), as in the normal and intermediate conditions there is a reduction in the number of tryptic peptides identified in the LiP samples comparative to the respective controls. This behaviour is not reflected in the heated condition and would appear to be more likely to reflect reasonable variation across a typical proteomics experiment.

In the semi-tryptic case, it was initially expected that there should be very few semi-tryptic peptides present in the control conditions, as there is no LiP applied to generate them. However, as shown in Figure 4.1(b), this is not the case, with over 5000 semi-tryptic peptides identified in each of the control samples. Whilst initially concerning, a review of the literature revealed that, perhaps surprisingly, this is a phenomenon observed in many proteomics experiments in which the only enzyme added is trypsin [197]. Assessment of semi-tryptic search results from MS of both complex biological samples and synthetic protein mixes revealed that such peptides are a bi-product of both endogenous protease activity and experimental sample preparation. For example, trypsin specificity is shown to be influenced by solvent choice and digestion time, with aqueous solvents and overnight digestion resulting in the production of more semi-tryptic peptides when compared to acetonitrile based buffers [198]. Other factors include the potential for in-instrument cleavage, and false identification due to low mass accuracy or the parameters of the search engine used [199]. Additionally, native proteases in the cell can induce cleavages as either part of normal cellular activity (e.g. pro-peptide cleavages and activation steps in cascades) or potentially during degradation as part of normal proteostasis, e.g. turnover.

Despite these issues, one encouraging aspect is that Figure 4.1(b) shows that more semi-tryptic peptides were identified in the LiP samples than the control and that this is consistent over each of the experimental conditions.

Returning to the tryptic peptides, a potential source for concern is highlighted in Figure 4.1(a) by the increased variability of the LiP treated samples in the normal condition. It appears that one sample (referred to as C1b_3) contained far fewer peptides than the others of the same condition, and indeed across the entire experiment. This variability, and the variability of the MS data as a whole was further investigated by performing Principal Component Analysis on

the tryptic peptide intensity data (Figure 4.2).

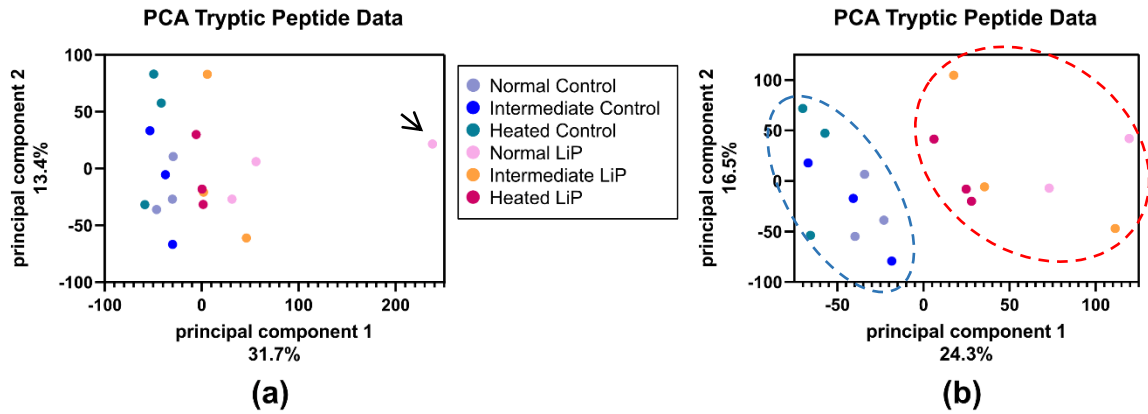


Figure 4.2: Principal Component Analysis on tryptic peptide intensity data. The percentages shown on each axis represent the explained variance ratio. (a) PCA of full tryptic peptide dataset. (b) PCA of dataset with outlying replicate (C1b_3) removed.

The most striking result from this analysis is the difference between C1b_3 and the rest of the dataset (Figure 4.2(a), indicated by arrow). Given that these are technical replicates, there should be no biological variability between replicates of the same condition and so it is likely that this variability is due to the quality of this particular samples MS run, or sampling preparation error. However, it is difficult to show this. Considering the rest of the data, the PCA revealed a clustering based on LiP treatment, as noted by the separation of these two dataset classes by PC1 in Figure 4.2(a). Principal component analysis was also repeated on the peptide intensity dataset after removal of the outlying replicate C1b_3, as shown in Figure 4.2(b). The clear separation of LiP treated samples and control samples was maintained. Variance separation based on LiP treatment implies that LiP has been successfully achieved in these cell lysates and that this can be seen in the MS signal at the level of the peptide data. There is little discernible variance identified between heat conditions on the PCA plots. However, this is not entirely unexpected as the effect of these heat perturbations is believed to be modest on the overall abundance of peptides in the proteome and likely dominated by the effect of LiP treatment itself. Indeed, this is reassuring as LiP changes should exceed those induced by heat shock and the technique itself sets out to investigate changes in protein structure rather than abundance.

4.1.1 Distribution of abundance values

The next step in this analysis is to consider changes at the level of peptide abundance. For each peptide identified, the log₂ fold change of its reported intensity in the LiP treated sample compared to that of the control sample was calculated, for each of the 3 conditions.

For example, for each tryptic peptide i , $X_i = \log_2 \left(\frac{\text{Intensity of } i \text{ in the LiP samples}}{\text{Intensity of } i \text{ in the control samples}} \right)$. The distribution of these X_i values was assessed through grouping them into bins between -4

and 4 in increments of 0.1, as shown in Figure 4.3. In this instance, for the first condition the C1b_3 run was excluded and instead the analysis was completed with the remaining 2 replicates.

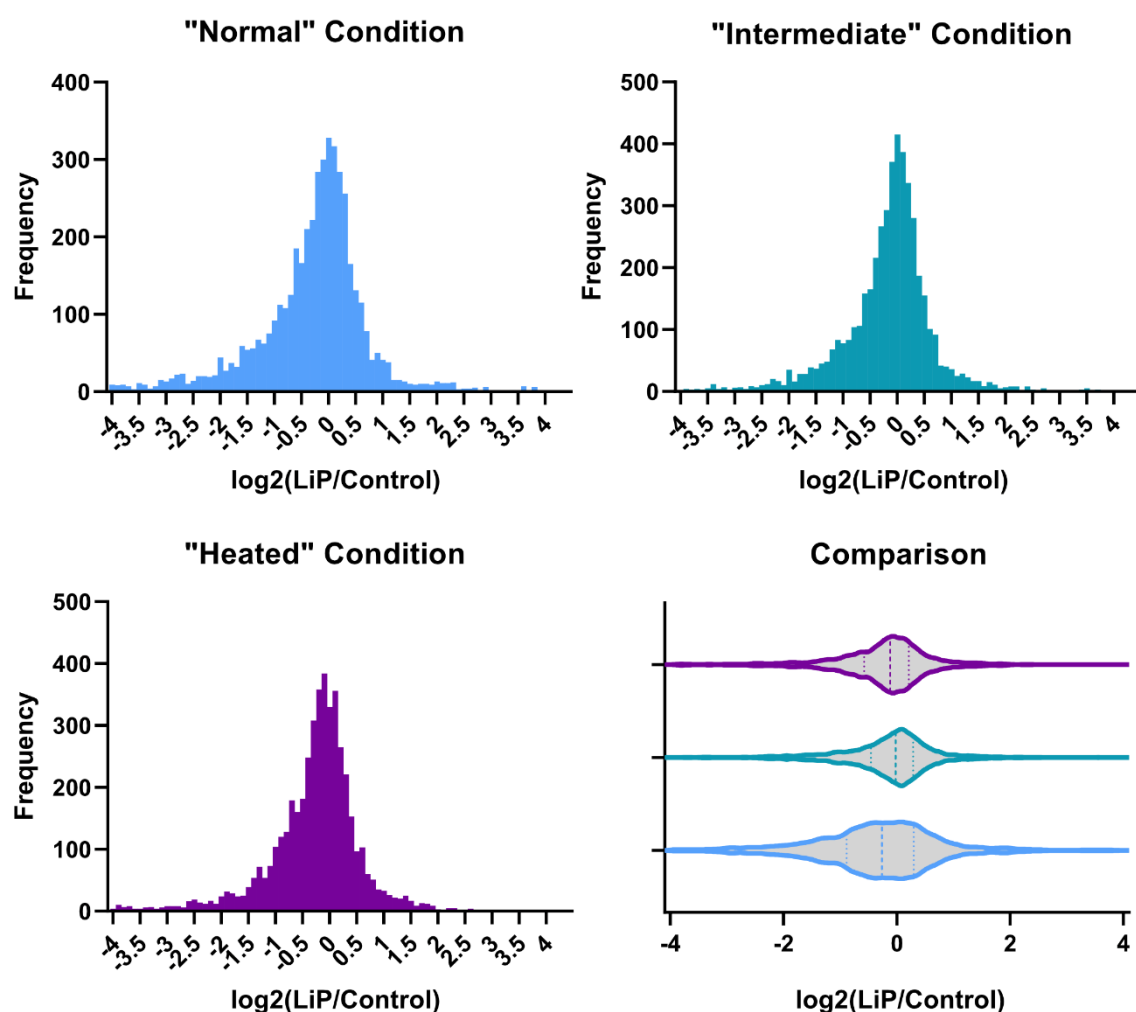


Figure 4.3: Tryptic peptide intensity \log_2 distribution. For the normal condition, 78% of X_i values fell between -1 and 1. This number was 83% in the intermediate condition and 83% in the heated condition. 17% of peptides in the normal condition had $X_i < -1$, 12% in the intermediate condition and 13% in the heated condition. The graph on the bottom right shows the X_i values for all conditions plotted together, this time displayed as a violin plot.

In each of the heat conditions, the vast majority of these values fell between -1 and 1 (an average of 81% of all peptides). This aligns with expectation; if the majority of tryptic peptides stem from protein regions unaffected by PK proteolysis, which is appropriately limited to accessible/flexible regions of protein structure, it might be expected that all peptides mapping to these unaffected areas should present similarly in both the control and LiP treated samples. Also evident in Figure 4.3 is a leftwards skew to the \log_2 distribution. Across the different heat conditions an average of 14% of all X_i values were below -1, leaving just 5% with $X_i > 1$. This shows that there is a strong subset of the tryptic dataset in

which the abundance of peptides decreases with LiP treatment. This is precisely the fingerprint one would expect from LiP-MS.

Considering the log₂ distribution alone does not provide direct insight into the effect of the heat conditions on LiP peptides. If anything, the slightly stronger leftward skew (Comparison, Figure 4.3) perhaps suggests more LiP activity in the “normal” condition, though this observation is dependent on normalisation and similar data processing pipelines. To learn more, the significance of these changes was investigated.

4.1.2 Significantly changing tryptic peptides

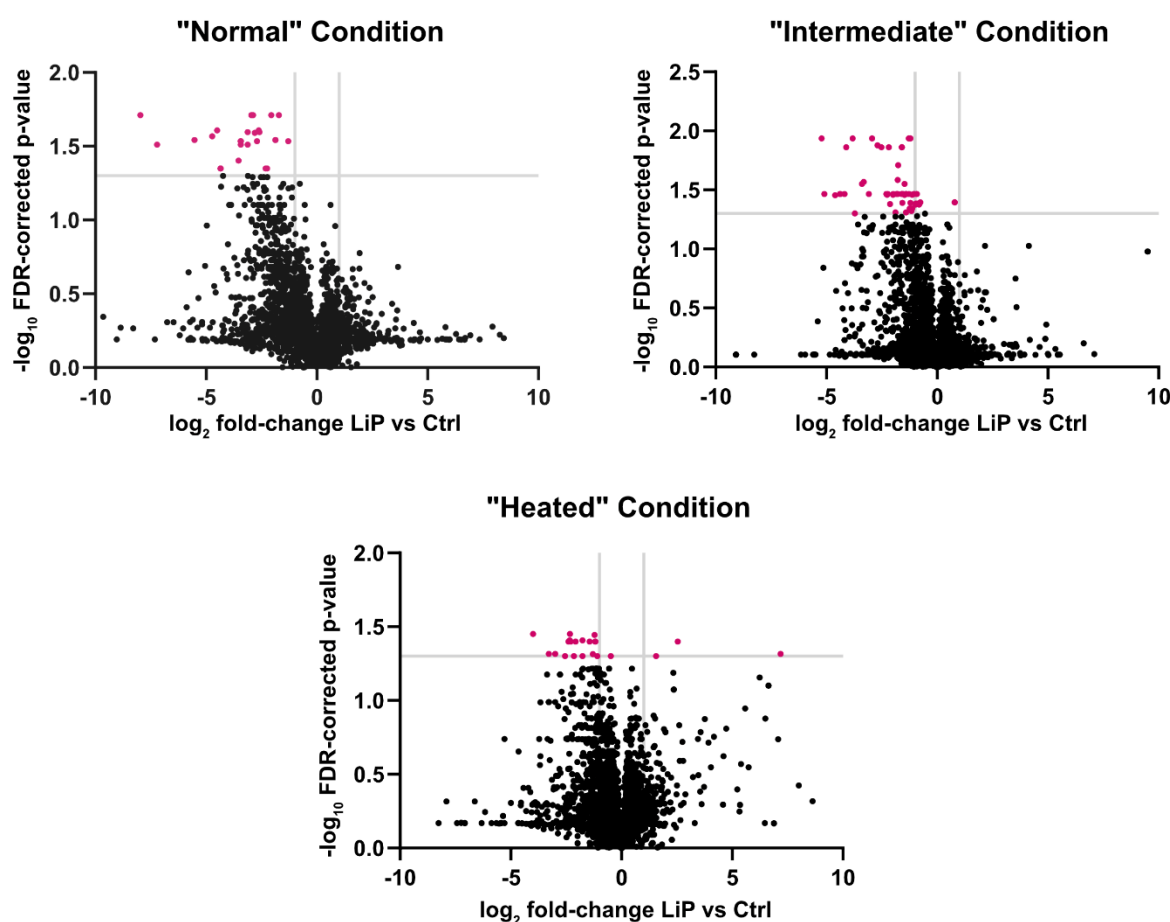


Figure 4.4: Tryptic peptide abundance changes. Volcano plots showing the log₂ fold change of the abundance of tryptic peptides in the LiP treated samples vs the control samples, plotted against the respective $-\log_{10}$ FDR-corrected p-value. P-values were determined using Welch t-test followed by the Benjamini-Hochberg multiple testing correction. The significance threshold of 0.05 is highlighted by a grey line on the y axis and significantly changing peptides are highlighted in pink. Log₂ fold change of -1 and 1 are also highlighted by grey lines on the x-axis.

Peptides with a **significant** change in abundance upon LiP treatment were determined using Welch’s t-tests and the Benjamini-Hochberg multiple testing correction. Welch’s t-test was chosen over a Student’s t-test as it does not presume equal variance between groups. However, if variances are equal, it performs identically to a Student’s t-test. Since it cannot

be guaranteed that the standard deviations for each group (control, LiP treated) are identical, Welch's t-test is a responsible choice. Benjamini-Hochberg is a commonly used multiple testing correction in omics sciences, due to its capacity to reduce Type 1 errors (false positives). This is also the method chosen by the Picotti lab specifically for LiP-MS experiments [167].

The results for tryptic peptides are shown as volcano plots in Figure 4.4, with significantly changing peptides ($p_{\text{adj}} < 0.05$) highlighted in pink. Under these test conditions, few peptides were found to meet the significance threshold. Peptides with significantly diminished abundance in the LiP treated samples compared to the control numbered 26 in the normal condition, 53 in the intermediate condition and 18 in the heated condition. However, the majority of significantly changing tryptic peptides are all reductions in abundance, consistent with the LiP hypothesis where they have been nicked and removed from detection. Although one expectation of the experiment was to find more LiP activity in the intermediate and heated conditions, this is not clearly demonstrated here. This analysis was then repeated with the semi-tryptic peptide dataset, to investigate this further (Figure 4.5, 4.6).

4.2 Semi-tryptic peptides

4.2.1 Distribution of abundance values

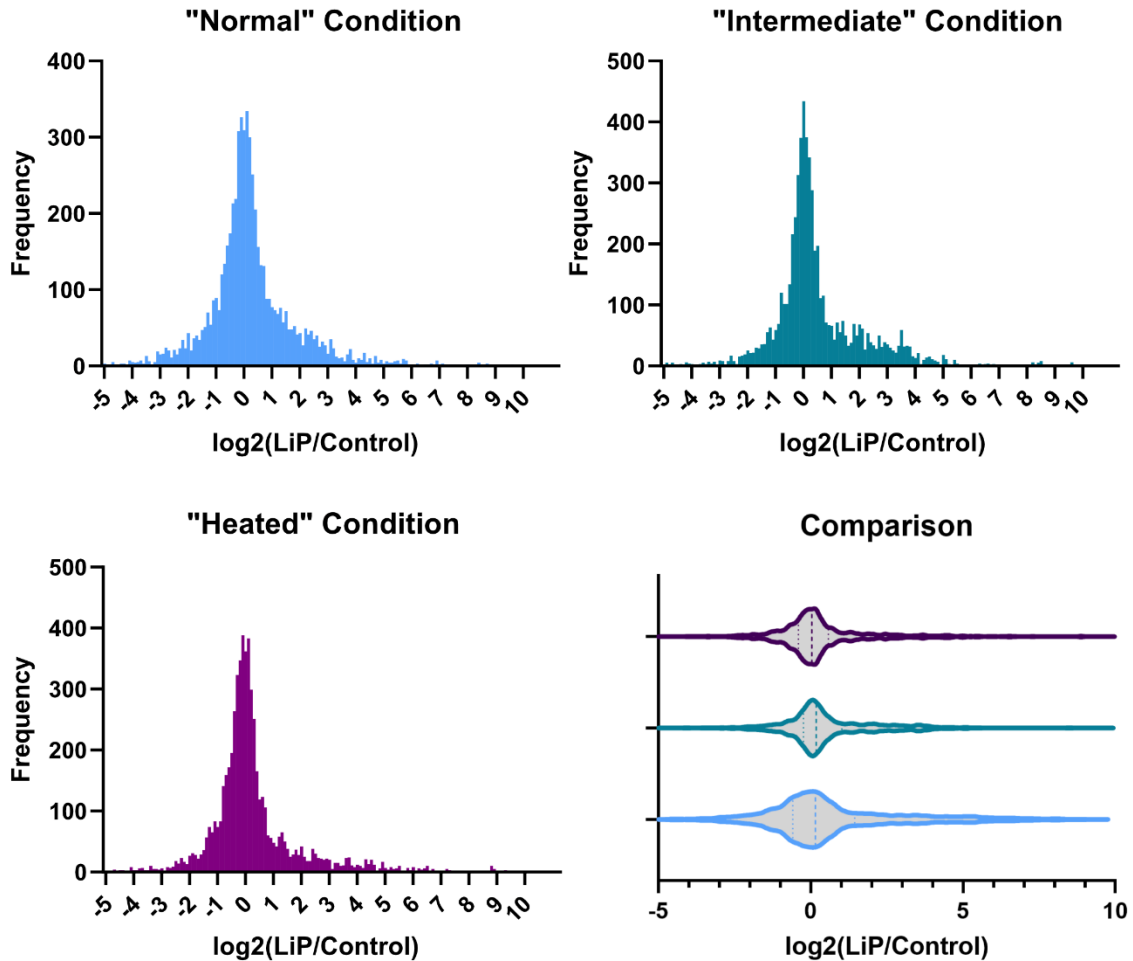


Figure 4.5: Semi-tryptic peptide intensity \log_2 distribution. For the normal condition, 66% of X_i values fell between -1 and 1. This number was 67% in the intermediate condition and 71% in the heated condition. 21% of peptides in the normal condition had $X_i > 1$, 24% in the intermediate condition and 19% in the heated condition. The graph on the bottom right shows the X_i values for all conditions plotted together, this time displayed as a violin plot.

As before for tryptic peptides, in each of the heat conditions the vast majority of X_i values fell between -1 and 1 (though slightly less with an average of 68% of all peptides, Figure 4.5). Assuming that these semi-tryptic peptide identifications are legitimate, and indeed have a baseline presence in tryptic-only proteomics experiments, this is similar to the above tryptic case and represents the peptides not affected by the LiP enzyme. Moreover, there is again a skew in the \log_2 distribution, but in the semi-tryptic case this is towards the right. Across the different heat conditions an average of 21% of all X_i values were above 1, leaving 11% with $X_i < -1$. This shows that there is a subset of the semi-tryptic dataset in which the abundance of peptides increases with LiP treatment, as predicted.

4.2.2 Significantly changing semi-tryptic peptides

Again, it is important to consider how consistently these peptides are identified across technical replicates, in order to determine how many have a **significant** change in abundance upon LiP treatment. This is performed in the same manner as for the tryptic case and the results shown in Figure 4.6.

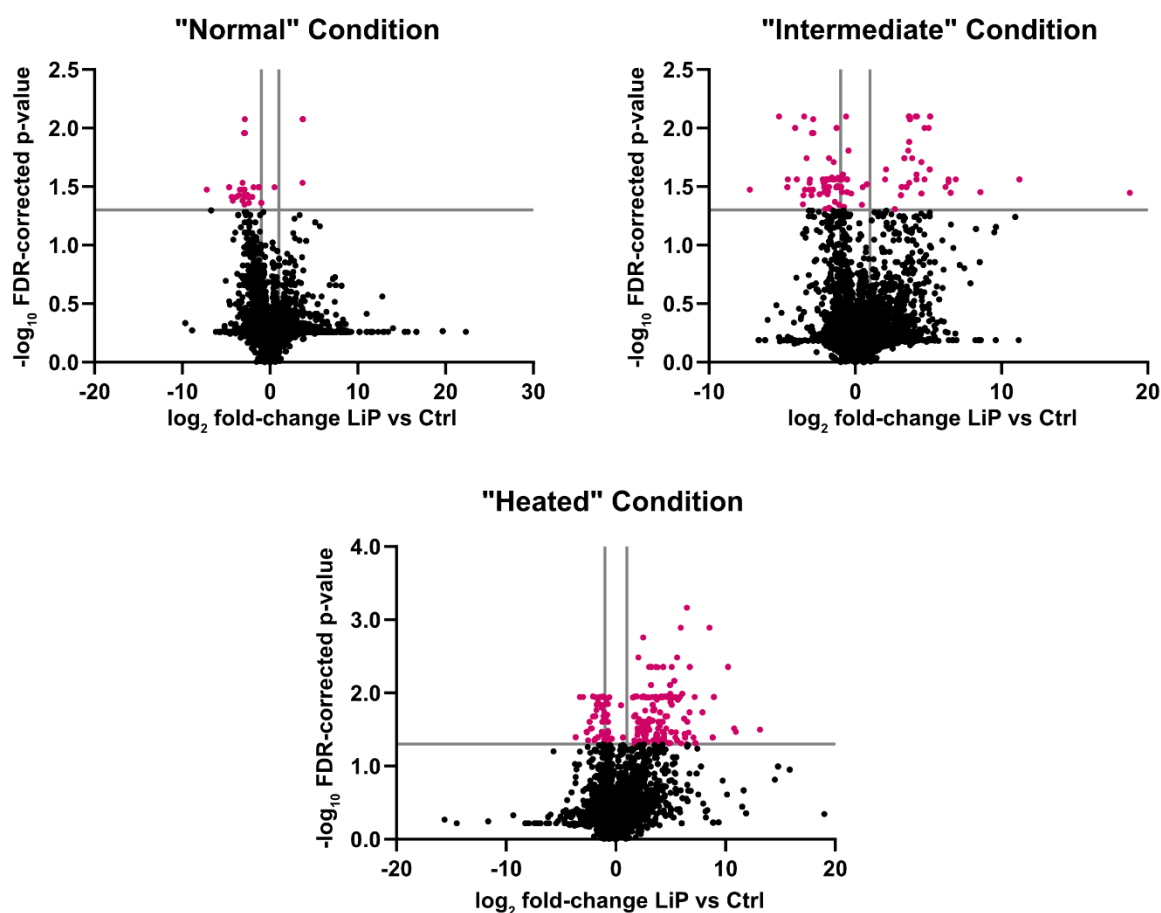


Figure 4.6: Semi-tryptic peptide abundance changes. Volcano plots showing the log₂ fold change of the abundance of semi-tryptic peptides in the LiP treated samples vs the control samples, plotted against the respective $-\log_{10}$ FDR-corrected p-value. P-values were determined using Welch t-test and the Benjamini-Hochberg multiple testing correction. The significance threshold of 0.05 is highlighted by a grey line on the y axis and significantly changing peptides are highlighted in pink. Log₂ fold change of -1 and 1 are also highlighted by grey lines on the x-axis.

This analysis revealed significant abundance changes in both directions. For example, in the normal condition 38 semi-tryptic peptides were found to have significantly higher abundance in the control sample compared to the LiP treated samples. This behaviour was echoed in the other heat conditions, with 81 and 93 semi-tryptic peptides with the same footprint identified in the intermediate and heat conditions respectively. Although the presence of semi-tryptic peptides in control samples has been discussed previously, it is not immediately obvious why their abundance would increase significantly following LiP treatment. A possible explanation as to why some semi-tryptic peptides display greater abundance in the control,

is that this effectively represents the inherent noise of the data, which the statistical model is unable to fully consider. A further consideration might be that in high temperatures and exposures to PK, the protein in question is degraded further, and the parent peptide that generates the semi-tryptic peptides is further digested thereby removing it from the pool. This idea is visualised in Figure 4.7. This is similar to that which was observed in the higher enzyme concentrations in Section 3 and would account for the increase in semi-tryptic peptides with this behaviour in the heated and incubated conditions.

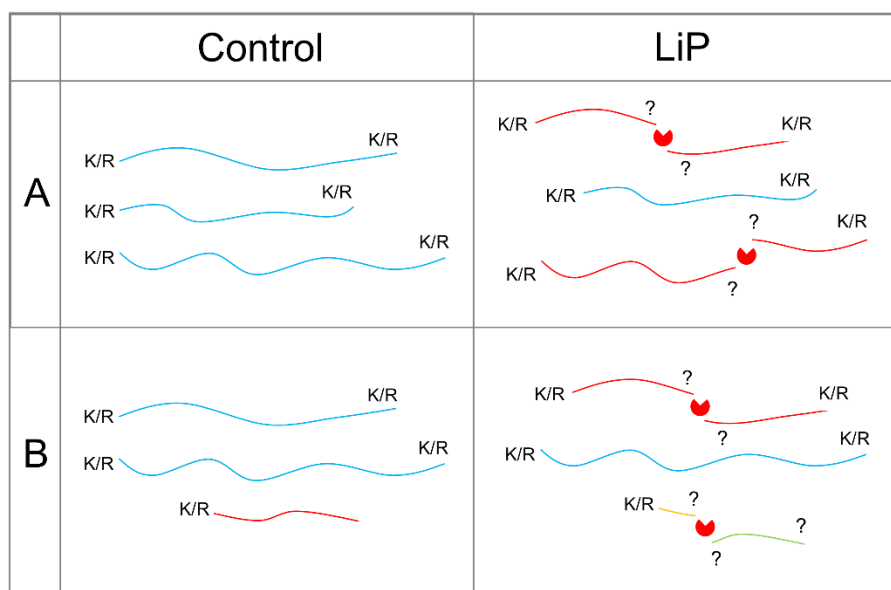


Figure 4.7: In an idealised LiP-MS experiment, no semi-tryptic peptides are present in the control condition (A). In the LiP condition, the additional cleavages (pac-man) will result in a decrease in the tryptic peptide signal (blue) and an increase in semi-tryptic peptides (red). However, in an experiment in which semi-tryptic peptides are also identified in the control (B), digestion of these peptides upon addition of the LiP enzyme could result in non-tryptic peptides (green), or semi-tryptic peptides too small for identification (orange).

However, considering changes in the other direction – that of significantly higher semi-tryptic peptide abundance in the LiP treated samples than the controls – 4 such peptides were identified in the normal condition, 44 in the intermediate condition and 214 in the heat condition. The number of significantly changing semi-tryptic peptides increases substantially with heat treatment, implying that there is an increase in enzyme accessible protein regions. The incubation of protein lysates at 37 °C (intermediate) and 42 °C (heat) for 30 minutes was expected to induce some protein denaturation, when compared to LiP performed immediately after lysis (normal). Increased protein denaturation/unfolding will result in more enzyme accessibility. Therefore, through analysing these semi-tryptic peptides, LiP-MS does seem capable of identifying protein unfolding over different experimental conditions.

As an additional note, this model presumes that increased denaturation leads to the generation of *the same* semi-tryptic peptide in every case. However, it is also likely that it leads to additional susceptible peptide bonds becoming accessible to the PK proteinase,

and so rather than increasing the abundance at a given bond, cleavages occur at several sites in the vicinity of the locally unfolded region which in turn would generate multiple semi-tryptic peptide species and diffusion of the attendant signal in the mass spectrometer across them. The net effect here could be a reduction in the abundance of an individual semi-tryptic peptide species.

Another point of consideration is how many of these significantly changing peptides are shared across the different conditions. Figure 4.8(a) displays the overlap between all significantly changing semi-tryptic peptides across the different heat conditions. This is then separated into those with a significantly higher abundance in the control samples (Figure 4.8(b)) and those with a significantly higher abundance in the LiP treated samples (Figure 4.8(c)). There is no overlap between the data shown in 4.8(b) and 4.8(c). In other words, the direction of change is always consistent across each of the heat conditions.

The increase in the number of peptides generated via LiP cleavage with greater sample perturbation (C1-3) is clear in Figure 4.8(c). There is also some demonstration of overlap across these different conditions, which is to be expected. For example, peptides significantly changing in all conditions (centre of the Venn diagram) represent regions of protein that are always enzyme-accessible, e.g. areas that do not undergo structural alterations with the different conditions. Peptides shared between the intermediate and heated conditions but absent from the first condition, potentially map to regions of protein that become more enzyme accessible (unfolded) with incubation at 37 °C (intermediate) and remain so at 42 °C (heated). Peptides unique to the heated condition (the largest subset of the data) are potentially protein regions that remain stable when incubated at 37 °C, but unfold at 42 °C. The possible interpretations of each segment of the Venn diagrams are presented in Figure 4.9.

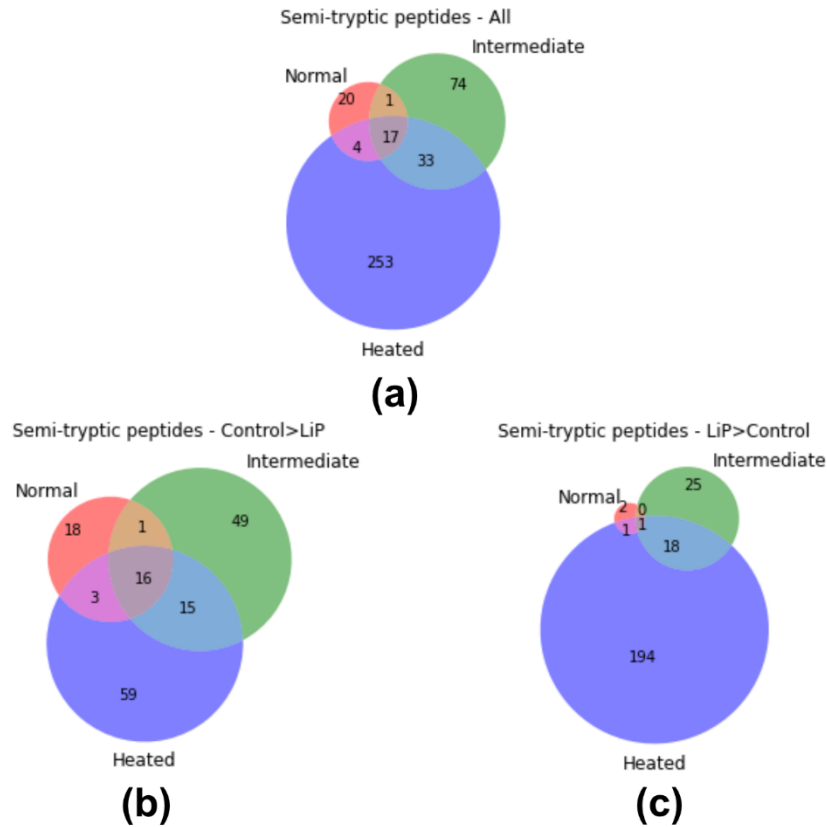


Figure 4.8: Weighted Venn Diagrams of significantly changing semi-tryptic peptides. (a) All significantly changing semi-tryptic peptides. (b) Semi-tryptic peptides with significantly higher abundance in the control samples. (c) Semi-tryptic peptides with significantly higher abundance in the LiP treated samples.

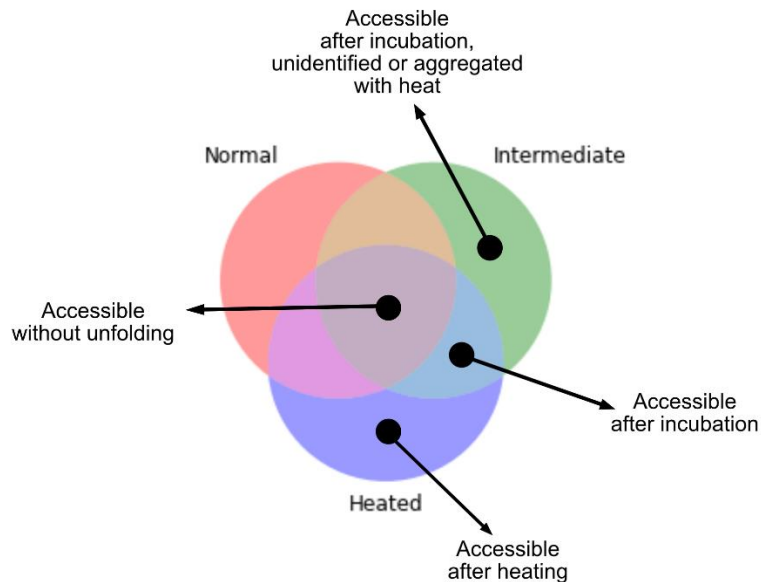


Figure 4.9: Possible interpretation of LiP peptide behaviours shown in Figure 4.8 - semi-tryptic peptides with significantly higher abundance in the LiP treated samples (LiP peptides). For example, LiP peptides found in all of the experimental conditions (centre) represent protein regions that are natively enzyme accessible. LiP peptides unique to the Intermediate condition may represent protein regions that become accessible after incubation but are unidentified or aggregated with heat. LiP peptides unique to the Heated condition can be interpreted as mapping to regions of heat-induced unfolding, etc.

4.3 Direct comparison of LiP treated samples

For completeness, the semi-tryptic peptides from LiP treated samples in the normal and heated conditions were compared directly, without normalisation to their respective controls. This comparison is shown in Figure 4.10. Despite large differences in semi-tryptic abundance (log2 change of generally between -10 and 10), none met the significance threshold when the multiple testing correction was applied (red dotted line, Figure 4.10). This is potentially due to semi-tryptic LiP peptides being identified in MS in a less robust manner, as discussed previously in Section 4.2.

Overall, due to the uncertainty surrounding semi-tryptic peptides in LiP-MS experiments, comparing LiP treated samples to control samples within experimental conditions appears to be a better controlled experiment. Given the lack of understanding regarding the presence of semi-tryptic peptides in MS data without LiP, one cannot fully account for the effect of different experimental conditions on the signal of these peptides. For example, heating and incubating protein lysates could lead to more endogenous protease activity and subsequent activation of different cleavage pathways, which lead to a different peptide species being produced. In this experiment, protease inhibitors were added to the lysis buffer, but that is not to say they are 100% effective. Furthermore, the differing experimental conditions could affect protein solubility and so the identification of peptides by MS. These factors further confound the identification and relative quantification of semi-tryptic peptides that are present in the MS data as a result of LiP cleavage. Therefore, if semi-tryptic peptides are to be formally considered as markers of protein unfolding, normalising LiP samples to their respective controls, within experimental conditions, offers the preferred route since it can alleviate these issues to a certain extent.

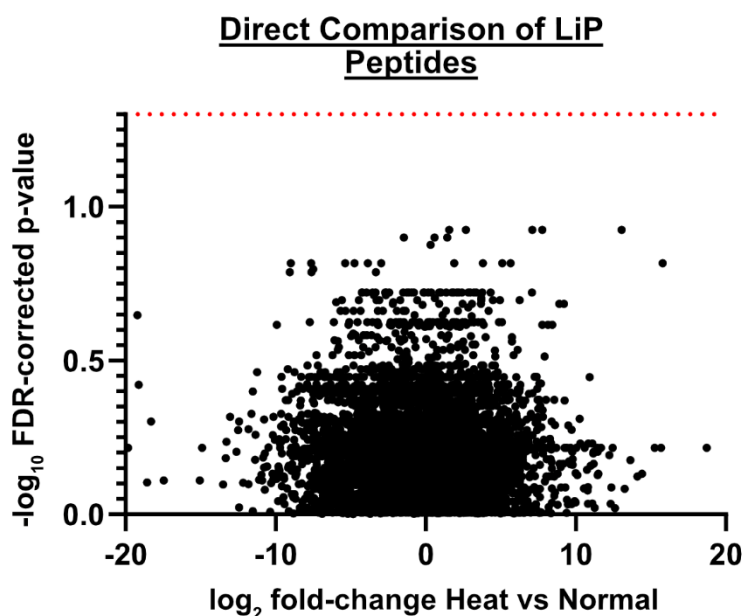


Figure 4.10: Semi-tryptic peptide abundance changes between the normal LiP treated sample and the heated LiP treated sample, shown as a volcano plot. FDR-corrected p-values were determined using Welch t-test and the Benjamini-Hochberg multiple testing correction. No semi-tryptic peptides were found to meet the significance threshold of $q < 0.05$, shown here by the red dotted line.

4.4 Protein examples

Once LiP peptides have been identified, these can then be mapped onto known protein structures. This allows us to investigate the LiP peptide fingerprint in more detail, in order to gain a better appreciation of how heat treatment may have effected which areas of protein are cleaved.

For example, the 18 semi-tryptic peptides found to be significantly increasing in abundance in both the intermediate and heated conditions (Figure 4.8(c), blue segment) were considered in their parental protein contexts. The positions of these peptides along their parent protein sequence was determined. Further, for each of the parent proteins highlighted, the rest of the LiP-MS data was searched in order to determine the level of sequence coverage provided by tryptic and semi-tryptic peptides. In order to understand structural transitions, high sequence coverage is desirable. Finally, these proteins were then further investigated for available structural information, such as whether PDB structures were available and, if so, the resolution and sequence coverage provided. This is exemplified in Table 4.2.

Table 4.2: Semi-tryptic peptides with significantly higher abundance in the LiP treated sample compared to control, in both the intermediate and heated conditions (Figure 4.8(c), blue segment). First 10 are displayed here. Displayed in the table is the semi-tryptic peptide, the corresponding parent protein (Uniprot ID and Gene Name) and the position of the peptide within the protein sequence. Start/End Pos reflect residue numbers, along the protein sequence from N to C terminus. Also detailed is the percentage sequence coverage of that parent protein from tryptic and semi-tryptic peptides in the LiP-MS data, PDB identifier name where a structure is available (not available is indicated by 'X') and the percentage of the protein structure that has been resolved in the PDB file. Highlighted in red is the protein investigated further in Figure 4.11.

Sequence	Protein	Gene Name	Start Pos	End Pos	LiP-MS Coverage (%)	PDB Identifier	PDB Coverage (%)
DDGRFPQVIK	ALDOA_HUMAN	ALDOA	88	99	31.32	1ALD	100
VLRPQVTAV	DYHC1_HUMAN	DYNC1H1	482	491	10.42	5NUG	100
QAGASQFETSAK	F8WCA0_HUMAN	VAMP2	72	85	38.98	X	X
KGVLKDPEIA	TAOK3_HUMAN	TAOK3	2	12	4.45	6BDN	35
SEISESHPNLNLDLNPQNI NK	A0A087X253_HUMAN	AP2B1	146	168	9.53	X	X
KLEGALGADTTEDGDEK	MYOF_HUMAN	MYOF	1092	1109	7.67	2DMH	6
TFDQLALDSPK	F8VUA6_HUMAN	RPL18	69	80	29.23	X	X
EKTVEVLEPEVTK	CYFP1_HUMAN	CYFIP1	108	121	7.58	3P8C	100
KMVGDTVGA	E9PQN9_HUMAN	IFITM2	66	75	16.07	X	X
SFSQPKPS	H0YH87_HUMAN	ATXN2	764	772	11.62	X	X

One protein highlighted by the above considerations is fructose-bisphosphate aldolase A, or aldolase (UniProt ID: P04075), an enzyme involved in glycolysis and implicated in several other cellular pathways such as signalling and motility [200]. Aldolase has an available X-ray crystallography structure encompassing the entire protein and relatively high sequence coverage in the MS dataset (31%), with a single semi-tryptic peptide found to be significantly changing in the intermediate and heated conditions, as shown in Figure 4.11. Indeed, in the intermediate condition the peptides identified for this protein show very little change with LiP treatment, aside from this single site in which the semi-tryptic peptide signal had a log2 fold change of eight (p value < 0.00103). This peptide maps to a surface accessible, short loop and alpha-helical region, highlighted in pink on the PDB structure displayed in Figure 4.11. The signal for this protein indicates a single LiP cleavage site, with the rest of the protein remaining stable. In the heated condition, the peptide signal associated with this protein was largely similar, with the caveat that there were 2 additional sites in which tryptic peptide signal appear to decrease substantially with LiP treatment, although these changes were not found to be significant with the multiple testing correction applied.

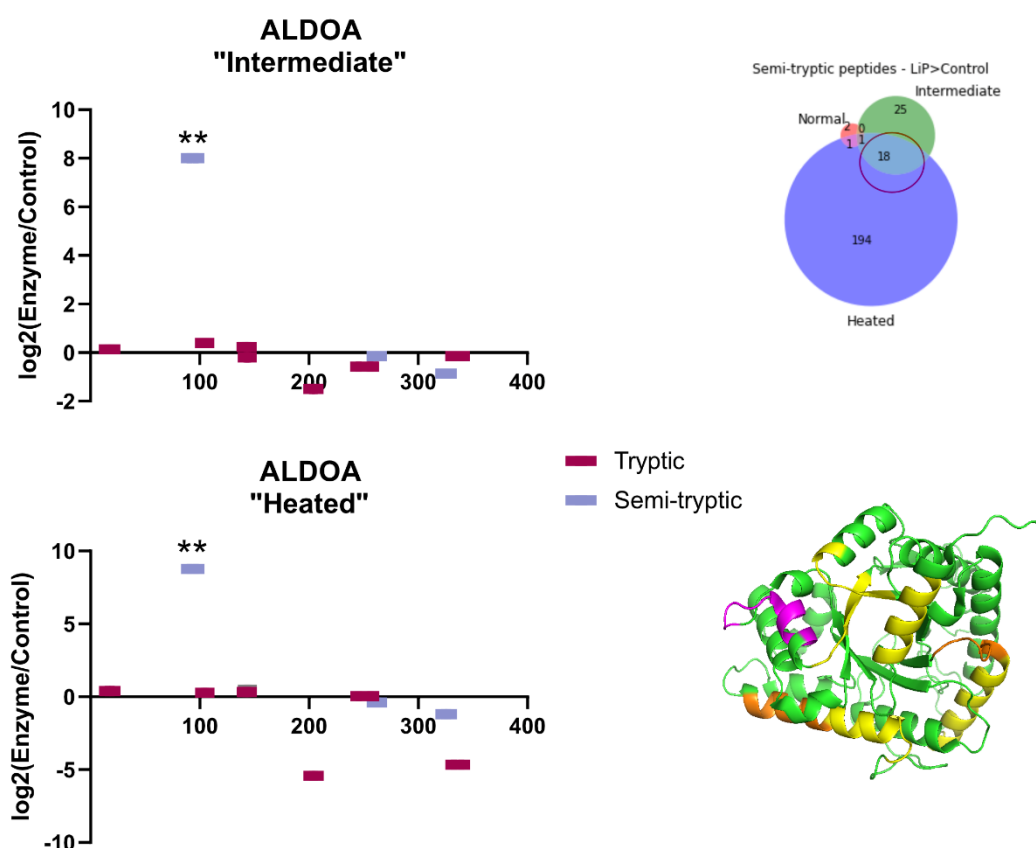


Figure 4.11: : Peptide signal for fructose-biphosphate aldolase A (UniProt ID: P04075) in the intermediate (top) and heated conditions (bottom). The y axis of the graphs represents the log2 fold change in peptide intensity between the control and LiP treated samples. This is plotted against the protein's sequence, so the peptide length is reflected. Tryptic peptides are displayed in pink and semi-tryptic peptides are displayed in lilac. Significantly changing peptides are highlighted with asterisks (** = Adjusted pvalue<0.001). Identified peptides are mapped onto a PDB structure (PDB Identifier = 1ALD [201]) of aldolase A, with tryptic peptides highlighted in yellow, semi-tryptic peptides highlighted in orange, and the significantly changing peptide region highlighted in pink. Venn diagram from Figure 4.8(c), with circle indicating the behavioural group this LiP peptide was identified in. A list of LiP peptides from this group is found in Table 4.2.

The 194 semi-tryptic peptides found to be significantly changing in only the heated condition (Figure 4.8(c), purple segment) were similarly investigated. A subset of these peptides is displayed in Table 4.3. An example protein from this behavioural subset was receptor of activated protein C kinase 1, or RACK (UniProt ID: P63244), which is thought to play a role in protein relocalisation and stabilisation and is known to interact with the ribosomal machinery [202]. Again, RACK has an available X-ray crystallography structure encompassing the entire protein and relatively good sequence coverage in the MS data, but in this instance none of the associated peptides are significantly changing in any condition apart from the heated condition. The peptide signal for the heated condition is shown in Figure 4.12 and displays a semi-tryptic peptide with significantly increased signal in the LiP treated samples and thus a suspected site of LiP cleavage. The structural features of this

cleavage site were investigated further using both Nickpred and through direct mapping onto the protein's structure (Figure 4.12). Unlike in some of the previous examples using Nickpred, this structure is not predicted but determined using X-ray crystallography (PDB Identifier = 4AOW) [203]. The LiP-susceptibility score for this cleavage site, as determined by Nickpred, was relatively low at 0.32 (scores range from 0 to 1, with 1 being the highest likelihood of cleavage). Furthermore, direct mapping to the protein structure revealed this cleavage site to be amid a β -strand, which has previously been determined to be the least LiP cleavable form of secondary structure [158]. Therefore, if this is a true LiP cleavage site, this implies that the protein has undergone heat-induced unfolding.

Table 4.3: Semi-tryptic peptides with significantly higher abundance in the LiP treated sample compared to control, in the heated condition only (Figure 4.8(c), purple segment). First 10 are displayed here. Displayed in the table is the semi-tryptic peptide, the corresponding parent protein (Uniprot ID and Gene Name) and the position of the peptide within the protein sequence. Start/End Pos reflect residue numbers, along the protein sequence from N to C terminus. Also detailed is the percentage sequence coverage of that parent protein from tryptic and semi-tryptic peptides in the LiP-MS data, PDB identifier name where a structure is available (not available is indicated by 'X') and the percentage of the protein structure that has been resolved in the PDB file. Highlighted in red is the protein investigated further in Figure 4.12.

Sequence	Protein	Gene Name	Start Pos	End Pos	LiP-MS Coverage (%)	PDB Identifier	PDB Coverage (%)
TNHIGHTGYLN	RACK1_HUMAN	RACK1	185	196	22.72	4AOW	100
FKTPELA	RBP2_HUMAN	RANBP2	2125	2132	4.62	4GAO	5
NVIGEPIDER	Q0QEN7_HUMAN	ATP5B	81	91	23.82	X	X
QAPYFGAEDYR	A0A0A0MSX9_HUMAN	IARS1	323	334	16.10	X	X
SVAELEELSGAK	A0A0A0MSX9_HUMAN	IARS1	482	494	16.10	X	X
SNETAAYK	STT3B_HUMAN	STT3B	639	647	7.99	6S7T	100
SARPSILPSK	H7C4C5_HUMAN	MAP4	170	180	22.92	X	X
SNSCFIR	FRG1_HUMAN	FRG1	151	158	10.85	6ZYM	100
AKYPDYEVT	Q6FIE5_HUMAN	PHP14	110	119	16.8	X	X
AEANNLAAAASAK	ATLA3_HUMAN	ATL3	347	360	8.13	5VGR	80

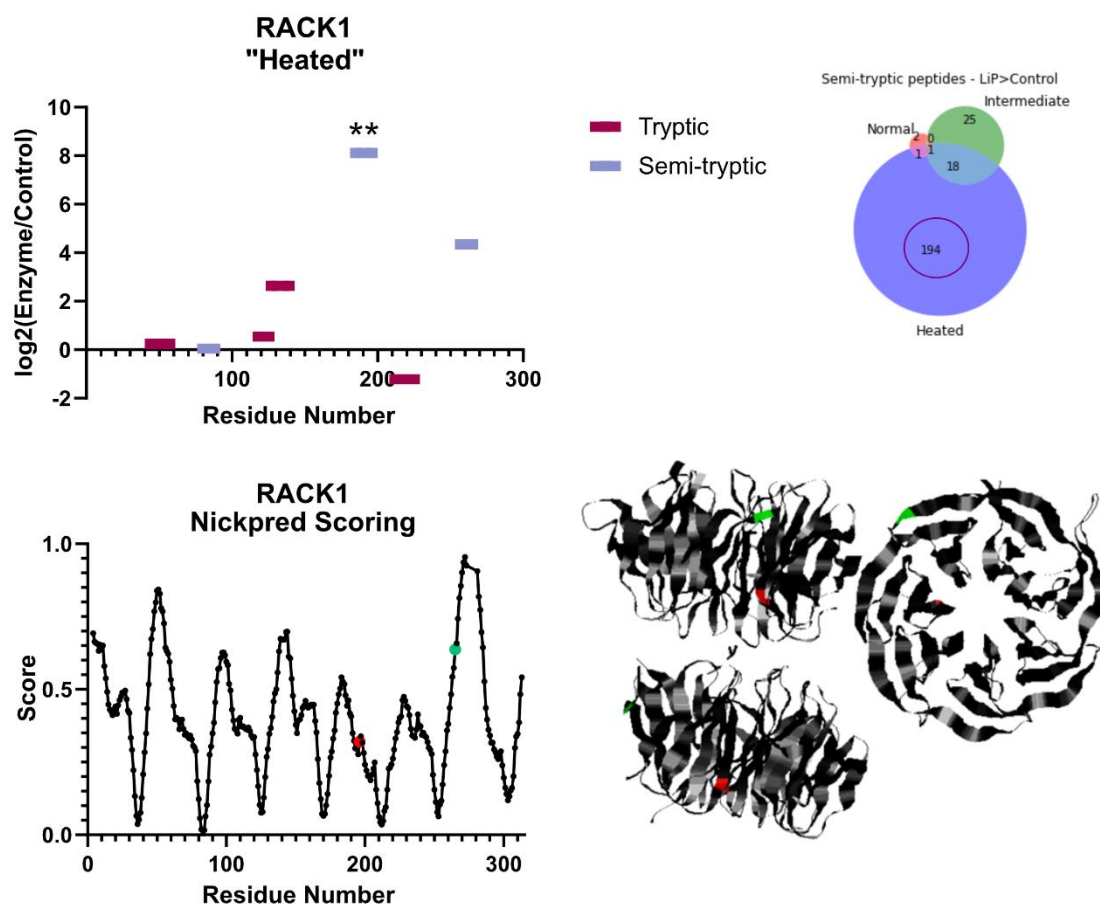


Figure 4.12: Top - Peptide signal for receptor of activated protein C kinase 1 (UniProt ID: P63244) in the heated condition. The y axis of the first graph represents the log2 fold change in peptide intensity between the control and LiP treated samples. This is plotted against the protein's sequence, so the peptide length is reflected. Tryptic peptides are displayed in pink and semi-tryptic peptides are displayed in lilac. Significantly changing peptides are highlighted with asterisks (** = Adjusted pvalue<0.01). Bottom left - Nickpred weighted norm scores for each residue of RACK. The cleavage site for the significantly changing semi-tryptic peptide is highlighted in red. The cleavage site for a second semi-tryptic peptide that is substantially increased in abundance upon LiP-treatment, but not significantly so, is highlighted in green. These peptides are mapped onto the PDB structure of RACK (PDB Identifier = 4AOW) and again highlighted in red and green. Venn diagram from Figure 4.8(c), with circle indicating the behavioural group this LiP peptide was identified in. A list of LiP peptides from this group is found in Table 4.3.

4.5 Further heat stress: 60 °C experiment

In the interests of completeness, a further heat stress experiment was also conducted, intended to further increase protein unfolding events and to determine if this was detectable by LiP-MS, and to examine whether the LiP events follow a linear pattern as temperature increases. In this instance, cell lysates were heated to 60 °C for 30 minutes, before either LiP treatment or control. This was compared to a no heat condition, where LiP treatment was performed immediately post-lysis, as above. In this instance, 60 °C was chosen following evidence by Leuenberger et al. [166] suggesting that this is the most common melting

temperature for the human proteome, and hence would be likely to induce substantial unfolding in many proteins of interest.

First, principal component analysis was performed, as shown in Figure 4.13(a). This revealed separation of the control and LiP treated samples in the no heat condition, in principal component 1. In the heated samples, the control clusters well with the control of the no heat condition, but the LiP treated samples seem to have more variability and are less well separated from their control.

The number of tryptic peptides identified in each sample is quantified in Figure 4.13(b), here defined as a non-zero reported intensity value in that replicate. Overall, little difference in the number of tryptic peptides identified in each experimental condition was observed, at least not at a significant level. This was in general agreement with the previous studies at lower temperatures, though previously a small reduction in tryptic peptide identifications was observed which might be expected to increase with heat treatment.

However, as before, evidence of LiP activity in the no heat condition was found via the log₂ fold change of reported tryptic peptide intensity in the LiP treated sample compared to the control sample, shown in blue in Figure 4.13(c). Here one can observe a skew of the log₂ distribution towards the left, indicating that for a number of peptides, the signal is reduced in the LiP treated samples. However, these changes mainly did not meet the significance threshold with the multiple testing correction applied, as shown in Figure 4.13(d). Further, analysis of the same log₂ distribution in the heated condition revealed very little difference between the reported peptide signal in control and LiP treated samples. This was further reflected in the differential abundance analysis shown in Figure 4.13(e), where the similarity of the samples can be seen in the higher p-values (less significant), when compared to the same analysis in the no heat condition.

Overall, these results suggested that LiP was unsuccessful in the heated condition. Subsequent reflection has led to the suggestion that this is due to flawed experimental design. A temperature of 60 °C was chosen based on evidence that this is the most common melting temperature of human proteins [166]. Therefore, the expectation was to observe mass unfolding of the proteome. However, the incubation time of 30 minutes may well have subsequently enabled misfolded proteins to aggregate, which would be confounding to the LiP experiment. Aggregation infers proteolytic resistance, the opposite of the greater proteolytic susceptibility inferred from protein unfolding. In this instance, one would expect to observe greater LiP activity in the no heated condition, than the heated condition, as was observed here. For these reasons, no further experiments were conducted at this higher temperature.

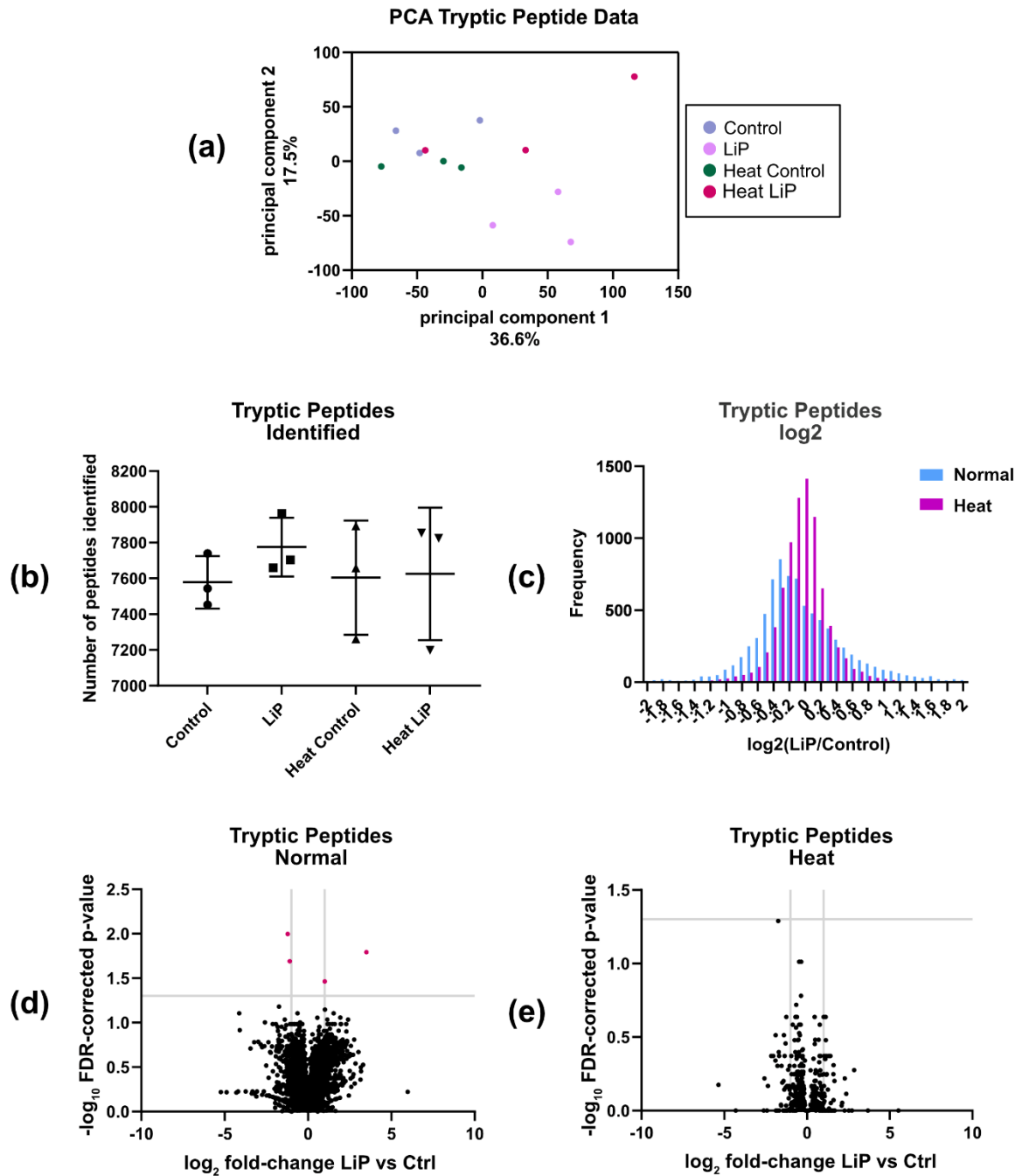


Figure 4.13: Analysis of 60 °C LiP-MS experiment. (a) Principal Component Analysis on tryptic peptide intensity data. The percentages shown on each axis represent the explained variance ratio. (b) Number of tryptic peptides identified in each run, defined here as having a non-zero reported intensity value. Error bars display the mean and standard deviation of the 3 replicates. (c) Tryptic peptide intensity log₂ distribution. For each tryptic peptide log₂ (LiP intensity/Control intensity) was calculated and organised into bins of 0.1 to show the distribution of values between -2 and 2. This is shown for the normal condition in blue and the 60 °C heated condition in purple. (d)+(e) Tryptic peptide abundance changes. Volcano plots showing the log₂ fold change of the abundance of tryptic peptides in the LiP treated samples vs the control samples, plotted against the respective -log₁₀ FDR corrected p-value, as determined by the Welch t-test and Benjamini Hochberg multiple testing correction. The significance threshold of 0.05 is highlighted by a grey line on the y axis and significantly changing peptides are highlighted in pink. Log₂ fold change of -1 and 1 are also highlighted by grey lines on the x-axis. (d) shows this analysis for the normal condition and (e) for the heated 60 °C condition.

4.6 Data processing

One possible cause of the presence of semi-tryptic peptide identifications, particularly in the control conditions, could be search engine error. Of particular concern was the apparent weighting towards semi-tryptic peptides shown in the Mascot peptide output, as shown in Figure 4.1. Of all peptides identified using semi-specific search parameters, 56% were found to be semi-tryptic. Endogenous protease activity, in-instrument cleavage and trypsin miscleavages may result in a baseline of semi-tryptic peptides in proteomics experiments, but none of these factors could account for a prevalence of semi-tryptic over tryptic. Another suggested cause was search engine error and indeed, despite their prevalence, semi-tryptic peptides were also identified with generally lower confidence, as far as can be inferred from the Mascot “score” metric, which is associated with the quality of the identification. This was determined through only considering peptides that fell into the top 20% of Mascot scores for the whole dataset. Such a cut-off saw the number of semi-tryptic peptides reduced by 93%, whereas the number of tryptic peptides was reduced by only 63%, demonstrating the reduced confidence placed in the semi-tryptic identifications in general.

To investigate whether this is a feature particular to Mascot, the MS RAW files were run again, with an alternative search engine, MaxQuant. Of all peptides identified by MaxQuant using semi-specific inclusive search parameters, just 17% were found to be semi-tryptic. Furthermore, when considering only peptides that fell into the top 20% of MaxQuant scores (a metric similarly associated with quality of identification [204]) for the whole dataset, the number of semi-tryptic peptides was reduced by 88% and the number of tryptic peptides by 79%. Although the scoring systems are not directly comparable, this suggests that the number of incorrectly identified semi-tryptic peptides may be reduced when using MaxQuant software. Consequently, the MaxQuant dataset was subsequently analysed to compare the derived results.

As in the Mascot processed dataset, few significantly changing LiP peptides were identified in the normal and intermediate conditions (a single tryptic peptide in each, and 1 and 5 semi-tryptic peptides respectively). However, LiP treatment was found to have a substantially larger effect in the heated condition, again matching the previous results. In the heated condition, 11 tryptic peptides were found to have significantly reduced signal with LiP treatment (Figure 4.14(a)) and this behaviour was also reflected in the distribution of fold changes (Appendix A.1, A.2), as shown previously. Moreover, 885 semi-tryptic peptides were found to have significantly increased signal with LiP treatment and 25 had significantly decreased signal (Figure 4.14(b)). Of the significantly increasing semi-tryptic peptides, 713 were only identified in the LiP treated sample (Figure 4.14(c)). Direct comparison with the

Mascot analysis showed that 33 of these peptides were shared, 161 significant only in Mascot and 852 significant only in MaxQuant.

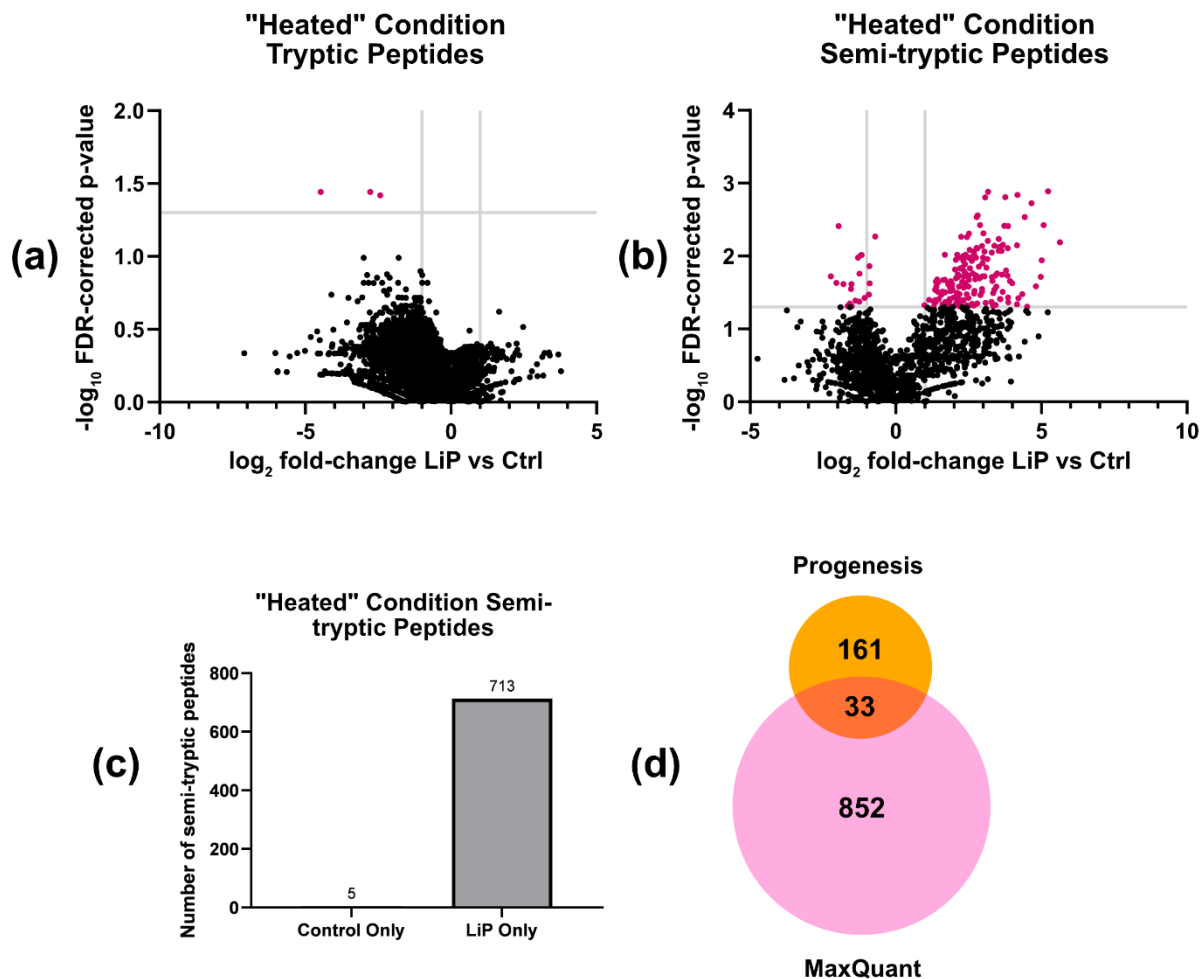


Figure 4.14: (a) Tryptic peptide abundance changes in the MaxQuant processed dataset, shown for the heated condition. Volcano plots showing the log₂ fold change of the abundance of tryptic peptides in the LiP treated samples vs the control samples, plotted against the respective $-\log_{10}$ FDR-corrected p-value. P-values were determined using Welch t-test and the Benjamini-Hochberg multiple testing correction. Significantly changing peptides are highlighted in pink. Here 11 tryptic peptides were found to be significantly changing, 8 of which were only identified in the control samples (not shown on plot). The grey lines on the x-axis indicate log₂ fold changes of -1 and 1, and the grey line on the y-axis indicates the corrected p-value significant cut-off of 0.05. (b) The same for semi-tryptic peptides. (c) Of the 885 semi-tryptic peptides with significantly increased signal with LiP treatment, 713 were identified only in the LiP sample (zero value in control). Of the 25 semi-tryptic peptides with significantly decreased signal with LiP treatment, 5 were identified only in the control sample (zero value in LiP). (d) Comparison between LiP peptides with the same behaviour identified in the Progenesis dataset (orange) and MaxQuant dataset (pink). 33 were shared.

The results of this analysis demonstrates that processing with MaxQuant did not change the overall conclusions of the study. The main takeaway – LiP activity increases with heat treatment and so unfolding – was preserved. However, despite fewer semi-tryptic peptides identified overall by MaxQuant, the number of LiP peptides identified in the heat condition

quadrupled. Further, a substantial proportion of these semi-tryptic peptides were unique to the LiP condition, which is more consistent with the original model of the presence of semi-tryptic peptides in LiP-MS data, and potentially indicative of a reduced level of noise in the control data under this processing condition.

For this reason, in the subsequent chapter MS data analysis will be conducted using MaxQuant.

4.7 Summary

The aim of this chapter was to establish the capacity of LiP-MS to detect changes to protein structure within complex cell lysates. This was tested through perturbing the proteome with heat treatment and then investigating the effect of such perturbation on the LiP-MS peptide signal.

Firstly, the success of LiP was verified. Principal component analysis revealed a clustering of the MS peptide signal based on LiP treatment, implying that LiP of the cell lysates has been successfully achieved and that this is reflected in the MS data. The effect of LiP on peptide signal was also investigated through assessing the log₂ fold change in reported peptide intensity between the control and LiP treated samples, for each of the heat conditions. In the tryptic dataset, the vast majority (an average of 81%) of peptides did not demonstrate a substantial change in signal with LiP treatment. If PK proteolysis is appropriately limited to flexible, accessible regions of protein structure, then the majority of the proteome will be unaffected. However, also evident was a strong subset of the tryptic dataset which demonstrated a decrease in signal with LiP (an average of 14% of all peptides). Likewise, in the semi-tryptic dataset, most peptides were not substantially effected by LiP treatment (an average of 68% of semi-tryptic peptides) but a sizeable subset demonstrated a gain in signal in the LiP samples (an average of 21%). Assuming that these semi-tryptic peptides are legitimately present, and indeed have a baseline presence in tryptic-only proteomics experiments, then these are precisely the results one would expect from successful LiP-MS.

Next, peptides with a significant change in abundance upon LiP treatment were determined. Exposure of the proteome to a temperature of 42 °C is sufficient to induce heat shock [196] and this can quickly lead to increased protein unfolding events. Increased protein unfolding should result in increased LiP enzyme activity. Therefore, if LiP is indeed capable of probing protein structure in human cell lysates, this would be identifiable by an increase in significantly changing LiP peptides in the heated condition. Whilst this was not reflected in the tryptic peptide analysis, the number of semi-tryptic peptides with a significant change in signal was consistently increased from the normal to heated experimental conditions, thus implying that protein unfolding can be detected through LiP-MS.

Significant, differential peptide abundance was determined using Welch's t-tests with the Benjamini-Hochberg multiple testing correction and a significance threshold of FDR-corrected p-value 0.05. Benjamini-Hochberg is often less conservative than the Bonferroni multiple testing correction and has been used previously in successful LiP-MS experiments [167]. Reducing the significance threshold (for example to an FDR-corrected p-value of 0.2) would have had the benefit of increasing the number of significantly changing tryptic peptides identified. However, the variance and uncertainty observed in semi-tryptic peptides calls for more robust statistical cut-offs to facilitate noise reduction.

Identified LiP peptides were then mapped onto known protein structures, in order to gain a more detailed understanding of the effect of heat treatment on the characteristics of LiP cleavage sites. For example, receptor of activated protein C kinase 1 (RACK) had a single LiP peptide mapping to a region that was only accessible in the heated condition. Further investigation using Nickpred and visualisation through direct peptide mapping onto the protein's crystal structure revealed that this cleavage site did not possess the structural characteristics associated with LiP. This could indicate that this protein has undergone heat-induced local unfolding.

Finally, the complications associated with semi-tryptic peptides have been highlighted throughout this chapter. It was hoped that the inclusion of technical replicates may ameliorate some of the issues caused by the presence of these peptides in the control samples, but analysis instead revealed that for each condition, some semi-tryptic peptides were actually present in significantly higher abundance in the control sample, when compared to the LiP treated sample. If this represents the level of noise expected in semi-tryptic peptide analysis, this will impair the confidence with which LiP cleavage sites can be identified. An alternative suggestion was that these peptides may represent regions of protein that have been excessively digested by PK. In this situation, as discussed in Section 3.1.3, parent peptides may be cleaved multiple times leading to loss in the identification of semi-tryptic peptides, either because the peptides generated are now non-tryptic, or too small for accurate MS identification.

A further possible explanation is that a proportion of these semi-tryptic peptides may have been misidentified. Of particular concern was the apparent weighting towards identification of semi-tryptic peptides over tryptic, shown by the Progenesis/Mascot output. This was investigated further by re-running the MS output files through an alternative search engine, MaxQuant. Switching to MaxQuant both reduced the prevalence of semi-tryptic peptides in the semi-specific search dataset, and increased their general identification confidence. Further, complete reprocessing of the LiP-MS data did not change the overall conclusions of

the study as the main takeaway – LiP activity increase with heat treatment and so unfolding – was preserved. However, despite fewer semi-tryptic peptides identified overall by MaxQuant, the number of semi-tryptic peptides found to be significantly changing with heat treatment increased substantially, with a high proportion found to be unique to the LiP condition. This is more in consistent with the original LiP-MS model, which presumes that the majority of semi-tryptic peptides identified in the dataset are created by the activity of the LiP protease.

Chapter Five: LiP-MS can provide structural insights into the cellular response to stress

In the previous chapter, LiP-MS was used as a probe of heat-induced protein unfolding in an immortalized hMSC line. However, previous LiP-MS experiments, reviewed in detail in Section 1.6.2, have demonstrated that the capacity of the technique to identify post-translational alterations to protein structure is not limited to protein unfolding. Indeed, the latest studies have shown the efficacy of LiP-MS in providing insight into many functionally critical processes [167, 169, 205].

For example, LiP-MS has been successfully utilized to probe the proteostatic response to thermal and oxidative stress in yeast [167]. This analysis highlighted many hundreds of proteins with potential stress-induced structural alterations, and functional enrichment analysis of these proteins revealed an over-representation of known stress responses. Moreover, LiP peptides (those with differential enzyme accessibility pre and post-stress, identifiable by the LiP “fingerprint” discussed above) demonstrated a strong correlation with sites of functional importance. Examples were provided of LiP peptides mapping to protein active sites, phosphosites, and areas of known conformational change. This information was then used to gain functional insight into the yeast stress response, for example by inferring the activation of molecular chaperones.

In addition, LiP-MS has also recently been used to assess protein-metabolite interactions in *E.coli*, by detecting conformatypic peptides observed following the addition of a known metabolite [169]. The LiP peptides identified in this study mapped to known protein-metabolite interaction sites, identified novel interactions and also indicated metabolite-induced higher order structural changes such as protein complex formation or dissociation.

Therefore, the next aim of this thesis, and the subject of this chapter, was to see whether LiP-MS can provide a similar global structural readout of the response to stress in hMSCs. The chosen stress was heat shock, induced through incubating the cells at 42 °C for 2 hours, with additional stress caused by treatment with a chaperone inhibitor (for details see Section 2.13-14). LiP-MS was performed on the perturbed and unperturbed cells in order to identify perturbation-induced protein structural changes. In this instance, two different LiP concentrations (E:S 1:2000, 1:500) were used, as well as the usual trypsin-only control, for each condition. The full experiment design is shown in Figure 5.1.

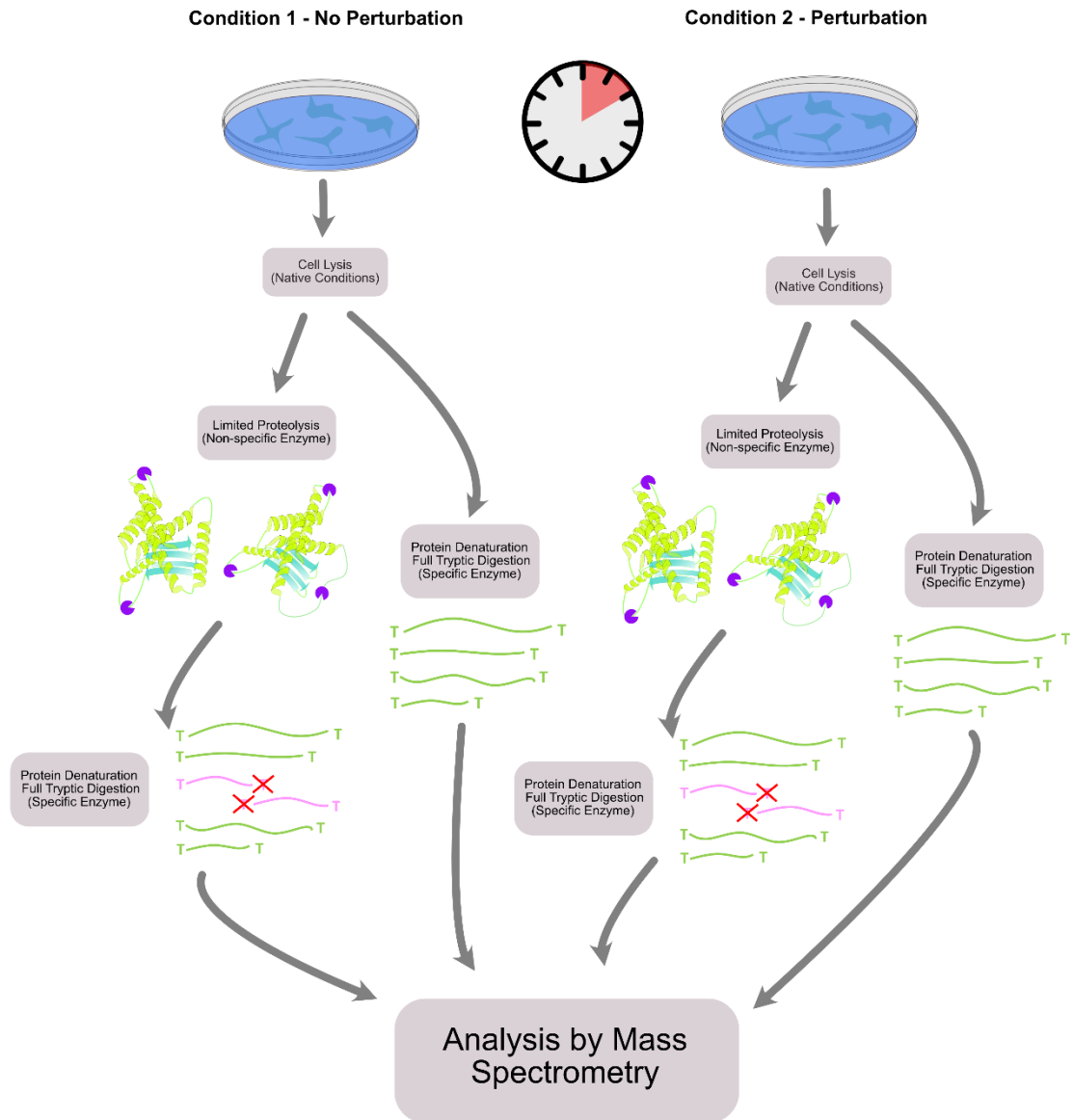


Figure 5.1: LiP-MS workflow. In condition 1, hMSCs were incubated at 37°C. In condition 2, hMSCs were incubated for 2 hours at 42°C and treated with a chaperone inhibitor (Section 2.13). Immediately following heat shock, flasks were removed from both incubators and cells were removed (Section 2.7) and lysed (Section 2.9.3). For each condition, 4 samples were incubated with PK at an E:S ratio of 1:2000, for 2 minutes followed by quenching of the enzymatic reaction by boiling. Once cooled, proteins were then denatured, digested with trypsin and subjected to analysis by MS (Section 2.15). 4 further samples for each conditions were subject to the same steps, but with PK ratio of 1:500. 4 samples for each condition were not treated with PK, but analysed by MS following a trypsin-only digest.

Previous work in the Swift lab has highlighted the need for a structural proteomics approach to the study of this system. In this work, heat shock was applied to donor-matched early (EP) and late passage (senescent) (LP) hMSCs and the stress response characterized using a combination of transcriptomics, proteomics and immunofluorescence imaging [10]. This analysis revealed a good correlation between the transcriptional stress response in EP and

LP hMSCs, but demonstrated a dampening of this response with senescence at the protein level, evidenced from changes in protein expression in key stress response pathways. This indicated that damage to the proteostasis network in senescent cells is predominantly post-transcriptional. Incorporation of these findings into a simple mathematical model of four key stress responders (HSP70, CHIP, HSF1, misfolded protein burden) led to a prediction of a greater burden of misfolded protein on senescent cells. However, testing of this prediction was not performed in this study and would benefit from an alternative approach that is also able to assess changes in protein structure.

In addition to the loss of proteostatic competence in chaperone modules and in the protein synthetic machinery (e.g. ribosomes), a significant lag between transcription and translation was also shown in all cells, suggesting that a complete understanding of proteostatic defects in the early stress response cannot be captured by measures of protein abundance alone. Indeed, perhaps other regulators of proteostasis such as protein relocation and post-translational protein remodelling – protein-protein interactions, PTMs, protein-small molecule interactions – are quicker to enact and so form the crux of the early stress response. The techniques used in this study were not capable of providing insight into these features. Instead, a structural proteomics approach is required to complement and enhance these findings.

Furthermore, there is a specific need to understand proteostasis in hMSCs as they are often used in regenerative medicine, with over 350 clinical trials conducted between 2004 and 2018 that involved the use of bone marrow-derived MSCs. Due to the small numbers attainable *in vivo* and the high dosage that is often required for hMSC-based therapeutics, expansion of hMSCs *in vitro* is very often required [15]. A common use of hMSCs in regenerative medicine is their transplantation to sites of damage, such as the delivery of bone marrow derived hMSCs to the brain in rat and mouse models of Parkinson's, Alzheimer's and Huntington's Disease [11-13]. MSC transplantation carries the risk of inflammation, damage to tissue and potentially infection, all of which are triggers of the heat shock response [14]. Therefore, an improved understanding of the effects of cell culture, *in vitro* passaging and stress in hMSCs will be beneficial towards the development of effective stem cell based therapies [16].

The LiP-MS experiment detailed in Chapter 4 utilized a heat shock applied to cell lysates, to perturb protein structure. Whilst this can provide interesting insights into protein stability, it is also interesting to ask whether LiP-MS can be used to provide functional insight into the **cellular** response to heat shock. This is a greater test of the sensitivity of LiP-MS, as the LiP step must be performed post lysis. Therefore, if the lysis method used (Triton X-100, Section

2.9.3) is denaturing, this could obscure stress-induced protein structural changes and decrease the capacity of LiP-MS to differentiate between stress conditions.

Therefore, the experiment detailed above was designed to answer the question – can LiP-MS provide insight into the cellular response to stress? This will be the subject of Chapter 5.

5.1 Stress-induced changes to protein abundance

Firstly, the control samples from each condition were used to determine whether the stress had induced any significant changes to protein abundance levels. The data was processed using Protti's inbuilt functions, including normalization and Welch's T-test with the Benjamini Hochberg multiple testing corrections (Section 2.18). This analysis revealed little to no significant stress-induced change in protein abundance, with only a single protein observed to change significantly at a FDR cutoff of 0.05, as shown in Figure 5.2. As mentioned above, previous data produced in the Swift lab suggests that sampling at this time point, e.g immediately post 2 hours heat shock, may be too early to see a translational response and consequently a significant change in most of the proteome [10]. This work was carried out by Jack Llewellyn and also involved treating hMSCs with heat shock for 2 hours at 42 °C. A notable difference is that this work used 4 biological replicates (hMSCs from 4 different donors), whereas the work carried out here uses 4 technical replicates (same donor, 4 repeats). In this case, the variance attributable to individuals is not present and does not need to be accounted for, simplifying data analysis.

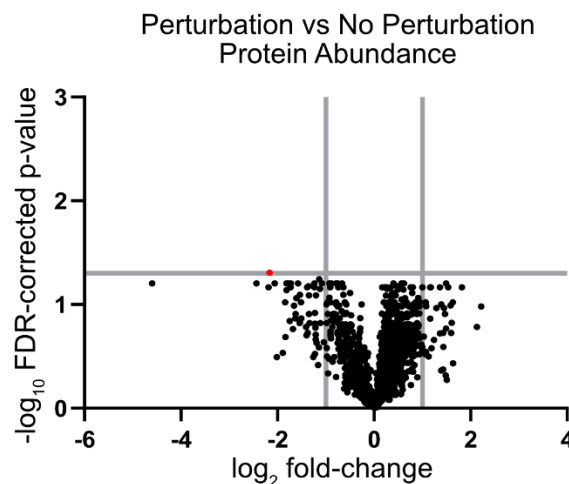


Figure 5.2: Volcano plot showing the \log_2 fold change of the abundance of proteins in the Perturbation condition vs the No Perturbation condition, plotted against the respected $-\log_{10}$ FDR-corrected p-value. P-values were determined using Welch t-test and the Benjamini-Hochberg multiple testing correction. Significantly changing proteins are highlighted in red, using a significance threshold of 0.05 (shown by grey line y axis). The grey lines on the x-axis represent \log_2 fold changes of -1 and 1 respectively.

In this previous work, the transcriptomic and proteomic response of hMSCs to heat stress were measured. Firstly, the similarity between the transcriptional and translational stress

response was investigated, through measuring protein and RNA levels pre and post 2 hours heat stress. The fold change data (unstressed vs stressed) produced from RNA-seq was compared to the protein fold change data from MS analysis and the corresponding Pearson product-moment correlation coefficient calculated. Briefly, this acts as a measure of the association between 2 variables. A correlation of $R^2 = 0.0048$ was found between the log2 fold change data (unstressed vs stressed) of all transcripts and the same data for all proteins, suggesting very little correlation between the transcriptional and translation response over the 2 hour time period (Figure 1.5). This is consistent with a faster or slower response, as well as a totally uncoupled system.

To get a fuller view, specific proteins were selected for investigation at further time points. One such protein was HSPA1A (stress-inducible HSP70), this time utilizing IF to measure antibody tagged HSPA1A protein and RT-qPCR to measure transcript abundance before, during and after the administration of heat shock. This data showed that the log2 fold change of HSPA1A transcript relative to its level before stress reaches its peak very quickly post heat shock (Figure 1.6). However, the protein level response is markedly slower, not reaching its maximum level until 4 hours post heat shock.

Taken together, this data suggests that sampling immediately post-stress is too early to see a significant response in protein abundance, due to a time delay between transcription and translation producing their gene products. However, structural protein changes are potentially quicker to enact, thus inviting investigation into whether structural alterations play a crucial role in the early cellular response to stress.

5.2 Stress-induced changes to protein structure

Given that the data is consistent with only a modest proteome change induced by this stress, in order to extract structural information, the full LiP-MS dataset can be applied (including the LiP treated sample, not just the controls). Firstly, a PCA was performed on the protein LFQ intensity data in order to get an overview of variation in the samples. The results, shown in Figure 5.3, indicate that the biggest source of variation is LiP condition, which separates the samples in the first principal component from control through to highest LiP enzyme (PK) concentration. There is apparently comparatively little variation induced by stress, at least in the first two components. However, this aligns with expectation as the extra limited proteolytic digestion step should significantly change the peptides (and by extension, to a lesser degree, the proteins) identified comparative to a trypsin-only digest. Therefore, it is encouraging that the samples group well by LiP condition, particularly as the strongest LiP condition (1:500) is the most different to the control, as one would expect.

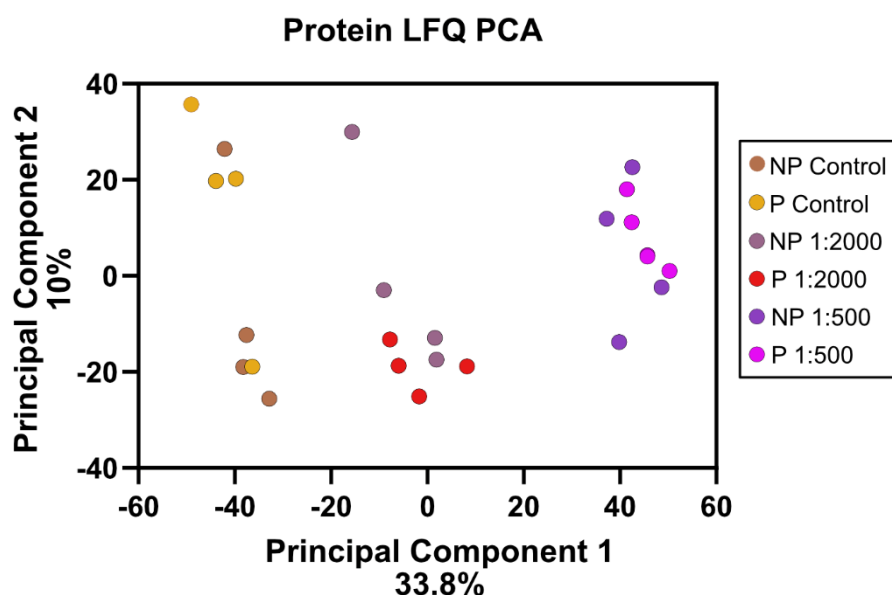


Figure 5.3: Principal Component Analysis of Protein LFQs from complete LiP-MS dataset. NP = No Perturbation. P = Perturbation.

Another method to assess the effects of the LiP processing on the samples is to simply quantify the number of peptides that have been identified in each sample. This is shown in Figure 5.4 for both tryptic and semi-tryptic peptides and also shows a significant separation of the number of peptides identified by LiP treatment. The manner of this separation is also consistent with a model where increased LiP would remove tryptic peptides from the pool, and increase semi-tryptic ones, with a negative correlation demonstrated between enzyme concentration and number of tryptic peptides identified (Figure 5.4(a)). Equally encouragingly, enzyme concentration and the number of the semi-tryptic peptides identified are positively correlated, though once again a sizeable number of semi-tryptic peptides are identified in each control condition (Figure 5.4(b)). Both measures also show an encouraging consistency amongst replicates, providing further evidence that LiP sites are predominantly determined by structural features in the native/near-native protein folds under investigation and potentially less by random cleavage.

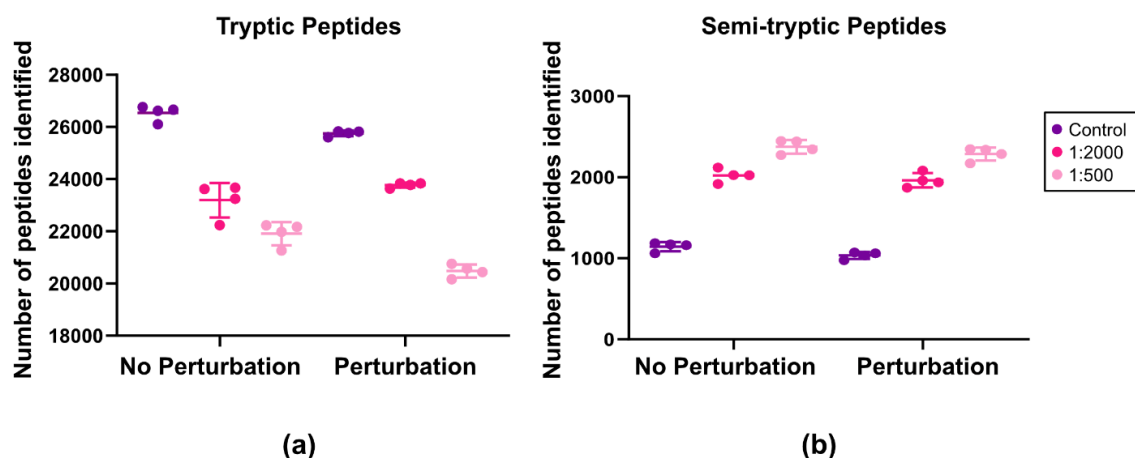


Figure 5.4: Number of peptides identified in each sample. Tryptic (a) and semi-tryptic (b) peptides identified in each sample (non-zero reported intensity). Error bars indicate the mean and standard deviation of the 4 replicates per condition.

5.2.1 LiP-induced alterations to protein signal

In order to investigate the hypothesis that these decreases in tryptic peptides and increases in semi-tryptic peptides are indeed associated with the structural properties of their parent proteins, few example cases were examined in more detail. Each of the below examples display only the unstressed data, in order to provide a baseline LiP readout. In line with expectation, the effect of LiP treatment on tryptic peptide signal varied depending on the protein being considered. For example, twinfilin (Uniprot ID: Q12792), shown in Figure 5.5 (top), is a protein primarily associated with actin binding. The tryptic peptides mapping to twinfilin in the unstressed condition, of which there were several, do not demonstrate any significant change in signal with LiP treatment. Therefore, this provides an example of a protein that appears to be largely resistant to proteolysis. In fulfilling its role as an actin binding protein, it is often held tightly in complex, as demonstrated by the crystal structure of the actin filament uncapping complex [206, 207].

The second protein shown in Figure 5.5 (bottom) is serine-threonine kinase receptor-associated protein “Strap” (Uniprot ID: Q9Y3F4). Strap has been shown to regulate the translation of type I collagen mRNAs [208] and as an inhibitor of transforming growth factor beta (TGF β) signalling [209]. The tryptic peptide signal of this protein was mainly unchanged with LiP treatment, but one tryptic peptide did demonstrate a significant loss of signal with the 1:500 LiP treatment. Mapping of this LiP peptide onto the AlphaFold PDB structure for Strap [188, 189] revealed that this region of protein is both surface accessible and contains a loop region. Whilst this structure is a prediction, the regions in question were all assigned “Very High” in AlphaFold’s inbuilt per-residue confidence score metric. Therefore, the tryptic peptide signal for this protein indicates a single LiP cleavage site, with the rest of the protein

remaining stable.

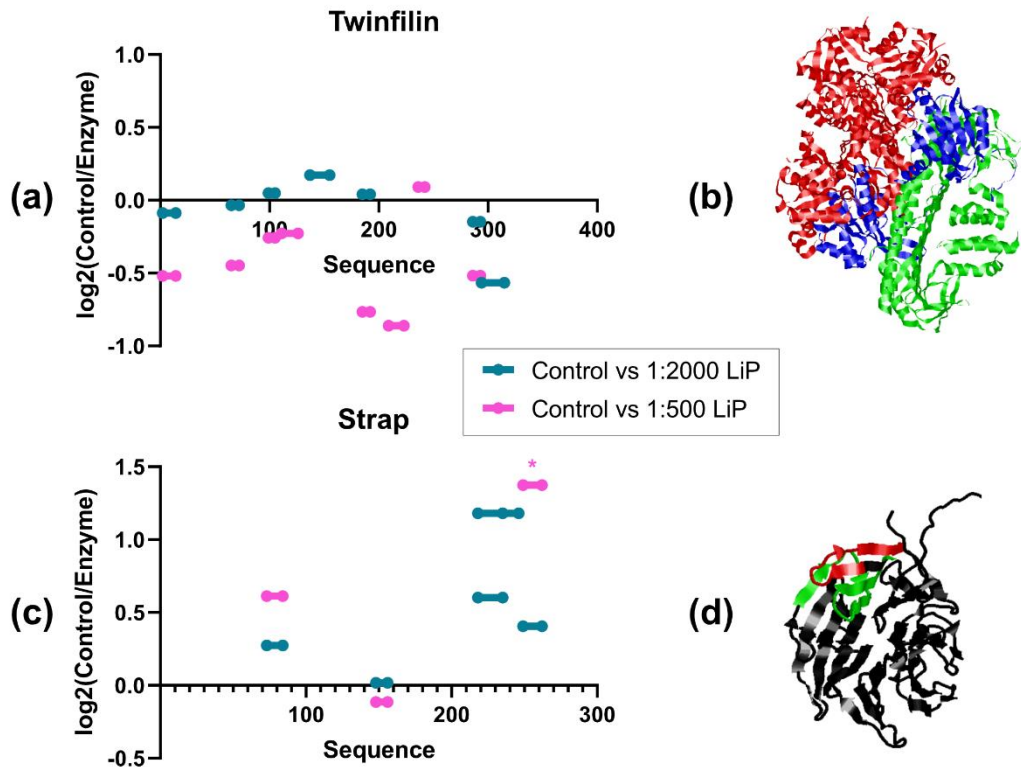


Figure 5.5: Tryptic peptide signal shown for 2 proteins in the unstressed experimental condition. The y axis represents the log2 fold change in peptide intensity between the control and LiP treated samples. This is plotted against the protein's sequence, so the peptide length is reflected. Significantly changing peptides are highlighted with an asterisk*, indicating an adjusted p-value of <0.05. (a) Twinfilin-1 (UniProt ID: Q12792). (b) X-ray crystallography structure of the actin filament uncapping complex [206, 207]. Here actin is highlighted in red, twinfilin in blue and the F-actin capping protein subunits in green. (c) Serine-threonine kinase receptor-associated protein (Strap) (UniProt ID: Q9Y3F4). (d) AlphaFold PDB structure for Strap. Here the significantly changing tryptic peptide (1:500 sample) is highlighted in red, and the tryptic peptide most changing (not significant) in the 1:2000 sample is highlighted in green. Displayed is the residues with "Very High" confidence score ((pLDDT > 90)).

Finally, the tryptic peptide signal of myotrophin (UniProt ID: P58546), a protein thought to play a role in the regulation of nuclear factor kappa b (NFkB) transcription factor activity [210], is considered. In this case, a signal was detected for seven tryptic peptides in the control sample, whereas this signal was lost entirely for 5/7 peptides in the 1:2000 LiP treated samples, with the remaining two peptides exhibiting significantly reduced signal (Figure 5.6(a)). In the 1:500 LiP treated sample, tryptic peptide signal was lost entirely for 6/7 peptides and reduced to near zero in the only remaining peptide. This protein thus provides an example of complete destabilisation following LiP treatment, the hypothesis here being that, when subject to LiP, this protein loses its structure and is subsequently digested extensively by the LiP enzyme [156]. This is further evidenced in the protein signal. Figure

5.6(c) displays the MaxQuant LFQ protein intensity data for myotrophin in each MS run and shows the complete loss of signal for this protein in the 1:500 LiP treated runs.

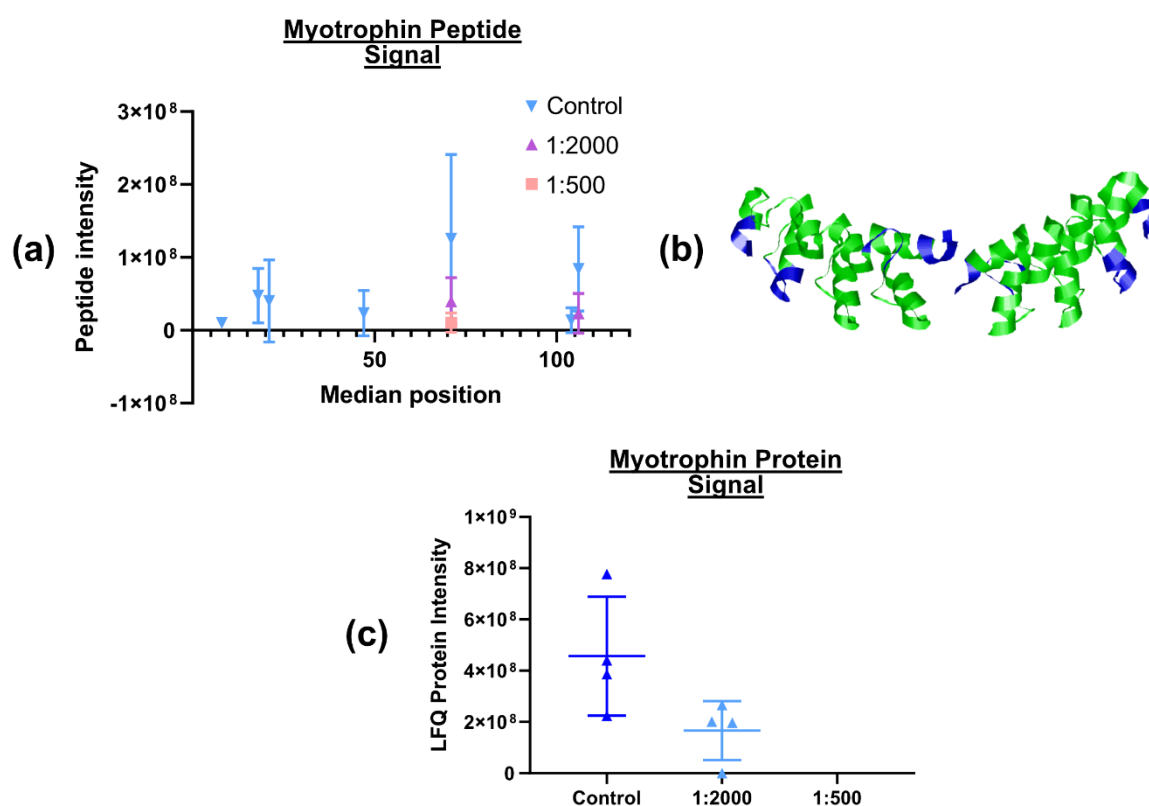


Figure 5.6: MS signal for myotrophin (UniProt ID: P58546). (a) The intensity of each tryptic peptide identified for myotrophin is displayed relative to its position along the protein's sequence. Error bars represent the mean and standard deviation of the 4 replicates for each LiP condition. (b) PDB structure 7DF7 for myotrophin [211]. Residues for which a tryptic peptide was identified in the control condition are highlighted in green, the overall sequence is in blue. (c) Protein intensity readout for myotrophin in each of the MS runs. Error bars again represent the mean and standard deviation of the 4 technical replicates for each LiP condition.

5.2.2 Identification of LiP peptides

Thus far the efficacy of LiP to generate detectable features in the MS data has been demonstrated. What is yet to be shown is whether any stress-induced changes in protein structure that are observed at the peptide level can subsequently be detected. To examine this, the peptide data output from MaxQuant was processed through LiP-MS specialist software Protti (Section 2.18), in order to identify significant changes in peptide abundance. Following on from results in previous chapters, and the apparent high variance observed in semi-tryptic data, further data analysis was restricted to the set of tryptic peptides, due to the higher confidence in their identification and their increased reliability as a measure of LiP cleavage. Peptides pertaining to LiP cleavage sites were identified as those with a significant

difference in abundance between the control and LiP treated samples. Such peptides are highlighted in pink in Figure 5.7.

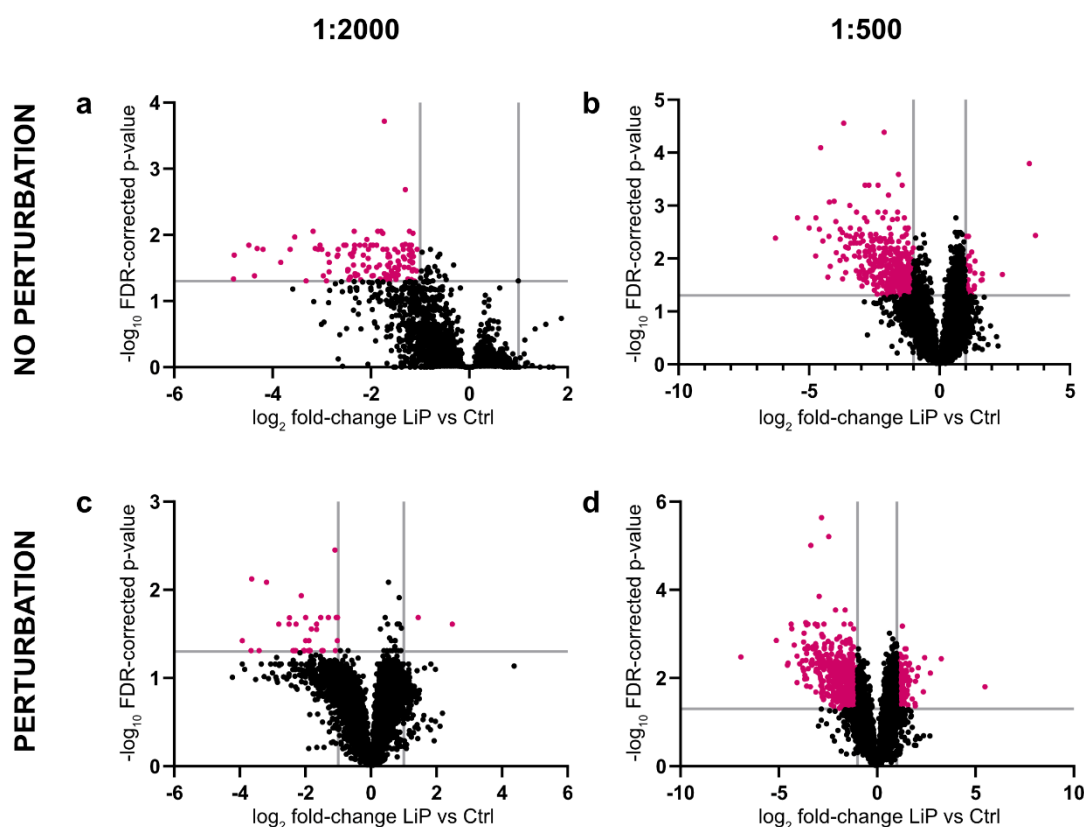


Figure 5.7: Peptide abundance changes. Volcano plots showing the \log_2 fold change of the abundance of tryptic peptides in the LiP treated samples vs the control samples, plotted against the respective $-\log_{10}$ FDR-corrected p-value. P-values were determined using Welch t-test and the Benjamini-Hochberg multiple testing correction. Significantly changing peptides are highlighted in pink. (a)+(b) No Perturbation samples, 1:2000 (Enzyme:Substrate) vs control, 1:500 vs control respectively. (c)+(d) Perturbed samples, 1:2000 vs control, 1:500 vs control respectively.

Typically, LiP cleavage is expected to result in a **decrease** in the abundance of a tryptic peptide comparative to the control (here the left hand side of the volcano plots displayed in Figure 5.7). The presence of such LiP peptides increases with enzyme concentration in both biological conditions, increasing from 124 (Figure 5.7(a)) to 447 (Figure 5.7(b)) in the No Perturbation condition and from 35 (Figure 5.7(c)) to 651 (Figure 5.7(d)) in the Perturbation condition. Not only does the number of LiP peptides increase with enzyme concentration but the scale of such change also appears to be increased, as is demonstrated by the wider spread of data in Figure 5.7. This provides further support for the hypothesis that the changes observed in these peptides are the results of LiP cleavage.

Next, LiP peptide changes due to perturbation were considered. The number of LiP peptides that are consistent or different before and after stress are displayed using Venn diagrams in Figure 5.8. For example, Figure 5.8(a) shows there are 249 LiP peptides in common to the

No Perturbation and Perturbation conditions. Following the logic of LiP cleavage sites, these peptides represent regions of protein structure that are presumably always enzyme accessible, and that this access is not further affected stress.

Figure 5.8(a) also highlights the 198 LiP peptides only found to be changing significantly in the No Perturbation condition, suggesting these peptides map to regions of protein structure that are only enzyme accessible **before** stress. Similarly, the 402 significant LiP-induced peptide changes only present in the Perturbation condition are representative of protein regions that are only enzyme accessible **after** stress. Taken together, these 600 peptides indicate that their parent proteins that have undergone some structural alteration in response to stress, which alters the accessibility of local peptide bonds to limited proteolysis.

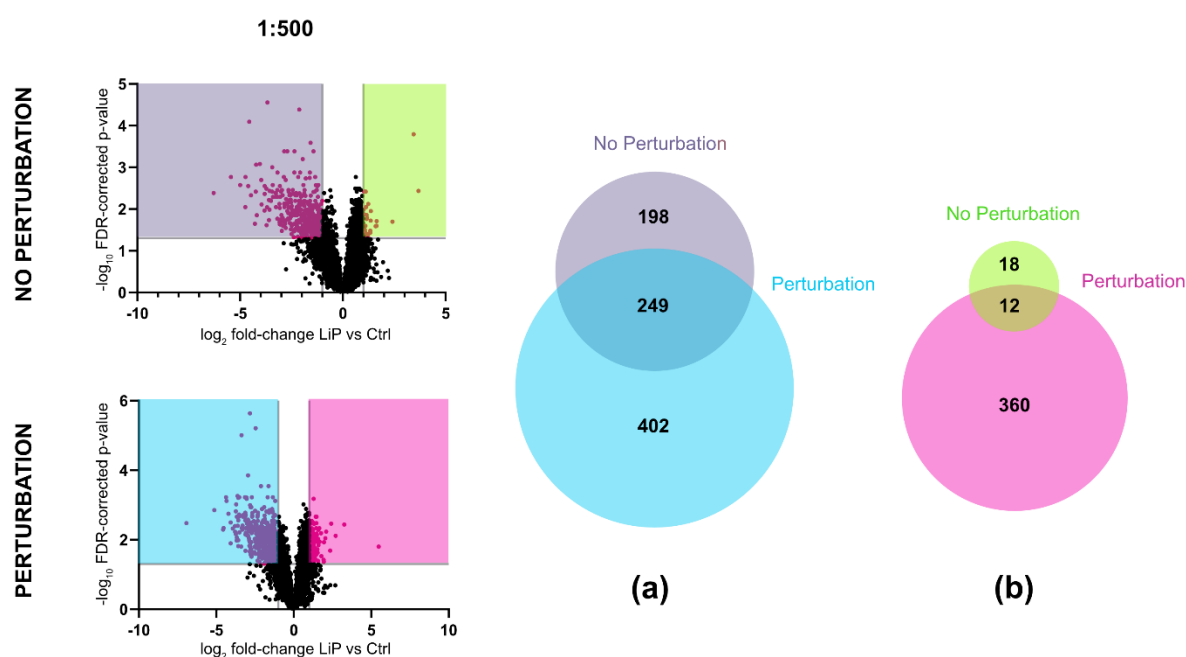


Figure 5.8: Volcano plots of peptide abundance changes for the 1:500 LiP condition, as in Figure 5.7 (a). Venn diagram of the overlap between significantly changing peptides in the No Perturbation and Perturbation conditions in which the abundance of said peptide is higher in the control than lip treated sample. (b) Venn diagram of the overlap between significantly changing peptides in the No Perturbation and Perturbation conditions in which the abundance of said peptide is higher in the LiP treated sample than the control. Colours correspond to the highlighted sections of the volcano plots. There were no instances in which the direction of abundance change was altered between the No Perturbation and Perturbation conditions.

Another behaviour highlighted in this analysis is that of significantly higher tryptic peptide abundance in the 1:500 LiP treated sample than in the control. This result was initially concerning, as it opposes the classical expectation that defines the peptide fingerprint of LiP cleavage sites. However, a working hypothesis consistent with this observation is that this may be caused by protein aggregation, or at least protein site-inaccessibility to trypsin. During the MS sample preparation protocol, steps are taken to ensure the effective digestion of proteins into peptides by trypsin (denaturation, reduction and alkylation prior to application

of trypsin, Section 2.15.1). However, some tightly folded proteins, and indeed some regions of proteins, are more resistant to digestion, particularly if they form aggregates. In such cases, digestion with trypsin alone may not be effective in producing MS-amenable peptides. However, it has been demonstrated that the inclusion of a second enzyme (often Lys-C/Trypsin mix) to the protocol can significantly increase digestion efficiency and so the number of identifiable tryptic peptides [212]. It is possible that a similar effect could occur here. Should the addition of PK result in extra cleavage(s) of trypsin-resistant proteins at stability-critical sites, this could lead to increased accessibility for trypsin in the subsequent digestion step. Under these conditions, said tryptic peptides would be more predominant in the double digestion (LiP) treated samples than the control. Figure 5.8(b) shows that the presence of significantly changing peptides with this behaviour is increased twentyfold after stress. This supports the hypothesis, as heat stress is known to induce protein aggregation and thus will likely increase the prevalence of proteolytic-resistance proteins.

5.2.3 Structural characteristics of LiP peptides

Nickpred

The structural characteristics of these significantly changing peptides were then further investigated using Nickpred, a software tool that predicted limited proteolytic sites [170], and through secondary structure mapping. A snapshot of the full results (1239 LiP peptides total) is given here in Table 5.1.

Briefly, Nickpred assesses several conformational parameters from a given protein structure in order to predict the susceptibility of each residue of that protein to cleavage by a given protease, under the assumption that said protein is occupying its native or near-native state. This susceptibility is outputted as a “score” with a value between 0 and 1, with 1 representing the highest likelihood of cleavage. A fuller explanation of the Nickpred software is provided in Section 3.1.2.

Here, the AlphaFold predicted structure for each of the potentially structurally-altered proteins highlighted in Figure 5.8 was downloaded and run through Nickpred. From this, Nickpred scores for every fully tryptic peptide were calculated and those for the significantly changing peptides were then extracted. Scores were obtained by averaging scores for each protein residue over the length of the peptide, to give a single readout. Nickpred provides four scores, which are averages of normalized feature scores, and which are either weighted/unweighted, or further normalized or raw.

Table 5.1: Structural characteristics of significantly changing peptides. For each peptide the parent protein, the experimental condition in which this peptide was found to be significantly changed, and the direction of said change is given. Further, each peptide was run through the Nickpred software and four scores assigned. These scores are calculated based on the individual residues and then averaged along the length of the peptide. Score 1 = weighted norm score, Score 2

= weighted score, Score 3 = Norm score, Score 4 = Prediction score. Each score is between 0 and 1, with 1 representing highest protease susceptibility. Finally, SS represents Secondary Structure information, with 0 = strand, 0.5 = helix, 1 = loop. This is again based on each residue, and therefore peptides mapping to multiple secondary structure features will have more than one score. This table is a small excerpt of the complete list of 1239 significantly changing peptides, which was used for the analysis detailed in the rest of this chapter.

Sequence	Protein	Behaviour	Score 1	Score 2	Score 3	Score 4	SS
VEVTEFEDIK	sp Q01105 SET_HUMAN	Perturbation Control>LiP	0.1	0.45	0.115	0.357	['0', '1']
DIHDDQDYLSLGLK	sp O75955 FLOT1_HUMAN	Perturbation Control>LiP	0.643	0.578	0.221	0.482	['0', '0.5', '1']
FFVGGNWK	sp P60174 TPIS_HUMAN	Perturbation Control>LiP	0.329	0.579	0.34	0.496	['0', '1']
SADEVLTGVK	sp Q9Y5X3 SNX5_HUMAN	Perturbation Control>LiP	0.507	0.548	0.679	0.599	['0.5', '1']
AICTEAGLMALR	sp P62191 PRS4_HUMAN	Perturbation Control>LiP	0.308	0.501	0.36	0.505	['0.5']
WIYEDVER	sp O43852 CALU_HUMAN	Perturbation Control>LiP	0.344	0.429	0.187	0.343	['0.5']
NAGNCLSPAVIVGLLK	tr A0A286YF22 A0A286YF22_HUMAN	Perturbation Control>LiP	0.241	0.468	0.284	0.435	['1']
TAMNVNEIFMAIAK	sp P51148 RAB5C_HUMAN	No Perturbation Control>LiP	0.323	0.519	0.228	0.431	['0.5', '1']
FYGAEIVSALDYHSEK	sp P31749 AKT1_HUMAN	No Perturbation Control>LiP	0.141	0.422	0.104	0.358	['0.5']
LPVDILGR	sp P42566 EPS15_HUMAN	No Perturbation Control>LiP	0.515	0.556	0.454	0.535	['0.5', '1']
YYSFFDLNPK	sp Q96AC1 FERM2_HUMAN	No Perturbation Control>LiP	0.574	0.579	0.409	0.51	['1']

This information can be used to investigate whether each of the different LiP peptide behaviours identified here have any defining structural characteristics. Figure 5.9 displays the Nickpred weighted norm score for each LiP peptide, separated into the different behavioural groups. The same analysis was carried out for all other peptides identified in this MS experiment, in order to provide a baseline for comparison.

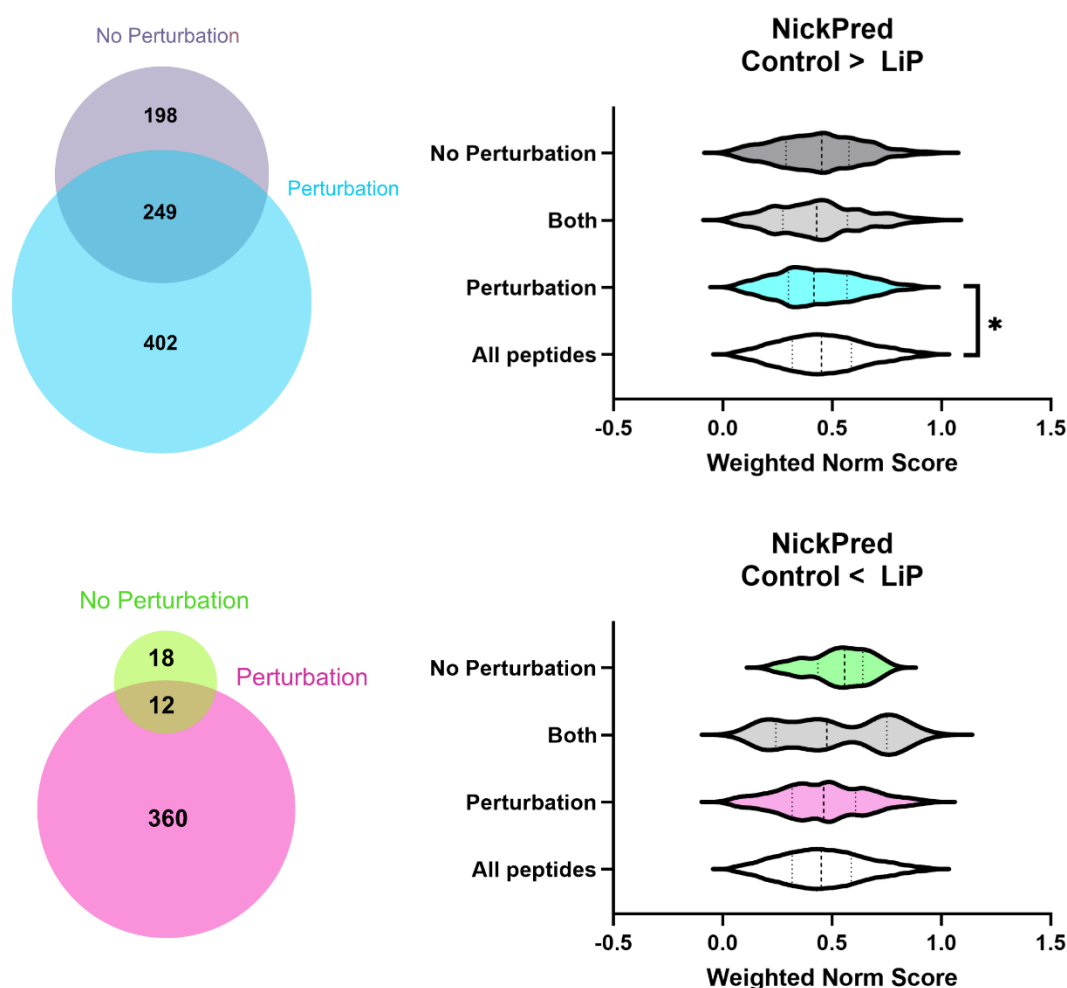


Figure 5.9: Distribution of Nickpred scores for each peptide group. For each protein identified in the MS data, the corresponding AlphaFold predicted structure was downloaded and run through Nickpred. This produced a “protease susceptibility” score from 0 to 1 for each residue, with 1 representing the highest likelihood of proteolytic cleavage. These scores were then summed and averaged in order to create a single score for each peptide. The violin plots on the right display the grouping of these scores based on the different peptide “behaviours” discussed above. Each behavioural group was statistically compared to the “All peptides” readout, representing Nickpred scores for the rest of the peptides identified in this LiP-MS experiment and so provides a dataset baseline. The group of peptides that are only enzyme accessible following perturbation (Control > LiP, Perturbation, blue) were found have significantly lower Nickpred scores than the baseline, following a Mann Whitney T-test (*= P value < 0.05).

As discussed above, the peptides represented by the shared area of the first Venn diagram (Figure 5.9) are those that are enzyme accessible both before and after perturbation.

Therefore, these peptides would be expected to map to protein regions that are stably susceptible to protease cleavage and so with a generally higher Nickpred score. However, this does not seem to be the case, with both the No Perturbation and shared peptides exhibiting non-significant variation from the overall peptides score set. This could be for a variety of different reasons. Firstly, the protein structures used in this analysis were all AlphaFold predictions. Whilst many AlphaFold predicted structures have been found to be

exceedingly accurate, some proteins and protein regions are predicted with low confidence. Due to the volume of proteins being considered in this analysis, no confidence cut off was applied to the structural data the peptides mapped to. Therefore, some Nickpred scores may have been generated based on structures that are not physiologically accurate. Furthermore, AlphaFold protein structures do not take into account post-translational modifications, an important feature governing *in vivo* protease susceptibility. For example, protein-protein interactions and the presence of post-translational modifications will impact PK accessibility to affected regions of the proteins structure, which is not necessarily reflected in this analysis. Moreover, B-values (temperature factors) are missing from AlphaFold, and so important information regarding in solution dynamics is also lost here.

However, LiP peptides only protease-accessible after the perturbation (blue area of Venn diagram) were found to have significantly lower Nickpred scores when compared to the whole-dataset baseline. This could be seen to imply that, in order for the protease to gain access to these proteins regions, some local unfolding is likely to have taken place. This is precisely what is expected from heat shock. Therefore, LiP-MS here may have captured heat-induced protein unfolding.

Another behaviour highlighted in Figure 5.9 is apparent higher Nickpred scores for peptides with significantly higher abundance in the LiP sample compared to the control in the No Perturbation condition (green section, bottom panel Figure 5.9). In some instances, protein regions that are highly susceptible to protease cleavage are also susceptible to aggregation, due to a lack of structural stability. However, due to the small number of peptides represented in this group this increase was not found to be significant when compared to the whole dataset baseline.

Secondary structure mapping

A further possible way to gain insight into the structural characteristics of LiP peptides is to consider what regions of secondary structure they map to. As discussed above, limited proteolysis has been shown many times to correlate with secondary structure features, with β -strand, α -helix and loop regions representing the least to most likely cleavage sites respectively [155, 156, 170]. Therefore, mapping LiP peptides to secondary structure can both verify successful LiP and potentially indicate local protein unfolding.

This was achieved through again utilising predicted structural information from AlphaFold to assign each residue as strand, helix or loop. However, this raised potential issues as, whilst studying tryptic peptide behaviour is more reliable than that of semi-tryptic peptides, this only provides a general area where LiP cleavage has occurred, as theoretically PK could have cleaved anywhere along the length of the peptide. When conducting this analysing it was

found that tryptic peptides very often encompass multiple forms of secondary structure, and so the structure at the actual site of cleavage was obscured. Considering only peptides that mapped to a single form of secondary structure depleted the dataset to too great an extent. The first alternative approach when analysing peptides en masse was to include all secondary structure forms identified for each LiP peptide. For example, when considering the percentage of peptides that mapped to strand, helix or loop regions, a peptide that mapped to both a helix and loop region would be considered a hit for both groups. The second approach was to preferentially assign secondary structure hits, based on the evidence of previous studies. For example, if a peptide mapped to a helix and a loop, this would be considered loop cleavage site. Peptides mapping to a strand and helix would be considered a helix cleavage site, and so on. The results of this analysis are shown in Figure 5.10.

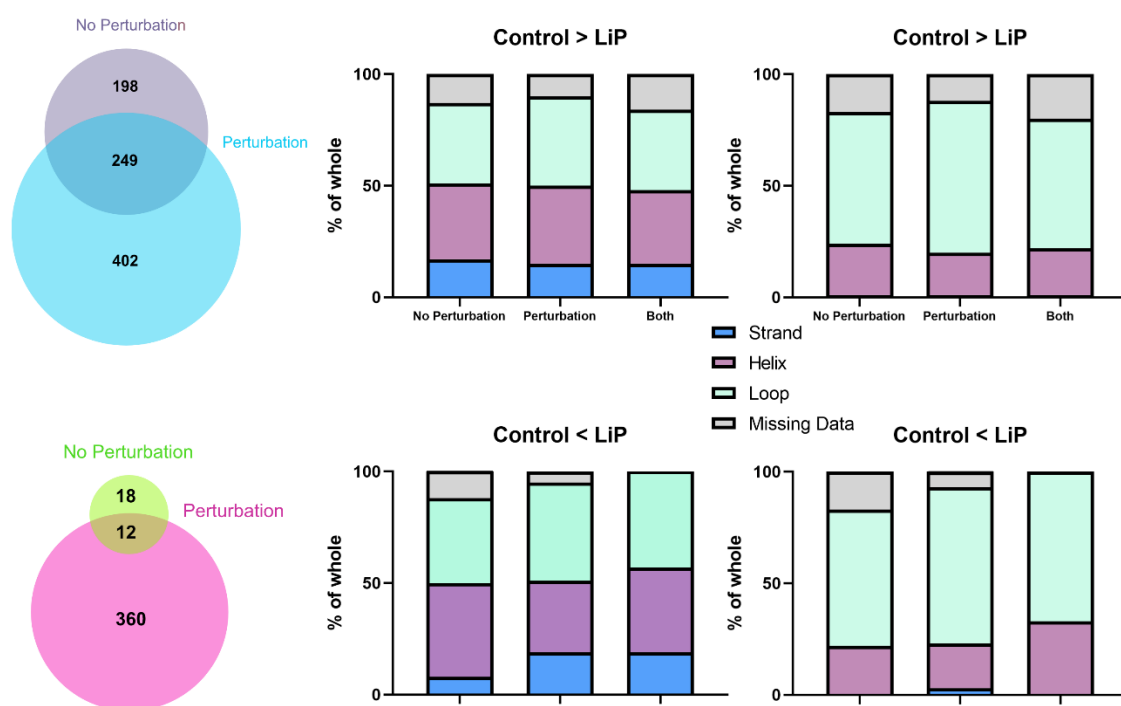


Figure 5.10: Secondary structure mapping of LiP peptides. In the first instance (plots to the left), regions of secondary structure are considered to potentially harbour cleavage sites if a LiP peptide overlaps with them in any way. For example, a LiP peptide that maps to protein region containing both a helix and a loop would count as a potential hit for both. In the second instance (plots to the right), each LiP peptide is considered as only mapping to one type of secondary structure, based on a preferential order of loop, helix and strand.

Overall, there was little difference across the different LiP peptides behaviours. There was a slight increase in the prevalence of peptides mapping to β -strands in the Perturbation (Control<LiP) condition, but not markedly. When mapped preferentially, potential LiP cleavage sites were most commonly associated with loop regions, followed by α -helixes and to a substantially lesser degree to β -strands. This was consistent across all experimental

conditions and is consistent with the literature. However, given the lack of specificity, this analysis has very limited utility.

MobiDB

To take a different approach, another way of assessing the structural characteristics of these peptides of interest is through utilizing MobiDB, a database of intrinsically disordered proteins and intrinsically disordered protein regions [174, 175]. This information has been collated over 10 years and is informed by both experimental evidence and, more recently, from incorporating AlphaFold structural predictions. The annotations available in MobiDB are divided into a four level “annotation pyramid” reflecting the quality of the evidence compared to proteome coverage attained. The first level, “curated”, is the most reliable but least comprehensive, and contains only annotations from manually curated databases. The fourth level, “predicted”, utilizes AlphaFold predicted protein structures to find potential regions of disorder, based on prediction accuracy score information and the predicted relative solvent accessibility for each residue.

The main features MobiDB reports on can be divided into 3 groups; intrinsic disorder, linear interacting peptides and binding modes. Regions of intrinsic disorder are those without native tertiary structure and can be inferred from missing or mobile regions in X-ray or NMR structures, or through sequence analysis such as enrichment in charged residues. Linear interacting peptides is the term used by MobiDB to group binding IDRs, regions that transition between order and disorder upon protein binding. These are identified through database searching or predicted from PDB structures using FLIPPER (Fast Linear Interacting Peptides Predictor [173]). IDR binding is crucial in many biological processes such as signaling and involves structural alterations. Binding modes in MobiDB are identified from comparison of PDB structures of proteins in monomeric and complex form.

Figure 5.11 displays the results of MobiDB analysis for this dataset, again divided into the different behavioral subgroups of LiP peptides. All identified peptides were also analyzed to provide a baseline percentage of IDRs. Peptides were counted as having an IDR hit if they overlapped in anyway with the reported IDR and these results were then filtered to include only hits from the top 2 levels of the “annotation pyramid”, these being curated and derived hits. This decision was made as including prediction hits seemed to hugely overestimate the number of IDRs in the dataset, with 87% of all reported peptides being deemed to have some crossover with an IDR.

Overall, 14% of peptides identified in the MS data contained IDRs. This increased to 26% in the No Perturbation condition. The hypothesis is that the peptides in this condition represent protein regions that are natively enzyme accessible. Therefore, it is consistent with this

hypothesis that a higher than general percentage of these peptides would map to IDRs, which are highly susceptible to enzymatic cleavage. This figure is lessened to 20% in the Perturbation condition. The hypothesis is that peptides in this condition represent protein regions that are only enzyme accessible after heat shock. Therefore, it is expected that this figure would be lower than before heat shock, as one might expect that a number of these cleavages map to regions of heat-induced, non-native unfolding, rather than innate IDRs, as is reported by MobiDB. However, this figure is still higher than the baseline, which could be also be reflective of functional structural change, such as heat-induced change in binding IDRs. Of peptides that were found to be significantly altered in both conditions, 24% were found to overlap with IDRs. These are thought to map to protein regions with enzyme accessibility unaffected by the heat treatment. That this number would represent a middle ground is also consistent, as these potentially reflect the subset of IDRs that remain structurally stable under stress.

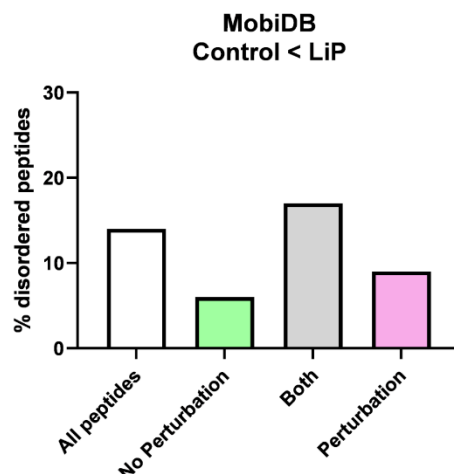
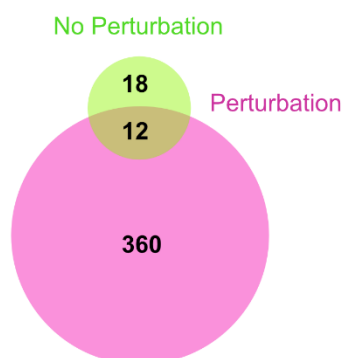
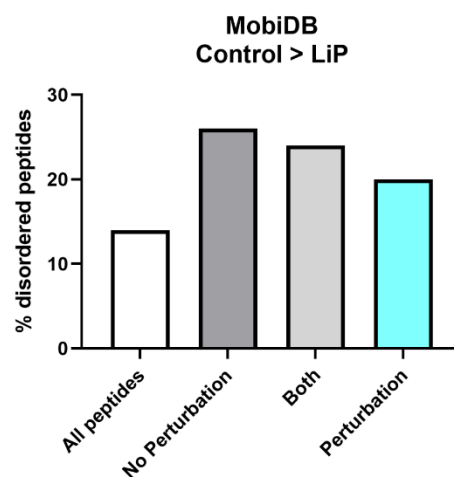
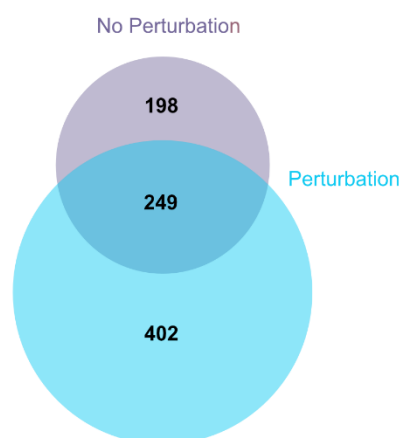


Figure 5.11: Assessment of intrinsically disordered protein regions using MobiDB. All proteins identified in the MS data were run through MobiDB software to identify intrinsically disordered regions. All identified peptides were then compared to these regions and investigated for any overlap. Peptides with an overlap of at least 2 residues were considered to have an IDR. This analysis was again split into the different LiP peptide behaviours

Another behaviour highlighted in Figure 5.11 (bottom) is a lower proportion of overlapping IDRs for peptides with significantly higher abundance in the LiP sample compared to the control in the No Perturbation condition. Here, 6% and 9% of peptides from the No Perturbation and Perturbation condition respectively mapped to IDRs, a decrease from the dataset total of 14%. This is somewhat in support of the previously suggested theory that these peptides may represent less accessible protein regions, that potentially require multiple proteases in order to create peptides that are then identifiable in MS.

Protein solubility

Finally, to test the hypothesis that LiP peptides with significantly higher tryptic peptide abundance in the LiP treated sample than the control potentially map to aggregated proteins, one must consider protein-level structural features.

CamSol is freely available, open source software from the University of Cambridge that calculates protein aggregation propensity based on amino acid sequence [172]. Through analysis of a particular set of physicochemical properties of amino acids, proteins of interest are assigned a solubility score, with lower scores representing higher aggregation propensity and less predicted solubility. In total, 390 peptides displayed the aforementioned behavior, mapping to 293 parent proteins (here referred to as LiP proteins). These proteins were analysed in CamSol to gain a “protein solubility score”, and this was compared to the same analysis conducted for all other proteins identified in the LiP-MS data, to provide a dataset baseline. The results are shown in Figure 5.12 and show a significant reduction in protein solubility score for LiP proteins when compared to the dataset average. Therefore, this supports the hypothesis that these LiP peptides are indicative of heat-induced protein aggregation.

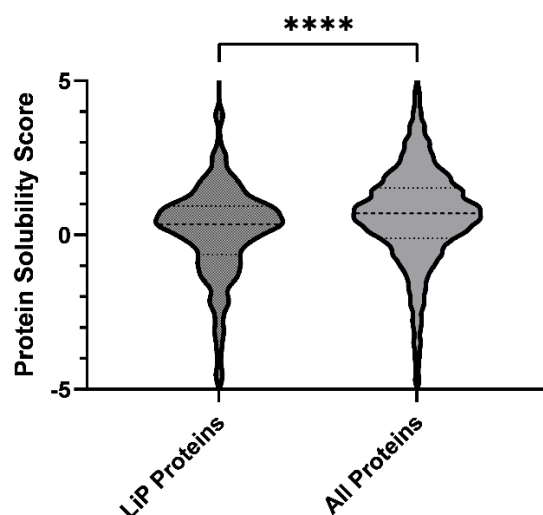


Figure 5.12: CamSol [172] analysis of LiP proteins. The 390 LiP peptides with significantly higher abundance in the LiP treated samples than the controls mapped to 293 parent proteins. These proteins were analysed in CamSol and given a protein solubility score, reflective of their aggregation propensity. The same analysis was also conducted for all other proteins identified in the LiP-MS data, in order to provide a dataset baseline. Scores for each group are displayed here as violin plots, and demonstrate a significant reduction in the mean protein solubility of LiP proteins (t-test, P -value<0.0001) compared to the dataset baseline.

Summary of structural investigations

LiP-MS analysis of hMSCs before and directly after heat stress resulted in the identification of several hundred LiP peptides that indicate stress-induced protein structural alterations.

Of these, ~200 were only present in the No Perturbation condition, implying that these are representative of protein regions that are only enzyme-accessible **before** stress.

Examination of the structural properties of these peptides revealed an enrichment for IDRs, with 26% mapping to MobiDB-defined IDRs, compared to a dataset average of 14%. This supports the idea that these LiP peptides represent protein regions that are enzyme-accessible in native conditions, as IDRs are considered to be highly susceptible to proteolysis. Furthermore, secondary structure mapping demonstrated that these peptides most commonly mapped to loop regions, followed by α -helices, which is also consistent with the known structural determinants of LiP under native conditions.

A further 400 LiP peptides were only present in the Perturbation condition, implying that these are representative of protein regions that are only enzyme-accessible **after** stress.

Examination of the structural properties of these peptides using Nickpred revealed significantly lower protease-susceptibility scores when compared to the dataset average. This result is indicative of local unfolding in these protein regions, a well-documented consequence of heat shock.

An alternative behavior highlighted in this experiment was that of significantly higher tryptic peptide abundance in the LiP treated samples compared to the controls, which is the opposite of the conventional LiP peptide fingerprint. It was hypothesized that this behavior may be indicative of protein aggregation or, more generally, of sites inaccessible to trypsin. Investigation of these peptides using MobiDB found a decreased prevalence of IDRs, with 6% and 9% mapping to IDRs before and after heat shock respectively, compared to the dataset average of 14%. This is supportive of the hypothesis that these peptides map to protein regions with decreased accessibility. Furthermore, analysis of the parent proteins of these peptides of interest using CalSol demonstrated a significant reduction in protein solubility score (increase in aggregation propensity) when compared to the dataset average. Therefore, this supports the hypothesis that these LiP peptides are indicative of heat-induced protein aggregation.

Overall, examination of the structural features of the LiP peptides identified in this study has demonstrated good correlation between the known structural determinants of LiP sites under native conditions, and the effects of heat shock on protein structure. However, much remains unexplained by considerations of structure alone. For example, secondary structure mapping revealed little difference across conditions. Further, IDR “hits” from the MobiDB analysis not only include regions of stable, native disorder, but also sites of functional structural changes such as binding IDRs, which have not yet been investigated.

Therefore, the next step in this analysis is to look for structural alterations that are indicative of cellular dynamics.

5.2.4 GO enrichment analysis of structurally altered proteins

The 600 LiP peptides identified as potentially mapping to regions of stress-induced protein structural change were further investigated, using a Gene Ontology (GO) Enrichment Analysis of their parent proteins. This analysis was processed through Protti and involved the comparison of GO terms assigned to these potentially structurally altered proteins to the GO terms for all proteins detected in this experiment. Whether or not these GO terms were significantly over- or under-represented in our protein fraction of interest was determined using a two-sided Fisher’s exact test. The results of such analysis are shown below, for Biological Process enrichment (Figure 5.13) and Molecular Function enrichment (Figure 5.14).

The proteins LiP-MS highlighted as undergoing stress-induced protein structural change (“LiP proteins”) are significantly enriched for GO terms associated with the cellular response to loss of proteostasis (highlighted in red, Figures 5.13, 5.14). For example, “protein folding”, “heat shock protein binding”, “protein folding chaperone” and “unfolded protein binding” are

all identified as overrepresented in the LiP proteins. This correlates well with the literature, as molecular chaperones are well known to be a crucial part of the response to stress. Studies have often focused on demonstrating the upregulation of heat shock proteins in stress conditions, but the mode of action of HSPs very often requires small molecule binding, local unfolding or refolding, intracellular re-localisation or complex formation [47]. No significant change in molecular chaperone protein *abundance* was observed in this study, but LiP-MS also provides a readout of differences in these other mode of action conformational parameters. Therefore, the overrepresentation of molecular chaperone associated GO terms in LiP proteins indicates that HSPs are undergoing structural alterations in order to become functionally active in response to perturbation, as has been reported many times in the literature [48, 167, 213-215].

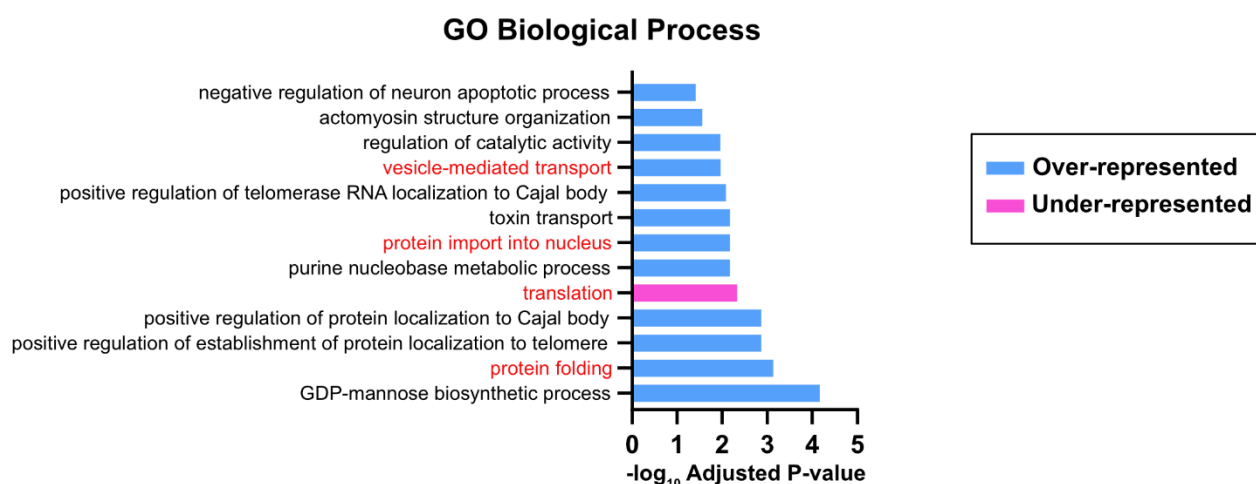


Figure 5.13: GO Biological Processes. GO enrichment analysis of LiP proteins. Significant over or under enrichment was determined using a two sided Fisher's exact test. Over-represented pathways are highlighted in blue, and under-represented highlighted in pink. GO biological processes associated with the cellular stress response are highlighted in red.

Another term highlighted in this analysis is “proteasome-activating activity”. The Ubiquitin Proteasome System (UPS) is another well documented protein quality control mechanism with an important role in the heat shock response, in this case degrading proteins that may have become unfolded under stress [216]. Proteasome activity can be regulated in three ways; transcriptionally, with post-translational modifications and through changes in complex composition [217]. No significant change in the expression of the proteasomal subunits identified in the MS data was observed. However, the presence or absence of PTMs and the interactions between subunits are factors that may affect LiP enzyme accessibility, and so the enrichment of this GO term in the LiP-MS structural data could be seen as an early indication of the latter two forms of proteasome activation.

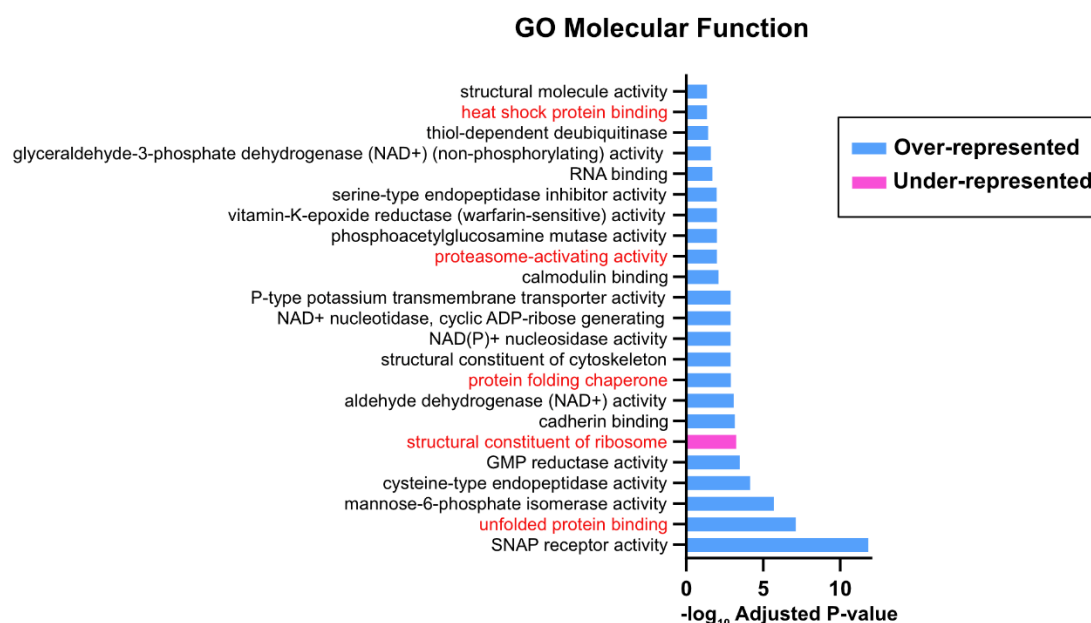


Figure 5.14: GO Molecular Function. GO enrichment analysis of LiP proteins. Significant over or under enrichment was determined using a two sided Fisher's exact test. Over-represented pathways are highlighted in blue, and under-represented highlighted in pink. GO biological processes associated with the cellular stress response are highlighted in red.

Finally, “protein import into the nucleus” has also been highlighted by this analysis. Again, this is a known component of the heat shock response and the mode of action of many of the associated proteins is reliant upon structural changes. For example, importin α has been shown to accumulate in the nucleus during heat shock at 42 °C, a process reliant upon complex formation with importin β and various cargo proteins [218]. Furthermore, the master regulator of the heat shock response, HSF1, must be able to relocate to the nucleus in order to instigate the transcription of heat shock proteins. This is also dependent upon the formation and dissociation of complexes and post-translational modifications [60, 219].

Overall, the GO enrichment analysis of LiP-peptide containing proteins indicates that a subset of the stress-induced structural alterations identified are over-represented in functionally important stress-response processes.

5.2.5 Post-Translational Modifications

One crucial category of functionally critical protein structural alterations are post-translational modifications (PTMs), one of the most abundant and important of which is phosphorylation [30]. Given this importance, it was prudent to investigate whether any of the 600 LiP peptides mapped to known phosphorylation sites (phosphosites). Phosphorylation could affect whether a peptide is detected, the signal intensity of the peptide in the MS, and could also correlate with a change in protein conformation which affects the accessibility to enzyme cleavage.

Table 5.2: LiP peptides map to phosphorylation sites. LiP peptides were compared to a curated human phosphoproteome [220] to find those that overlapped or were within proximity (within 10 amino acids). The curated phosphoproteome also contained a “function score” for each phosphosites, with scores between 0.5 and 1 representing phosphosites predicted to have particular functional importance. Table shows a subset of the LiP peptides which mapped to phosphosites, with information including the peptide sequence, parent protein, LiP peptide behaviour, peptide start and end position, phosphosites position and the phosphosites functional score. Sites of particular interest are highlighted here in red and discussed further in text.

Seq	Protein	Behaviour	Start	End	Phospho Site	Site Functional Score
SGDSEVYQLGDVS QK	sp Q04837 SSBP_HUMAN	No Perturbation Control>LiP	67	81	73	0.597613
ASGVAVSDGVIK	sp P23528 COF1_HUMAN	No Perturbation Control>LiP	2	13	3	0.990121
TVYGGGCSEMLM AHAVTQLANR	sp P78371 TCPB_HUMAN	No Perturbation Control>LiP	406	427	428	0.630489
EAVAMESYAK	sp P78371 TCPB_HUMAN	No Perturbation Control>LiP	432	441	428	0.630489
SSTETCYSAIPK	sp O75369 FLNB_HUMAN	No Perturbation Control>LiP	2496	2507	2502	0.581049
EASIDILHSIVK	sp O00231 PSD11_HUMAN	No Perturbation Control>LiP	21	32	14	0.77256
VEVTEFEDIK	sp Q01105 SET_HUMAN	Perturbation Control>LiP	123	132	133	0.557371
IYGGSVTGATCK	sp P60174 TPIS_HUMAN	Perturbation Control>LiP	207	219	215	0.556323
LSDGVAVLK	sp P10809 CH60_HUMAN	Perturbation Control>LiP	397	405	410	0.641583
IQEIIQLDVTSEY EKEK	sp P10809 CH60_HUMAN	Perturbation Control>LiP	371	389	385	0.528906

Ochoa et al. 2020 curated a reference phosphoproteome with information on approximately 120,000 human phosphosites [177, 220]. These phosphosites were searched to find those that overlapped or were in proximity (within 10 amino acids, as defined by Cappelletti, Hauser et al. 2021 [167]) to a LiP peptide. This analysis found phosphosite matches for 193 LiP peptides, representing 32% of all LiP peptides. This reference phosphoproteome also contained a “functional score” for each phosphosite, determined using machine learning trained on around 60 features that were deemed to be reflective of structural and regulatory importance. Scores were assigned between 0 and 1, with scores above 0.5 representing particularly functionally impactful phosphosites. In total 59 LiP peptides mapped to phosphosites meeting this criterion, a subset of which are displayed in Table 5.2.

Therefore, of the protein regions that were differentially accessible to the LiP protease before and after stress, a third mapped to reported phosphosites. This follows on from the aforementioned study utilizing LiP-MS to investigated the yeast osmotic stress response, which found a subset of LiP peptides overlapped or were in close proximity (within 10 amino acids) to a phosphopeptide [167]. The study then further investigated the highlighted phosphosites for links to the stress response. This was repeated here, with a particular focus on the subset of 59 predicted “functionally important” phosphosite matches. Several known

components of the cellular response to heat stress were highlighted, a subset of which is highlighted in red in Table 5.2.

For example, phosphorylation of 26S proteasome non-ATPase regulatory subunit 11 (PSMD11) at serine 14 has been shown to increase proteasomal activity and so the degradation of misfolded proteins under stress [221]. The LiP peptide found adjacent to this site (highlighted in red, Table 5.2) was identified in the No Perturbation condition only. Several other peptides were identified for this protein that were not significantly altered with stress, as shown in Figure 5.15.

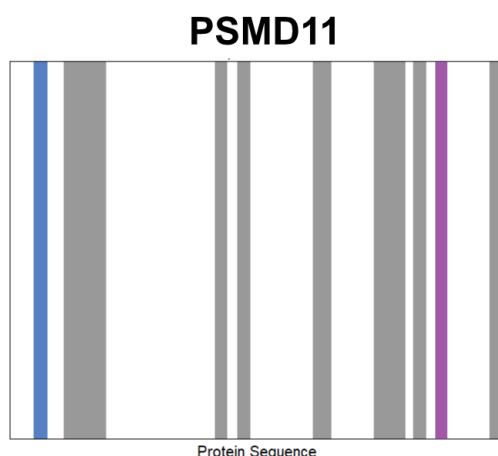


Figure 5.15: Barcode plot of 26S proteasome non-ATPase regulatory subunit 11 (PSMD11). Twelve peptides mapped to this protein, shown here as shaded regions along the protein sequence, displayed as a barcode plot. The peptide highlighted in blue was found to be significantly changing in the No Perturbation condition only. The peptide highlighted in purple was found to be significantly changing in both the Perturbation and No Perturbation conditions. All other peptides mapping to this protein were not significantly altered by stress or LiP treatment.

Another stress-responder highlighted was molecular chaperonin HSP60. Two LiP peptides identified in the Perturbation condition mapped to phosphosites with high predicted functional importance (“CH60_HUMAN”, Table 5.2). No further information regarding these phosphosites was found in the literature, but during the course of research evidence was found that HSP60 is ubiquitinated at lysine 396, a site between these two LiP peptides. Whilst the functional consequences of such ubiquitination are not clear, it has been suggested that it is stress associated [222], could indicate involvement with the proteasome system [223] and may impair HSP60’s chaperone functioning [213].

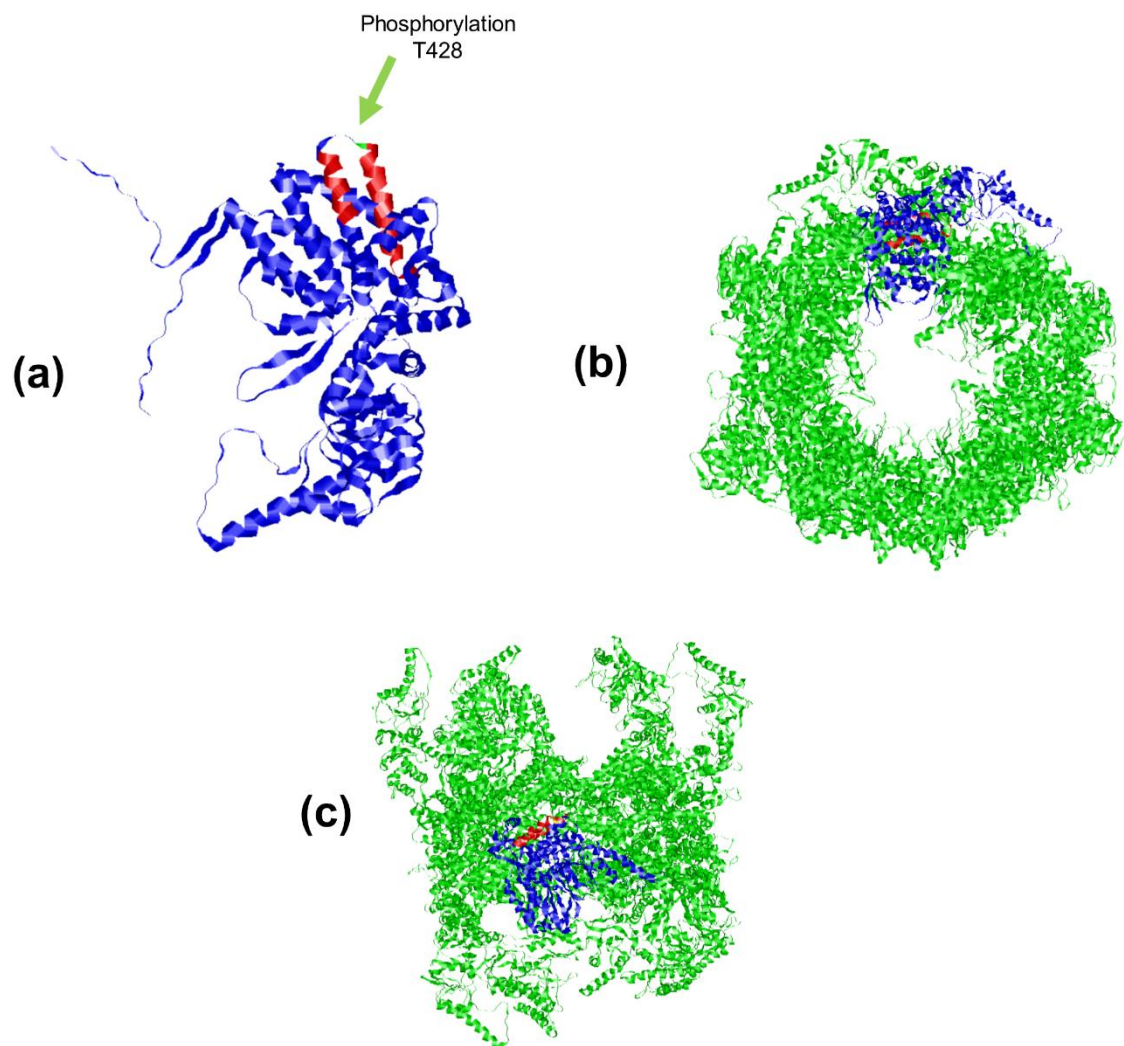


Figure 5.16: T-complex protein 1 subunit β (CCT2). (a) AlphaFold structure of CCT2, with LiP peptides (No Perturbation condition) highlighted in red. The phosphorylation site at T428 is shown in green and indicated by arrow. (b) + (c) Cryo-EM structure (PDB:5GW4 [214]) of the CCT/TRiC complex. Here CCT2 is highlighted in blue, and the LiP peptides are again highlighted in red. (b) View from above ring structure, (c) view from side.

Furthermore, T-complex protein 1 subunit β (also known as CCT2) is a component of the CCT/TRiC complex, an ATP-dependent molecular chaperone machine involved in folding of cytoskeletal proteins. A potential CCT2 phosphosite was highlighted at T428, which was found to be flanked by two LiP peptides (“TCPB_HUMAN”, Table 5.2). These LiP peptides were unique to the No Perturbation condition and so represent a protein region accessible only before heat stress. No functional information was found pertaining to this specific phosphosite. However, a recent study used cryo-EM structures to investigate the dynamics of the complex under ATP binding and identified CCT2 as a crucial subunit for complex formation and allosteric cooperativity, with CCT2 undergoing substantial conformational

change upon ATP binding [214]. These structures are displayed in Figure 5.16, from multiple angles, with the LiP peptides, phosphosite and CCT2 protein highlighted.

Substrate binding sites

The GO term enrichment analysis highlighted terms associated with molecular chaperones (“protein folding chaperone”, “heat shock protein binding”) as significantly enriched among LiP proteins. These results inspired the investigation of other abundant molecular chaperones for signs of stress-induced structural alterations.

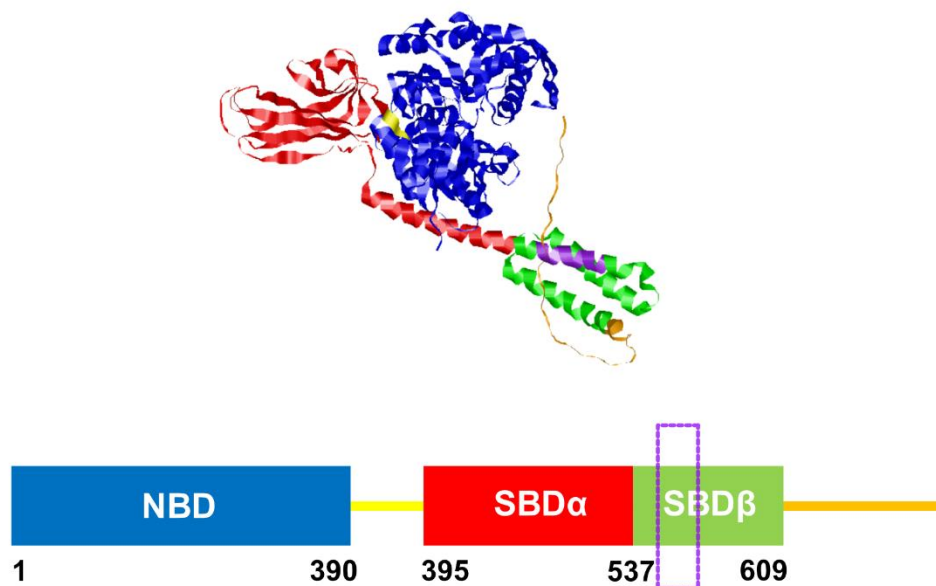


Figure 5.17: Adapted from [215]. AlphaFold structure of Hsp70 (1B). The nucleotide binding domain (NBD) is highlighted in blue, linker region in yellow, substrate binding domain (SBD) α in red, β in green and the disordered C-terminal in orange. The identified LiP peptide (No Perturbation only) is highlighted in purple. This is also shown in the colour matched schematic, with the position of the LiP peptide along the protein sequence highlighted.

HSP70 proteins are highly conserved, ubiquitous, and associated with several mechanisms of proteostasis such as protein import, assembly, refolding, and disaggregation [46]. The structure of HSP70 is characterized by a C-terminal substrate-binding domain (SBD) capable of recognizing extended hydrophobic protein regions. This process is mediated by ATP binding at the N-terminal nucleotide-binding domain (NBD). Delivery of client proteins to the ATP-bound, open confirmation of HSP70 is facilitated by co-chaperone HSP40. This process initiates hydrolysis of the bound ATP, resulting in a closed HSP70-substrate complex. The substrate is released following NEF-triggered ADP/ATP exchange [47].

In the LiP-MS data, 24 peptides mapped to HSP70, including one LiP peptide, which was identified as significantly changing in the No Perturbation condition only. This peptide mapped to the substrate binding domain, as shown in Figure 5.17. This can be seen as evidence that, following perturbation, this protein region is no longer accessible, as this HSP70 is now held in its active, client-bound formation. This is again similar to the aforementioned study from the Picotti lab, where LiP-MS analysis of the heat stress response in yeast also revealed an enrichment of molecular chaperones in structurally altered proteins, and an example was provided in which LiP peptides were mapped to the substrate binding pocket of disaggregase chaperone HSP104 [167].

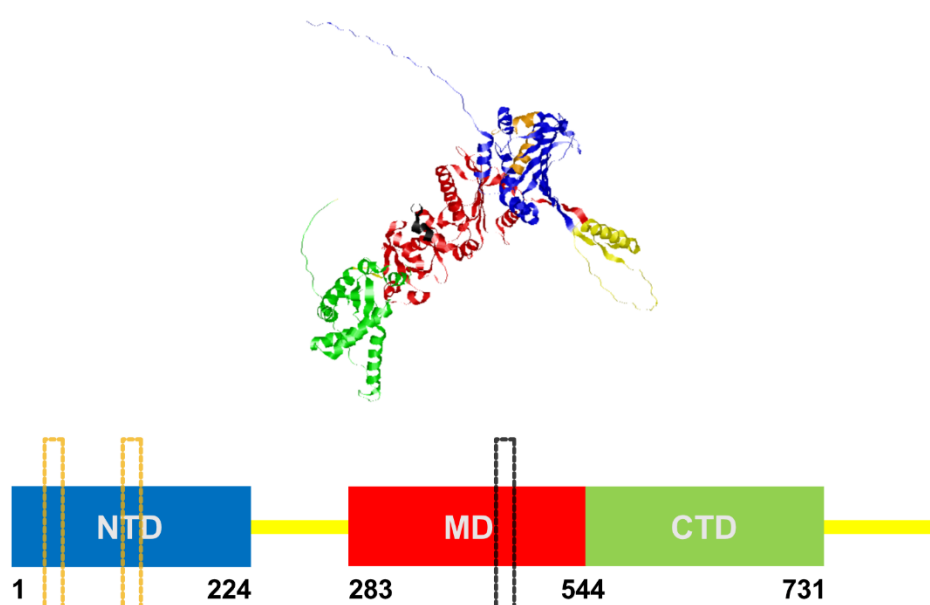


Figure 5.18: Adapted from [215]. AlphaFold structure of HSP90α. The N terminal domain (NTD) is highlighted in blue, linker region in yellow, middle domain (MD) in red, C-terminal domain in green. The identified LiP peptides (Perturbation only) are highlighted in orange, and the alternative LiP peptide is highlighted in black. This is also shown in the colour matched schematic, with the position of the LiP peptides along the protein sequence highlighted.

HSP90 chaperones are also highly conserved and ubiquitously expressed, with many hundreds of client proteins. The general structure of HSP90 chaperones consists of a N-terminal domain responsible for ATP binding, a middle domain involved in ATP hydrolysis and substrate binding and a C-terminal domain associated with dimerization, which is essential for HSP90 function. In eukaryotes, the N-terminal and middle domains are connected by a charged linker region. Client protein folding by HSP90 is a complex process involving interactions with several other chaperones and co-chaperones. For example, clients bound by the HSP70/HSP40 machinery can be additionally bound by HSP90 when facilitated by co-chaperones, such as Hop (STI1) [48, 49, 224]. Two LiP peptides were

identified in the Perturbation condition mapping to HSP90 α 's N-terminal domain, shown in Figure 5.18. This is possibly indicative of the substantial conformational change undergone by HSP90 when bound by ATP.

Interestingly, one of the peptides identified for HSP90 displayed significantly increased signal in the LiP condition compared to the control, in the Perturbation condition. The region occupied by this peptide is highlighted in black in Figure 5.18, and maps to the centre of the middle domain – the region of HSP90 most associated with substrate binding. As previously discussed, this is the opposite behaviour of the classical LiP peptide fingerprint and one hypothesis for the presence of peptides with such behavior was protein aggregation, which is supported to an extent by the CamSol protein solubility assessment (Section 5.2.3, Protein solubility).

However, this assessment was based solely on the proteins amino acid sequence, and fails to take into account functional protein-protein interactions. Therefore, an alternative, or further explanation could be that these peptides are indicators of protein-complex formation. In this scenario, PK cleavage destabilizes the protein complex, increasing solubility. Previously bound protein regions are then available for tryptic digestion.

Table 5.3: Complex Portal. Parent proteins of tryptic peptides with significantly higher signal in the LiP condition were compared to the Complex Portal manually curated list of human protein complexes [176]. 178/293 matches were found, with a selection of matches of particular interest displayed in table. Information regarding complex identifier, name, the accession numbers of the complex member proteins, the structure of the complex and the matches from the LiP-MS dataset

is displayed.

#Complex ac	Recommended name	Identifiers (and stoichiometry) of molecules in complex	Complex assembly	Matches
CPX-5993	26S Proteasome complex	O00231(1) O00232(1) O00487(1) O14818(1) O43242(1) P17980(1) P20618(1) P25786(1) P25787(1) P25788(1) ...	33-mer heteromer	O43242,P28074,P25786,O14818,P62191,P25789,P25788,P49721,O00232,P60900
CPX-6030	Chaperonin-containing T-complex	P17987(2) P40227(2) P48643(2) P49368(2) P50990(2) P50991(2) P78371(2) Q99832(2)	Heterohexadecamer	P40227
CPX-6036	Eukaryotic translation initiation factor 3 complex	O00303(1) O15371(1) O15372(1) O75821(0) O75822(1) P55884(1) P60228(1) Q13347(1) Q14152(1) Q7L2H7(1) Q99613(1) Q9UBQ5(1) Q9Y262(1)	Heterotridecamer	P55884,Q9Y262
CPX-5223	40S cytosolic small ribosomal subunit	P08708(0) P08865(0) P15880(0) P23396(0) P25398(0) P39019(0) P46781(0) P46782(0) P46783(0) P60866(0) ...	-	P63244,P46781,P62269,P62979,P62249,P62701,P62273,P15880
CPX-3288	HSP90A-CDC37 chaperone complex	P07900(2) Q16543(2)	Heterotetramer	P07900

This hypothesis was investigated using the open source Complex Portal, a manually curated database of macromolecular complexes [176]. In total, 390 significantly altered tryptic peptides had a higher abundance in the LiP treated sample and 360 of these were unique to the Perturbation condition (Figure 5.8). These peptides correspond to 293 proteins in total, of which 178 (61%) corresponded to known protein complexes, as defined by the human Complex Portal database (1400 macromolecules). For comparison, 15% of all proteins reported in the LiP-MS dataset were found to be members Complex Portal protein complexes. Overall, these proteins of interest are over-represented as protein complex components. Further, the prevalence of LiP peptides with this fingerprint increased twentyfold with stress and investigation of protein complex matches highlighted many known stress-responders, as demonstrated in Table 5.3. For example, ten LiP parent proteins were found to be members of the 26S proteasome complex.

5.3 Summary

The aim of this chapter was to assess the capacity of LiP-MS to provide a global structural readout of the response to stress in hMSCs. Stress was induced through the application of heat shock and a chaperone inhibitor and LiP-MS was performed on the perturbed and unperturbed cells in order to identify perturbation-induced protein structural changes. Analysis revealed no significant stress-induced alterations to protein abundance. This aligns with previous data produced in the Swift lab, that indicates the role of protein in the early stress response is predominantly post-translational.

However, the capacity of LiP to generate detectable features in the MS data of this experiment was demonstrated in several ways. Firstly, principal component analysis revealed a clustering of MS runs based on LiP treatment, with the strongest LiP condition exhibiting most variance to the control, as expected. Furthermore, the number of tryptic peptides identified by MS was shown to negatively correlate with LiP enzyme concentration, whilst the number of semi-tryptic peptides identified by MS positively correlated with enzyme concentration. Finally, the effect of LiP treatment on the MS signal for specific proteins and how this correlated to the protein's structural features was explored, providing examples of proteolytically resistant and proteolytically culpable proteins.

Moreover, LiP peptides (in this case referring to those with a significantly reduced tryptic peptide signal upon LiP treatment) were identified in both stress conditions and led to the identification of hundreds of proteins with potential stress-induced structural alterations.

Of these, ~200 LiP peptides were only present in the No Perturbation condition, implying that these are representative of protein regions that are only enzyme-accessible **before** stress. Analysis of the structural characteristics of these peptides revealed an over-representation of intrinsically disordered regions, which is consistent with the known structural determinants of LiP under native conditions.

A further 400 LiP peptides were only present in the Perturbation condition, implying that these are representative of protein regions that are only enzyme-accessible **after** stress. Examination of the structural properties of these peptides using Nickpred revealed significantly lower protease-susceptibility scores when compared to the dataset average. This result is indicative of local unfolding in these protein regions, a well-documented consequence of heat shock.

GO enrichment analysis of structurally altered proteins (parent proteins of the LiP peptides) found they were enriched for many known cellular stress response factors, such as “unfolded protein binding” and “proteasome activating activity”. This indicates that, despite the lack of protein abundance changes, LiP-MS is able to capture the dynamics of the cellular response to stress through the identification of functional protein conformational alterations that form a crucial part of the proteostasis network. Furthermore, 32% of LiP peptides that were differentially accessible before and after stress mapped to phosphorylation sites, a PTM that is functionally critical and plays a substantial role in the cellular response to stress. Stress response factors highlighted in the GO enrichment analysis were then further investigated through LiP peptide mapping, revealing stress-induced structural alterations (differential protease accessibility) at PTM sites, ATP binding sites and, in the case of molecular chaperones, client-protein binding sites.

An alternative behavior highlighted in this experiment was that of significantly higher tryptic peptide abundance in the LiP treated samples compared to the controls (~400 peptides), which is the opposite of the conventional LiP peptide fingerprint.

A working hypothesis consistent with this observation is that this may be caused by protein aggregation, or at least protein site-inaccessibility to trypsin. Some tightly folded proteins, and indeed some regions of proteins, are more resistant to digestion, particularly if they form aggregates. In such cases, digestion with trypsin alone may not be effective in producing MS-amenable peptides. However, should the addition of PK result in extra cleavage(/s) of trypsin-resistant proteins at stability-critical sites, this could lead to increased accessibility for trypsin in the subsequent digestion step. Under these conditions, said tryptic peptides would be more predominant in the double digestion (LiP) treated samples than the control.

Furthermore, the presence of significantly changing peptides with this behaviour was increased twentyfold after stress. This supports the hypothesis, as heat stress is known to induce protein aggregation and thus will likely increase the prevalence of proteolytic-resistant proteins.

Interrogation of this hypothesis through assessment of peptide structural characteristics found a decreased prevalence of IDRs, which is consistent with the idea that these peptides map to regions of decreased accessibility. Moreover, analysis of the parent proteins of these peptides of interest using CalSol demonstrated a significant reduction in protein solubility score (increase in aggregation propensity) when compared to the LiP-MS dataset average. Therefore, this supports the hypothesis that these LiP peptides are indicative of heat-induced protein aggregation.

A possible further explanation could be that these peptides are indicators of protein-complex formation. In this scenario, PK cleavage destabilizes the protein complex, increasing solubility. Previously bound protein regions are then available for tryptic digestion. Indeed, protein complex components were found to be over-represented in this group. Further, the protein complex components highlighted here corresponded to known stress-response macromolecules, such as the 26S proteasome complex.

Therefore, LiP-MS can provide structural insights into the hMSC response to stress.

Chapter Five: Addendum

Throughout the course of the LiP-MS data analysis in Chapter 5, only tryptic peptides were considered. This was due to the uncertainty surrounding the identification and quantification of semi-tryptic peptides discussed in previous chapters. Whilst global features of the datasets linked to the total numbers of tryptic and semi-tryptic peptides were entirely consistent with expected changes in the proteome of the cells under given conditions, local protein-specific changes were more challenging to detect. The principal method used to determine sites of potential limited proteolysis was observation of changes in tryptic peptide abundance, as shown on the volcano plots presented in previous chapters (e.g. Figure 5.7). Although this methodology has been used successfully in differential proteomics experiments and is standard in the field, the normalization algorithms applied are designed to detect changes at the protein level in the context of a proteome which is not expected to show global changes. This latter assumption is preserved here, namely that the change in global protein abundance is minor, and indeed most individual proteins do not show a significant change in relative abundance under the model. However, to detect sites of LiP at the peptide level a drop in relative signal is expected, as should be observed on a volcano plot.

However, this observation is made without directly considering the other peptides in that protein which would not be expected to change if they were not subject to limited proteolysis. Moreover, global changes to the observed protein signal are not formally considered in this model, such as those caused by total digestion or aggregation. Local normalization approaches can potentially account for such effects, by normalizing the peptide signal to the total protein signal. In order to gain a better understanding of this behaviour and to test these assumptions, as well as to generate a control dataset of expected normalized peptide signals for a standard proteome under different environmental conditions, an additional MS experiment was conducted.

Here, cells were subjected to 2 hours heat shock at 42 °C and then allowed to recover for 4 hours at 37 °C. Cells were taken before heat shock (C1), immediately after (C2), and following the 4 hours recovery period (C3), creating three experimental conditions, the first two of which match the experiment conducted in Chapter 5. However, in this experiment lysates from each condition were subject to digestion with trypsin only before analysis by MS. In other words, no LiP step was included. This was to generate a control ‘background’ data set. The cells used were the immortalized hMSC cell line Y201 (Section 2.5) and for each condition 7 technical replicates (sample preparation) were run.

An aim of this experiment was to further investigate the presence of semi-tryptic peptides in trypsin-only (LiP-MS control) experiments and how this may be impacted by different experimental conditions, using a higher number of replicates and cell line.

Firstly, the number of tryptic and semi-tryptic peptides identified in each condition is shown in Figure 5.19. An average of 1150 semi-tryptic peptides were identified, representing approximately 5% of all peptides. The number of semi-tryptic or tryptic peptides identified did not change significantly across the different experimental conditions.

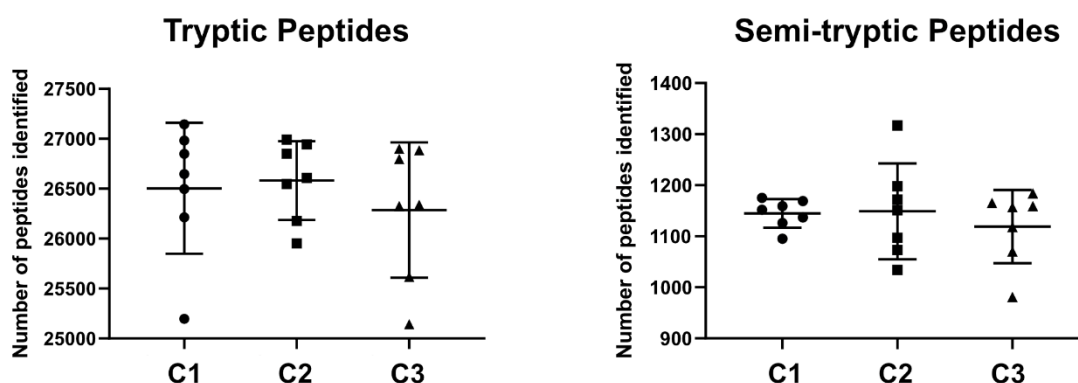


Figure 5.19: Number of peptides identified with $FDR < 0.01$. Left – number of tryptic peptides identified in each replicate of each experimental condition. Right – number of semi-tryptic peptides identified in each replicate of each experimental condition. Identified is defined as having a non-zero reported intensity. Bars represent the mean and standard deviation of the 7 technical replicates.

Overall, 1793 semi-tryptic peptides were identified in at least one replicate of one condition. These were mapped to their parent protein sequence and the non-tryptic termini investigated. This revealed that 502 semi-tryptic peptides (28%) had the non-tryptic termini at residue number 2 of the parent protein's sequence. This does not appear to be a feature at the protein C-terminal, just 14 semi-tryptic peptides map to within 2 residues of the C-terminal end. Therefore, these are not “true” semi-tryptic peptides, but instead a tryptic peptide mapping to the start of a protein sequence with a missing first residue. Indeed, in MS data processed through MaxQuant with the digestion parameters set to “specific”, the same feature was observed. In other words, when MaxQuant searches for tryptic peptides only, these same peptides are reported as tryptic, despite having a non-tryptic N-terminus. These could well be produced post-translationally as part of the normal maturation step and/or potentially as the result of some exoprotease processing, but would not be regarded as a natural limited proteolytic cleavage produced by a proteinase K digest. They could (and should) be discarded from consideration of candidate LiP peptides.

Next, the peptide signal consistency across the technical replicates was investigated, through calculating the coefficient of variance (ratio of standard deviation to mean) for each reported peptide, using the intensity reported by MaxQuant. This is shown in Figure 5.20,

and split into tryptic and semi-tryptic peptides for each experimental condition. Demonstrated here is a significantly larger coefficient of variance of semi-tryptic peptides, compared to tryptic, in each condition. Moreover, removing the semi-tryptic peptides with non-tryptic termini mapping to protein residue 2 further increases this difference between tryptic and semi-tryptic consistency (Figure 6.2(b)). Interestingly, the variance in semi-tryptic peptide signal also seems to be altered by experimental condition, with those identified in C2 displaying particularly high levels of inconsistency.

In other words, tryptic peptides have a much more consistent signal across replicates and experimental conditions than semi-tryptic peptides. This supports the general assertion that they represent a less reproducible measure for assessing limited proteolytic cleavage in (near) native folded proteins.

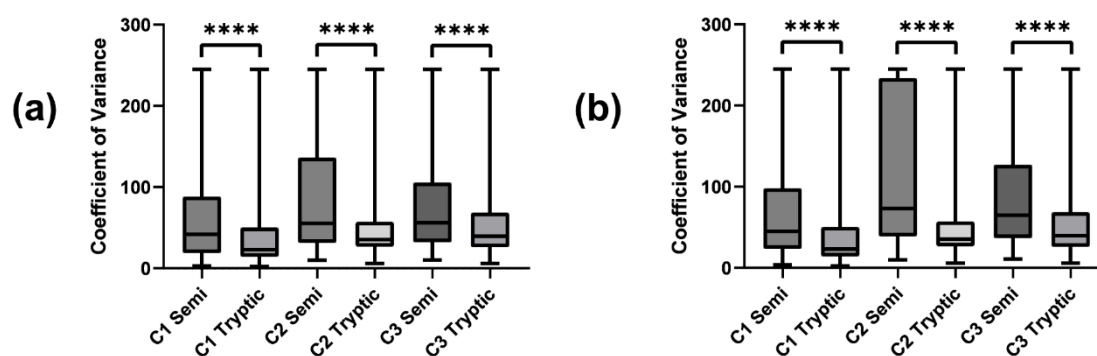


Figure 5.20 (a) Box and whisker plots displaying the coefficient of variance for semi-tryptic and tryptic peptides, for each experimental condition. (b) The analysis of (a) with the semi-tryptic peptides with non-tryptic termini mapping to protein residue 2 removed. Significance was determined using a Mann Whitney t-test, **** p -value < 0.0001.

Peptides normalized to protein signal

To investigate the contribution of individual tryptic and semi-tryptic peptides to total signal for a given protein, each reported peptide intensity was divided by the total reported signal for its parent protein, for each replicate in each condition. Therefore, the normalized peptide values reflect the contribution of that peptide to the overall protein signal.

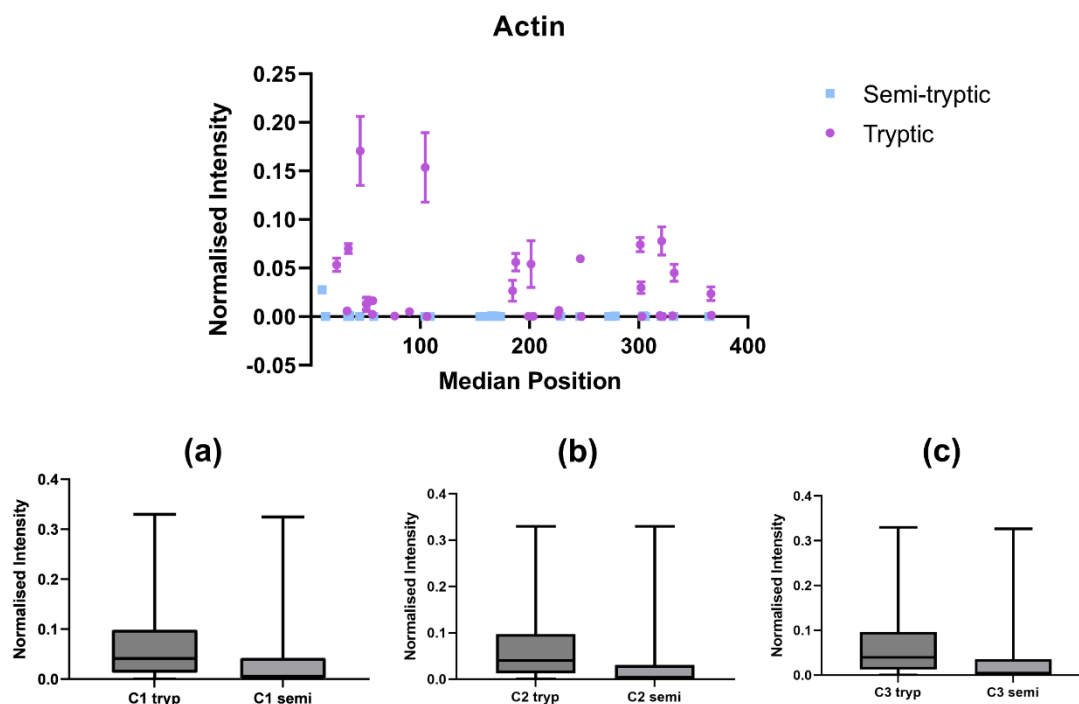


Figure 5.21: The normalised intensity of each peptide identified for actin is displayed along the protein's sequence. The mean and standard deviation of the 7 replicates is shown. Tryptic peptides are shown in pink and semi-tryptic peptides shown in light blue. Only non-zero values are shown. (a)-(c) Box and whisker plots showing the distribution of all normalised intensity values for tryptic and semi-tryptic peptides, for experimental conditions 1 to 3 respectively. Only proteins with at least 3 peptides were included.

The effect of this normalization was investigated in actin, as a well-studied, abundant protein. In C1, 32 tryptic peptides were identified for actin and 34 semi-tryptic peptides. However, despite their number, when normalised the semi-tryptic peptides all had a very small contribution to the protein signal, compared to their tryptic counterparts, as shown in Figure 5.21. Indeed, this is reflected in the dataset as a whole. For each condition, the distribution of normalised intensities for all tryptic peptides was compared to the same for all semi-tryptic peptides in Figure 5.21(a)-(c). Only proteins with at least 3 peptides were included. In each condition, the contribution of semi-tryptic peptides to protein signal was substantially lower than that of tryptic peptide signal.

Next, the impact of experimental condition on tryptic and semi-tryptic normalised peptide intensity was investigated. As shown in Figure 5.22, the mean normalized intensity was taken for each peptide and compared across conditions. Spearman's rank correlation coefficient (R) was calculated for each comparison and showed a remarkably strong linear relationship (all $R > 0.97$), for both tryptic and semi-tryptic peptides. Therefore, normalising peptides to protein signal confers consistency across experimental conditions, for both tryptic and semi-tryptic peptides. This offers good evidence this normalization procedure is not confounded by differing experimental conditions perturbing protein levels and negatively

influencing peptide levels, at least in terms of global trends in the correspondence in overall normalized peptide signal.

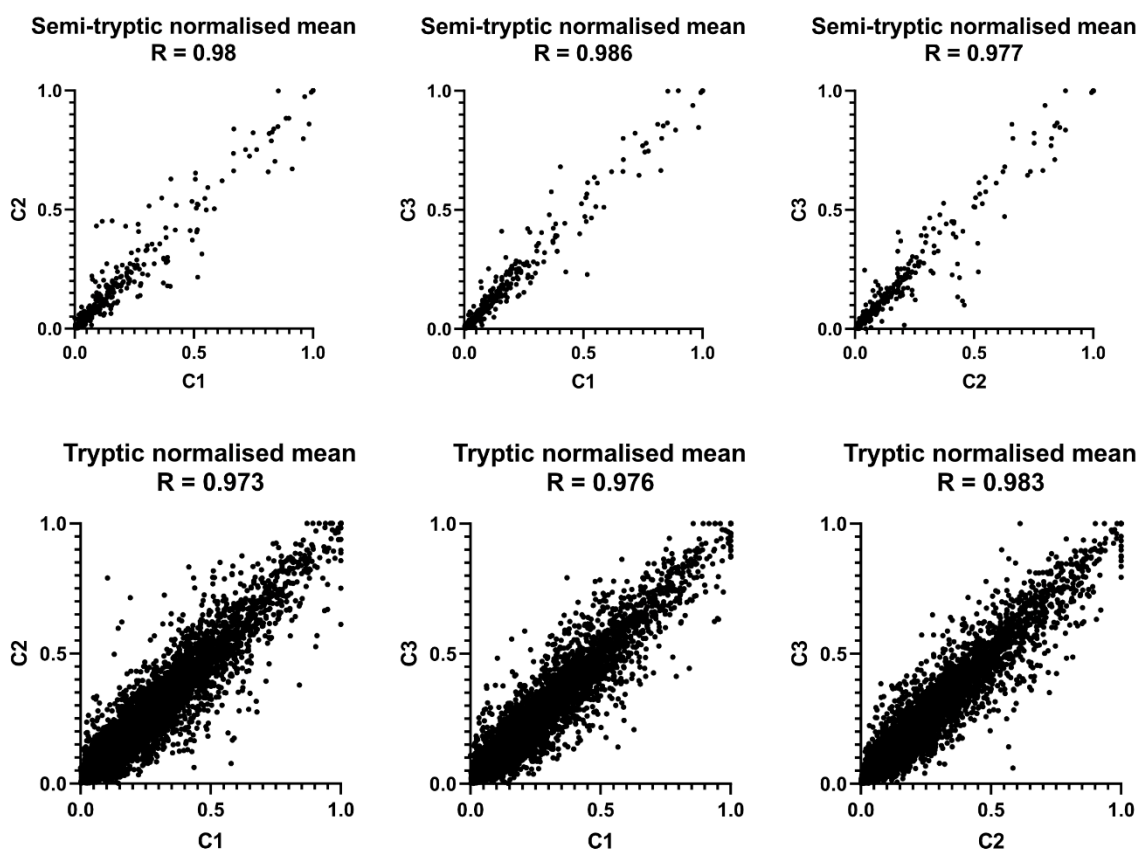


Figure 5.22: Correlation between normalised peptide intensity across experimental conditions. For each peptide, the mean normalised intensity of the 7 replicates was taken. This is then plotted against the same peptide in a different condition, as indicated by the axes. R represents Spearman's rank correlation coefficient.

Taking these findings into consideration, the LiP-MS experiment from Chapter 5 was re-analysed, this time with a semi-specific search. As before, the number of tryptic and semi-tryptic peptides identified in each condition was calculated, shown in Figure 5.23.

Similar to the 7 replicate experiment, approximately 1100 semi-tryptic peptides were identified in the control conditions, representing around 5% of all reported peptides. Furthermore, 483 semi-tryptic peptides were identified in the controls of both experiments. However, in the LiP-MS experiment of Chapter 5, the number of semi-tryptic peptides identified doubles in the samples treated with PK at an E:S of 1:2000, and increases further still at PK E:S 1:500. This is mirrored by the tryptic peptide identifications, with some 5000 fewer tryptic peptides reported in the LiP 1:500 samples.

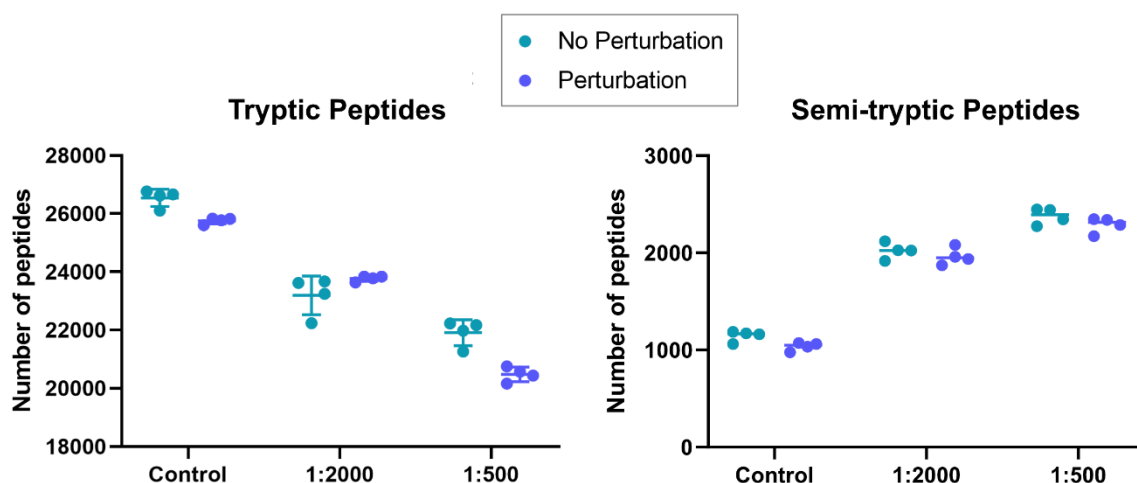


Figure 5.23: Number of tryptic and semi-tryptic peptides identified in the LiP-MS experiment of Chapter 5, processed with a semi-specific search. Identified is defined as having a non-zero reported intensity. Bars represent the mean and standard deviation of the 4 technical replicates. The experimental condition is indicated by colour (see legend) and the LiP treatment is shown on the x-axis.

Next, the contribution of semi-tryptic peptides to protein signal, and how this is changed by addition of the LiP enzyme, was investigated. In this instance, the intensities of all semi-tryptic peptides to a parent protein were summed and then divided by the total reported intensity for that protein, therefore giving a single score per protein. This is shown in Figure 5.24. As shown in the 7 replicate experiment, the signal contribution from semi-tryptic peptides is very low in the control conditions. However, also evident is an increase in the percentage signal from semi-tryptic peptides with LiP enzyme concentration, in a linear fashion (Control < 1:2000 (E1) < 1:500 (E2), Figure 5.24).

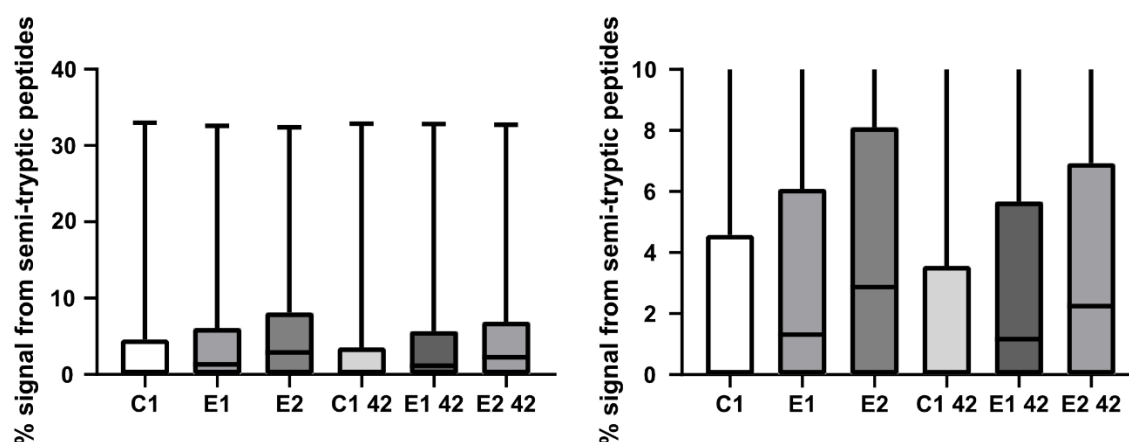


Figure 5.24: For each protein from the LiP-MS experiment of Chapter 5, the contribution of semi-tryptic peptides to the whole protein signal was calculated. Shown here is a box and whisker plot demonstrating the % protein signal (mean taken if ¼ replicates non-zero, if all replicates zero, reported as zero) from semi-tryptic peptides, split into experimental condition

and LiP treatment. At least 3 peptides (semi or tryptic) were required for a protein to be considered. The plot on the right displays the same data as the plot on the left, but with a smaller y-axis to better visualize the differences between groups.

Therefore, LiP generates more semi-tryptic peptides, these contribute more to total protein signal and this effect scales with LiP enzyme concentration. This was investigated at the individual protein level. As shown in Figure 5.25, the percentage contribution of semi-tryptic peptide to total protein signal in cytoplasmic dynein 1 heavy chain 1 (DYCH1) increases with LiP enzyme concentration, shown here for the No Perturbation condition.

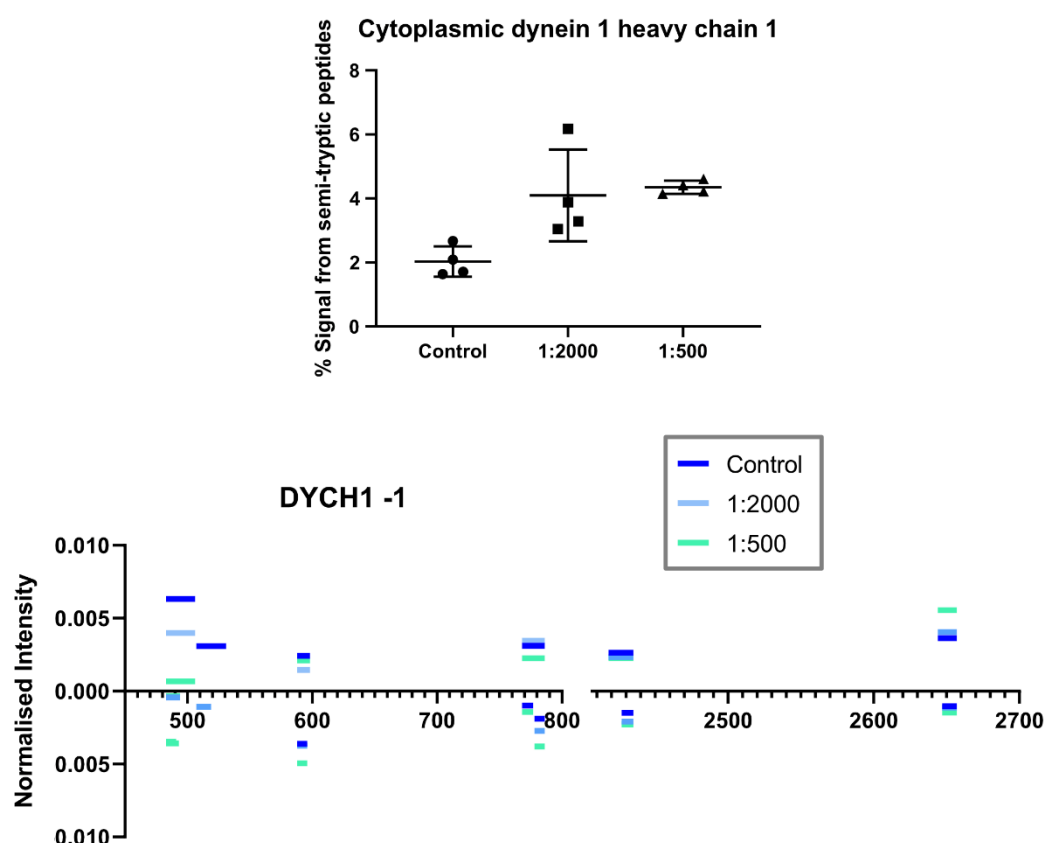


Figure 5.25: Cytoplasmic dynein 1 heavy chain 1 (DYCH1-1) normalised peptide signal. Top – percentage contribution of semi-tryptic peptides to total protein signal, by LiP treatment for No Perturbation condition. Bars show the mean and standard deviation of the 4 technical replicates. Bottom – normalised intensity of semi-tryptic peptides and their parent tryptic peptides. The x-axis displays the position of these peptides along the protein sequence and the y-axis the normalised intensity. Tryptic peptides are shown above the x-axis and semi-tryptic peptides below.

Also displayed are the semi-tryptic peptides mapping to DYCH1, for which the parent tryptic peptide was also identified. The position of these peptides within the protein sequence is displayed along the x-axis and their normalized intensity on the y-axis. Tryptic peptides are shown above the x-axis, and semi-tryptic below, for the No Perturbation condition and with the LiP treatment indicated by colour (see Legend). Demonstrated here is a reduction in normalized tryptic peptide signal with LiP concentration, with a corresponding increase in the signal contribution from the associated semi-tryptic peptide. Also demonstrated is little

change over conditions, corresponding to the peptides with a lower overall signal contribution, again representing the baseline semi-tryptic peptide signal.

Chapter Six: Conclusions and Future Work

The aim of this thesis was to develop and refine the capacity of LiP-MS to provide omics scale insight into the structural components of proteostasis in primary human cells.

A comprehensive understanding of the proteostasis network is paramount to a matched understanding of how these processes underpin human health, in particular due to the increasing prevalence of neurodegenerative diseases and the progressive loss of proteostasis with ageing. The maintenance of proteostasis is governed by dynamic changes to protein abundance, protein localization and protein structure. To gain a systems-level understanding of proteostasis and its breakdown, ideally, each of these features should be characterized at the omics scale. Therefore, this thesis was focused on the development of a technique, LiP-MS, as a proteomics method which can provide omics scale insight into the structural aspects of proteostasis in primary hMSCs.

LiP-MS was chosen over other structural proteomics techniques due to pre-existing evidence in the literature of its ability to provide physiologically relevant protein structure information in complex biological samples, at the spatial resolution required to investigate many of the dynamic post-translational remodeling events that form part of, or trigger, the proteostasis network [19]. However, to the best of the author's knowledge, at the outset of this study LiP-MS had not been applied or demonstrated at the systems-level to characterize functional processes in primary human cells.

The cells used in this study are primary bone marrow-derived hMSCs. They have the advantage of biological relevance and are well suited to studying the effect of external perturbation, such as heat shock, at the cellular level. In addition, repeated *in vitro* passaging of hMSCs induces a state of replicative senescence, which is relevant to ageing. The comparison of donor-matched early passage hMSCs to late passage, predominantly senescent hMSCs can potentially provide insight into senescence-associated discrepancies in proteostasis. Furthermore, there is a specific need to understand proteostasis in hMSCs as they are frequently used in regenerative medicine. A high dosage is often required for hMSC-based therapeutics and so, due to the small number attainable *in vivo*, *in vitro* expansion of the cells is necessary.

The first aim of this thesis was to establish the efficacy of LiP-MS in hMSCs. The meeting of this aim was contingent upon optimising aspects of the LiP-MS protocol for use in this

chosen system, focusing on the conditions of proteolysis (enzyme concentration, reaction time) and method of cell lysis. These are crucial in order to ensure that limited proteolysis is achieved, here meaning that LiP cleavage sites are dictated by protein structure, and an appropriate concentration is used where some cleavage occurs but not total degradation of the attendant proteins. This should manifest in the detected pools of proteolytic peptides, both tryptic and semi-tryptic (which are presumed to be predominantly the result of LiP-MS).

The success of this optimisation has been shown in many ways. For example, principal component analysis in Chapters 4 (Figure 4.2) and 5 (Figure 5.3) has shown a clear separation of control and LiP treated samples, with Figure 5.3 in particular showing increasing separation with increasing LiP concentration. This clearly demonstrates how LiP treatment generates a different pool of peptide signal which separates out the experimental datasets on a global basis. It was further demonstrated by differences in the number of peptides identified by LiP treatment. The manner of this separation was also consistent with the model in which increased LiP removes tryptic peptides from the pool, and increases semi-tryptic ones, with a negative correlation demonstrated between enzyme concentration and number of tryptic peptides identified (Figure 5.4(a)), and a positive correlation between enzyme concentration and the number of the semi-tryptic peptides identified (Figure 5.4(b), Figure 4.1(b)).

Moreover, evidence that these LiP sites were determined by protein structural features, and, as a corollary, that the cell lysis method used preserves protein structure, was provided both from investigation of the structural characteristics of the protein regions inhabited by LiP sites, and from investigation of the effect of external perturbation on the number and nature of LiP sites identified. For example, the LiP-MS experiment in Chapter 4 showed a substantial difference in the number of LiP peptides identified in proteome extracts treated with different heat conditions. Further to this, the LiP-MS experiment in Chapter 5 identified hundreds of proteins that were differentially accessible to the LiP protease before and after stress (Figure 5.8). The LiP peptides identified in each condition also had different structural characteristics, for example a prevalence of intrinsically disordered regions before stress (Figure 5.11) and lower protease-susceptibility scores (Nickpred) after stress (Figure 5.9).

The next aim of this thesis was to use LiP-MS to identify protein structural alterations on a proteome-wide scale. This was achieved by perturbing cell lysates with heat treatment and investigating the effect of such perturbation on the LiP-MS peptide signal. Exposure of the proteome to heat results in a shallower protein folding energy landscape and so results in a rise in protein unfolding events. Increased protein unfolding was expected to result in increased LiP enzyme activity. This experiment revealed that the number of semi-tryptic

peptides with a significant change in signal was consistently increased from the normal to heated experimental conditions, thus implying that protein unfolding can indeed be detected through LiP-MS.

The work detailed in Chapter 4 also highlighted the complications associated with semi-tryptic peptides, due to their significant presence in the control conditions. Despite a general acceptance that a baseline of semi-tryptic peptides may be present in trypsin-only proteomics experiments, due to factors such as in-instrument cleavage, trypsin miscleavage and native protease activity, their prevalence in the MS dataset detailed in Chapter 4 was hard to explain, with a subset of semi-tryptic peptides displaying significantly higher signal in the control sample, when compared to the LiP treated sample. A suggested explanation for this was that, if there is indeed a baseline level of semi-tryptic peptides in all samples, additional PK cleavage could result in the generation of peptides that are either non-tryptic or too small for identification by MS, and this would be observed as a loss of semi-tryptic peptide signal with LiP treatment (Figure 4.7). However, if this represents the level of noise expected in semi-tryptic peptide analysis, the confidence with which LiP cleavage sites can be identified against this background signal is significantly impaired.

The impact of search engine choice was determined through re-running the LiP-MS data through MaxQuant (previously, Mascot was used). This reduced the prevalence of reported semi-tryptic peptides in the dataset whilst maintaining the overall conclusions of the study; namely, that LiP activity increases with heat treatment and is expected to be due to additional unfolding. In fact, the number of semi-tryptic peptides found to be significantly changing with heat treatment increased substantially, with a high proportion found to be unique to the LiP condition. Despite this, a subset of semi-tryptic peptides still displayed significantly higher signal in the control sample.

Finally, the “baseline” presence of semi-tryptic peptides in trypsin-only (LiP-MS control) experiments, and how this may be impacted by different experimental conditions, was investigated using the seven replicate experiment conducted using an immortalized hMSC cell line (Chapter 5: Addendum). Here, an average of 1150 semi-tryptic peptides were identified, representing approximately 5% of all peptides and this was not significantly affected by different experimental conditions (Figure 5.19). Semi-tryptic peptides were found to have significantly greater variance than their tryptic counterparts (Figure 5.20), and contribute significantly less to total protein signal (Figure 5.21). However, the contribution of semi-tryptic peptides to total protein signal did linearly increase with application of LiP and LiP enzyme concentration (Figure 5.24), consistent with expectation that increased proteinase K presence generates more non-tryptic cleavage activity.

Overall, through comparing LiP-treated samples to internal controls (trypsin only) and including strict statistical filtering of abundance changes (Benjamini Hochberg multiple testing correction, $q < 0.05$), the global trends of semi-tryptic peptides in LiP-MS align with the LiP fingerprint model and can identify differences between perturbed samples. However, their signal is much noisier than their tryptic counterparts. Therefore, for the LiP-MS experiment conducted in Chapter 5, analysis was focused primarily on tryptic peptides.

The final aim of this thesis was to use LiP-MS to provide a global structural readout of the response to stress in hMSCs. Stress was induced through the application of heat shock and LiP-MS was performed on the perturbed and unperturbed cells. Relative quantification of protein abundance revealed no significant stress-induced changes, aligning with previous data produced in the Swift lab which indicates the role of protein in the early stress response is predominantly post-translational [10]. However, statistical analysis of the LiP peptide data led to the identification of hundreds of proteins with potential stress-induced protein structural alterations.

Examination of the structural properties of these peptides using Nickpred revealed significantly lower protease susceptibility scores in a subset of this data, suggesting that heat-induced protein unfolding has been captured in the stressed cells. Moreover, GO enrichment analysis of structurally altered proteins found a significant enrichment for many known cellular stress response pathways such as “protein folding chaperone” and “protein import into nucleus”.

This indicates that, despite the absence of significant changes in protein abundance in the cells, LiP-MS is able to capture the dynamics of the cellular response to stress through the identification of protein conformational alterations that form a crucial part of the proteostasis network. Furthermore, a subset of LiP peptides that were differentially accessible before and after stress mapped to phosphorylation sites, a PTM that is functionally critical and plays a substantial role in the cellular response to stress. Stress response factors highlighted in the GO enrichment analysis were then further investigated through LiP peptide mapping, revealing stress-induced structural alterations (differential protease accessibility) at PTM sites, ATP binding sites and, in the case of molecular chaperones, client-protein binding sites.

An alternative behavior highlighted in this experiment was that of significantly higher tryptic peptide abundance in the LiP treated samples compared to the controls (~400 peptides), which is the opposite of the conventional LiP peptide fingerprint. A working hypothesis consistent with this observation was that this may be indicative of protein aggregation or protein complex formation. This hypothesis was supported somewhat by analysis of the

intrinsic structural properties of the highlighted proteins and protein regions, with evidence of decreased disorder (Figure 5.11) and protein solubility (Figure 5.12). Furthermore, the presence of significantly changing peptides with this behaviour was increased twentyfold after stress. This supports the hypothesis, as heat stress is known to induce protein aggregation and thus will likely increase the prevalence of proteolytic-resistant protein aggregates. Finally, these proteins were over-represented as complex components, with stress response-associated protein complexes highlighted.

Historically, LiP has most frequently been used to assess the structure of individual proteins, either in purified samples or lysates in which the protein of interest has been spiked [161, 162, 165]. More recent advances saw the technique act as a proteome-wide probe of protein structure. For example, LiP-MS was used to investigate protein thermostability through heating cell lysates to increasing temperatures and quantification of the corresponding LiP signal [19, 166].

The most comprehensive LiP-MS study (published within the time frame of research conducted for this thesis) demonstrated the efficacy of the technique in probing the proteostatic response to thermal and oxidative stress in yeast and nutrient adaption in *E.coli* [167]. Analysis of the protein abundance data revealed little change, whilst many LiP peptides were identified, mapping to hundreds of proteins with potential stress-induced structural alterations, just as in the work conducted in this thesis. Also similarly, functional enrichment analysis performed for these proteins highlighted many pathways known to be correlated with yeast osmotic or heat stress. This was followed by comparison of LiP peptides with reported phosphosites, identifying a subset of LiP peptides that overlapped or were in proximity to (within 10 amino acids) of phosphopeptides. Finally, the activation of molecular chaperones was inferred through peptide mapping to the substrate or ATP binding site of disaggregase chaperone Hsp104.

The work conducted in this thesis, through protocol optimization and robust data handling, has expanded this principle further, now demonstrating the capacity of LiP-MS to provide structural insights into the stress response in primary human cells.

After the research conducted in this thesis was concluded, a study was published by the Picotti lab using LiP-MS to characterize structural changes to the proteome in the cerebrospinal fluid (CSF) of Parkinson's Disease (PD) patients [225]. LiP-MS analysis was performed on CSF from a well characterized cohort of patients with PD and an age matched cohort of healthy donors. In this instance, LiP peptides were identified using a multiple linear regression model, including coefficients to account for the variability resultant from age, sex, protein abundance, tryptic-only peptide abundance and cohort membership. This highlighted









88 LiP peptides from 76 parent proteins (out of a total 7144 peptides from 859 proteins) with a significant cohort effect (corrected p value < 0.05). GO enrichment analysis of these proteins revealed they were significantly enriched for processes that have been typically associated with PD, such as synapse maintenance and acetylcholine metabolism [226].

Furthermore, the CSF analysis was complemented by LiP-MS of post-mortem brain tissue samples from a separate, smaller cohort of 5 PD patients and 5 healthy controls. Of the 76 LiP proteins highlighted in CSF, 27 were also detected in the brain, with 16 also shown to have LiP-MS-indicated structural alterations between control and PD. Of the 88 LiP peptides highlighted in CSF, 11 were also detected in the brain and 8 showed similar patterns between control and PD. Further, three LiP proteins detected in the brain data were linked to PD associated genes, for example, apolipoprotein E (APOE) [227].

Another avenue of investigation explored in this study was to characterize the variability of the human CSF structural proteome. Overall, 6.5% of the proteins detected had at least one region that displayed high levels of structural variability across healthy patients. This was unique from changes due to sequence alterations, cleavages or PTMs and the structurally altered proteins were from a variety of cellular functional pathways.

Contrary to previous LiP-MS experiments performed in yeast and *E.coli*, the differential peptide abundance analysis in this study was performed only using fully tryptic peptides. The justification provided for this was that fully tryptic peptides had “lower noise than that of half-tryptic peptides”, but no more information was provided. This is supportive of the conclusions drawn in this thesis regarding the unreliability of semi-tryptic peptides in LiP-MS of human samples. This is exciting work that aligns with the conclusions of this thesis and recommendations from this paper can be incorporated into future work. Indeed, an interesting continuation of the work in this thesis would be to use LiP-MS to characterize structural proteome changes in early passage and senescent hMSCs. An outline of this future experiment is shown in Table 6.1.

Table 6.1: Future experimental design. Donor-matched early and late passage hMSCs would be subjected to heat stress (indicated in this schematic through the red/orange shading). Each sample of each condition would be subject to (a) LiP treatment followed by tryptic digestion or (b) tryptic digestion only (Control).

	Early Passage		Late Passage	
Control				
LiP				

Donor-matched early and late passage (replicative senescence) hMSCs from multiple donors would be subject to heat shock, with unstressed and stressed cells from each condition then either LiP treated or digested only with trypsin (control) before analysis by MS. LiP peptides in this instance would be identified through multiple regression analysis in order to account for donor to donor variance, as in the LiP-MS PD paper discussed above [225] and in previous work with hMSCs in the Swift lab [10]. Taking into account the lessons learned from this thesis, only tryptic peptides would be used in the LiP peptide analysis.

This would enable multiple structural comparisons. For example, LiP-MS of unstressed early passage and senescent hMSCs would enable assessment of both inter-donor structural proteome variability and senescence-induced alterations to native protein structure. Furthermore, one could compare the structural implications of stress in early passage and senescent cells, both with respect to protein unfolding and aggregation events and with respect to the stress response, as characterized through conformational and abundance changes in components of the proteostasis network.

Moving forward, LiP-MS has the potential to provide fresh insight into the mechanisms of the cellular response to stress through identifying previously undescribed, functionally critical protein structural changes, the presence and consequences of which can then be further investigated using orthogonal biochemical techniques.

Bibliography

1. Klaips, C.L., G.G. Jayaraj, and F.U. Hartl, *Pathways of cellular proteostasis in aging and disease*. The Journal of Cell Biology, 2018. **217**(1): p. 51-63.
2. Brownridge, P., C. Lawless, A.B. Payapilly, K. Lanthaler, S.W. Holman, V.M. Harman, C.M. Grant, R.J. Beynon, and S.J. Hubbard, *Quantitative analysis of chaperone network throughput in budding yeast*. Proteomics, 2013. **13**(8): p. 1276-91.
3. Lecker, S.H., A.L. Goldberg, and W.E. Mitch, *Protein Degradation by the Ubiquitin–Proteasome Pathway in Normal and Disease States*. Journal of the American Society of Nephrology, 2006. **17**(7): p. 1807-1819.
4. Glick, D., S. Barth, and K.F. Macleod, *Autophagy: cellular and molecular mechanisms*. J Pathol, 2010. **221**(1): p. 3-12.
5. Höhn, A., D. Weber, T. Jung, C. Ott, M. Hugo, B. Kochlik, R. Kehm, J. König, T. Grune, and J.P. Castro, *Happily (n)ever after: Aging in the context of oxidative stress, proteostasis loss and cellular senescence*. Redox Biol, 2017. **11**: p. 482-501.
6. Kikis, E.A., T. Gidalevitz, and R.I. Morimoto, *Protein homeostasis in models of aging and age-related conformational disease*. Adv Exp Med Biol, 2010. **694**: p. 138-59.
7. Kaushik, S. and A.M. Cuervo, *Proteostasis and aging*. Nat Med, 2015. **21**(12): p. 1406-15.
8. Labbadia, J. and R.I. Morimoto, *The biology of proteostasis in aging and disease*. Annu Rev Biochem, 2015. **84**: p. 435-64.
9. Cummings, J.L., T. Morstorf, and K. Zhong, *Alzheimer's disease drug-development pipeline: few candidates, frequent failures*. Alzheimers Res Ther, 2014. **6**(4): p. 37.
10. Llewellyn, J., V. Mallikarjun, E. Appleton, M. Osipova, H.T.J. Gilbert, S.M. Richardson, S.J. Hubbard, and J. Swift, *Loss of regulation of protein synthesis and turnover underpins an attenuated stress response in senescent human mesenchymal stem cells*. Proc Natl Acad Sci U S A, 2023. **120**(14): p. e2210745120.
11. Danielyan, L., S. Beer-Hammer, A. Stolzing, R. Schafer, G. Siegel, C. Fabian, P. Kahle, T. Biedermann, A. Lourhmati, M. Buadze, A. Novakovic, B. Proksch, C.H. Gleiter, W.H. Frey, and M. Schwab, *Intranasal delivery of bone marrow-derived mesenchymal stem cells, macrophages, and microglia to the brain in mouse models of Alzheimer's and Parkinson's disease*. Cell Transplant, 2014. **23 Suppl 1**: p. S123-39.
12. Danielyan, L., R. Schäfer, A. von Ameln-Mayerhofer, F. Bernhard, S. Verleysdonk, M. Buadze, A. Lourhmati, T. Klopfer, F. Schaumann, B. Schmid, C. Koehle, B. Proksch, R. Weissert, H.M. Reichardt, J. van den Brandt, G.H. Buniatian, M. Schwab, C.H. Gleiter, and W.H. Frey, 2nd, *Therapeutic efficacy of intranasally delivered mesenchymal stem cells in a rat model of Parkinson disease*. Rejuvenation Res, 2011. **14**(1): p. 3-16.
13. Yu-Taeger, L., J. Stricker-Shaver, K. Arnold, P. Bambynek-Dziuk, A. Novati, E. Singer, A. Lourhmati, C. Fabian, J. Magg, O. Riess, M. Schwab, A. Stolzing, L. Danielyan, and H.H.P. Nguyen, *Intranasal Administration of Mesenchymal Stem Cells Ameliorates the Abnormal Dopamine Transmission System and Inflammatory Reaction in the R6/2 Mouse Model of Huntington Disease*. Cells, 2019. **8**(6).

14. Morimoto, R.I., *Proteotoxic stress and inducible chaperone networks in neurodegenerative disease and aging*. Genes Dev, 2008. **22**(11): p. 1427-38.
15. Kabat, M., I. Bobkov, S. Kumar, and M. Grumet, *Trends in mesenchymal stem cell clinical trials 2004-2018: Is efficacy optimal in a narrow dose range?* Stem Cells Transl Med, 2020. **9**(1): p. 17-27.
16. Gao, F., S.M. Chiu, D.A. Motan, Z. Zhang, L. Chen, H.L. Ji, H.F. Tse, Q.L. Fu, and Q. Lian, *Mesenchymal stem cells and immunomodulation: current status and future prospects*. Cell Death Dis, 2016. **7**: p. e2062.
17. Herbig, U., M. Ferreira, L. Condel, D. Carey, and J.M. Sedivy, *Cellular senescence in aging primates*. Science, 2006. **311**(5765): p. 1257.
18. Wang, C., D. Jurk, M. Maddick, G. Nelson, C. Martin-Ruiz, and T. von Zglinicki, *DNA damage response and cellular senescence in tissues of aging mice*. Aging Cell, 2009. **8**(3): p. 311-23.
19. Schopper, S., A. Kahraman, P. Leuenberger, Y. Feng, I. Piazza, O. Muller, P.J. Boersema, and P. Picotti, *Measuring protein structural changes on a proteome-wide scale using limited proteolysis-coupled mass spectrometry*. Nat Protoc, 2017. **12**(11): p. 2391-2410.
20. Anfinsen, C.B., *Principles that govern the folding of protein chains*. Science, 1973. **181**(4096): p. 223-30.
21. Raschke, T.M., *Water structure and interactions with protein surfaces*. Curr Opin Struct Biol, 2006. **16**(2): p. 152-9.
22. Hartl, F.U. and M. Hayer-Hartl, *Molecular chaperones in the cytosol: from nascent chain to folded protein*. Science, 2002. **295**(5561): p. 1852-8.
23. Milo, R., *What is the total number of protein molecules per cell volume? A call to rethink some published values*. Bioessays, 2013. **35**(12): p. 1050-5.
24. Ellis, R.J., *Protein misassembly: macromolecular crowding and molecular chaperones*. Adv Exp Med Biol, 2007. **594**: p. 1-13.
25. Dobson, C.M., *Protein folding and misfolding*. Nature, 2003. **426**(6968): p. 884-90.
26. Hartl, F.U., A. Bracher, and M. Hayer-Hartl, *Molecular chaperones in protein folding and proteostasis*. Nature, 2011. **475**(7356): p. 324-32.
27. Englander, S.W. and L. Mayne, *The nature of protein folding pathways*. Proc Natl Acad Sci U S A, 2014. **111**(45): p. 15873-80.
28. Oldfield, C.J., Y. Cheng, M.S. Cortese, C.J. Brown, V.N. Uversky, and A.K. Dunker, *Comparing and combining predictors of mostly disordered proteins*. Biochemistry, 2005. **44**(6): p. 1989-2000.
29. Lynch, M., *The evolution of multimeric protein assemblages*. Mol Biol Evol, 2012. **29**(5): p. 1353-66.
30. Ardito, F., M. Giuliani, D. Perrone, G. Troiano, and L. Lo Muzio, *The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (Review)*. Int J Mol Med, 2017. **40**(2): p. 271-280.
31. Milroy, L.G., T.N. Grossmann, S. Hennig, L. Brunsveld, and C. Ottmann, *Modulators of protein-protein interactions*. Chem Rev, 2014. **114**(9): p. 4695-748.
32. Clare, D.K. and H.R. Saibil, *ATP-driven molecular chaperone machines*. Biopolymers, 2013. **99**(11): p. 846-59.
33. Lindsley, J.E. and J. Rutter, *Whence cometh the allosterome?* Proceedings of the National Academy of Sciences, 2006. **103**(28): p. 10533-10535.
34. Itzhak, D.N., S. Tyanova, J. Cox, and G.H.H. Borner, *Global, quantitative and dynamic mapping of protein subcellular localization*. eLife, 2016. **5**: p. e16950.

35. Alexander, P., Z. Eldar, P. Shiri, and S. Rony, *The MAPK cascades: Signaling components, nuclear roles and mechanisms of nuclear translocation*. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research, 2011. **1813**(9): p. 1619-1633.
36. Advani, V.M. and P. Ivanov, *Translational Control under Stress: Reshaping the Translatome*. Bioessays, 2019. **41**(5): p. e1900009.
37. Adomavicius, T., M. Guaita, Y. Zhou, M.D. Jennings, Z. Latif, A.M. Roseman, and G.D. Pavitt, *The structural basis of translational control by eIF2 phosphorylation*. Nat Commun, 2019. **10**(1): p. 2136.
38. Pavitt, G.D., *Regulation of translation initiation factor eIF2B at the hub of the integrated stress response*. Wiley Interdiscip Rev RNA, 2018. **9**(6): p. e1491.
39. Pakos-Zebrucka, K., I. Koryga, K. Mnich, M. Ljubic, A. Samali, and A.M. Gorman, *The integrated stress response*. EMBO Rep, 2016. **17**(10): p. 1374-1395.
40. Park, S., Y. Lim, D. Lee, R. Elvira, J.M. Lee, M.R. Lee, and J. Han, *Modulation of Protein Synthesis by eIF2 α Phosphorylation Protects Cell from Heat Stress-Mediated Apoptosis*. Cells, 2018. **7**(12).
41. Protter, D.S.W. and R. Parker, *Principles and Properties of Stress Granules*. Trends Cell Biol, 2016. **26**(9): p. 668-679.
42. Youn, J.Y., B.J.A. Dyakov, J. Zhang, J.D.R. Knight, R.M. Vernon, J.D. Forman-Kay, and A.C. Gingras, *Properties of Stress Granule and P-Body Proteomes*. Mol Cell, 2019. **76**(2): p. 286-294.
43. Kershaw, C.J., M.G. Nelson, J. Lui, C.P. Bates, M.D. Jennings, S.J. Hubbard, M.P. Ashe, and C.M. Grant, *Integrated multi-omics reveals common properties underlying stress granule and P-body formation*. RNA Biol, 2021. **18**(sup2): p. 655-673.
44. Morales-Polanco, F., C. Bates, J. Lui, J. Casson, C.A. Solari, M. Pizzinga, G. Forte, C. Griffin, K.E.L. Garner, H.E. Burt, H.L. Dixon, S. Hubbard, P. Portela, and M.P. Ashe, *Core Fermentation (CoFe) granules focus coordinated glycolytic mRNA localization and translation to fuel glucose fermentation*. iScience, 2021. **24**(2): p. 102069.
45. Brehme, M., C. Voisine, T. Rolland, S. Wachi, J.H. Soper, Y. Zhu, K. Orton, A. Vilella, D. Garza, M. Vidal, H. Ge, and R.I. Morimoto, *A chaperome subnetwork safeguards proteostasis in aging and neurodegenerative disease*. Cell Rep, 2014. **9**(3): p. 1135-50.
46. Radons, J., *The human HSP70 family of chaperones: where do we stand?* Cell Stress Chaperones, 2016. **21**(3): p. 379-404.
47. Balchin, D., M. Hayer-Hartl, and F.U. Hartl, *Recent advances in understanding catalysis of protein folding by molecular chaperones*. FEBS Letters, 2020. **594**(17): p. 2770-2781.
48. Hoter, A., M.E. El-Sabban, and H.Y. Naim, *The HSP90 Family: Structure, Regulation, Function, and Implications in Health and Disease*. Int J Mol Sci, 2018. **19**(9).
49. Schopf, F.H., M.M. Biebl, and J. Buchner, *The HSP90 chaperone machinery*. Nat Rev Mol Cell Biol, 2017. **18**(6): p. 345-360.
50. Levy-Rimler, G., P. Viitanen, C. Weiss, R. Sharkia, A. Greenberg, A. Niv, A. Lustig, Y. Delarea, and A. Azem, *The effect of nucleotides and mitochondrial chaperonin 10 on the structure and chaperone activity of mitochondrial chaperonin 60*. Eur J Biochem, 2001. **268**(12): p. 3465-72.
51. Christoph, S., S.M. Anne, R. Stefanie, and F. Judith, *Mechanism of the eukaryotic chaperonin: protein folding in the chamber of secrets*. Trends in Cell Biology, 2004. **14**(11): p. 598-604.

52. Thulasiraman, V., C.F. Yang, and J. Frydman, *In vivo newly translated polypeptides are sequestered in a protected folding environment*. *Embo j*, 1999. **18**(1): p. 85-95.
53. Gestaut, D., S.H. Roh, B. Ma, G. Pintilie, L.A. Joachimiak, A. Leitner, T. Walzthoeni, R. Aebersold, W. Chiu, and J. Frydman, *The Chaperonin TRiC/CCT Associates with Prefoldin through a Conserved Electrostatic Interface Essential for Cellular Proteostasis*. *Cell*, 2019. **177**(3): p. 751-765.e15.
54. Webster, J.M., A.L. Darling, V.N. Uversky, and L.J. Blair, *Small Heat Shock Proteins, Big Impact on Protein Aggregation in Neurodegenerative Disease*. *Front Pharmacol*, 2019. **10**: p. 1047.
55. Mogk, A., C. Ruger-Herreros, and B. Bukau, *Cellular Functions and Mechanisms of Action of Small Heat Shock Proteins*. *Annu Rev Microbiol*, 2019. **73**: p. 89-110.
56. Mogk, A., E. Kummer, and B. Bukau, *Cooperation of Hsp70 and Hsp100 chaperone machines in protein disaggregation*. *Frontiers in Molecular Biosciences*, 2015. **2**.
57. Franco, A., L. Velasco-Carneros, N. Alvarez, N. Orozco, F. Moro, A. Prado, and A. Muga, *Unzipping the Secrets of Amyloid Disassembly by the Human Disaggregase*. *Cells*, 2021. **10**(10): p. 2745.
58. Sousa, R., *Structural mechanisms of chaperone mediated protein disaggregation*. *Frontiers in Molecular Biosciences*, 2014. **1**.
59. Takaki, E. and A. Nakai, *Regulation of HSF Activation and Repression*, in *Heat Shock Factor*, A. Nakai, Editor. 2016, Springer Japan: Tokyo. p. 51-72.
60. Sivéry, A., E. Courtade, and Q. Thommen, *A minimal titration model of the mammalian dynamical heat shock response*. *Phys Biol*, 2016. **13**(6): p. 066008.
61. Zheng, X., J. Krakowiak, N. Patel, A. Beyzavi, J. Ezike, A.S. Khalil, and D. Pincus, *Dynamic control of Hsf1 during heat shock by a chaperone switch and phosphorylation*. *Elife*, 2016. **5**.
62. Li, Y., S. Li, and H. Wu, *Ubiquitination-Proteasome System (UPS) and Autophagy Two Main Protein Degradation Machineries in Response to Cell Stress*. *Cells*, 2022. **11**(5).
63. Yim, W.W. and N. Mizushima, *Lysosome biology in autophagy*. *Cell Discov*, 2020. **6**: p. 6.
64. Hipp, M.S., S.H. Park, and F.U. Hartl, *Proteostasis impairment in protein-misfolding and -aggregation diseases*. *Trends Cell Biol*, 2014. **24**(9): p. 506-14.
65. Douglas, P.M. and A. Dillin, *Protein homeostasis and aging in neurodegeneration*. *J Cell Biol*, 2010. **190**(5): p. 719-29.
66. Becker, R.E., N.H. Greig, and E. Giacobini, *Why do so many drugs for Alzheimer's disease fail in development? Time for new methods and new practices?* *J Alzheimers Dis*, 2008. **15**(2): p. 303-25.
67. Anderson, R.M., C. Hadjichrysanthou, S. Evans, and M.M. Wong, *Why do so many clinical trials of therapies for Alzheimer's disease fail?* *Lancet*, 2017. **390**(10110): p. 2327-2329.
68. Calderwood, S.K., A. Murshid, and T. Prince, *The shock of aging: molecular chaperones and the heat shock response in longevity and aging--a mini-review*. *Gerontology*, 2009. **55**(5): p. 550-8.
69. Soti, C. and P. Csermely, *Molecular chaperones and the aging process*. *Biogerontology*, 2000. **1**(3): p. 225-33.
70. Soti, C. and P. Csermely, *Aging and molecular chaperones*. *Exp Gerontol*, 2003. **38**(10): p. 1037-40.

71. Hall, D.M., L. Xu, V.J. Drake, L.W. Oberley, T.D. Oberley, P.L. Moseley, and K.C. Kregel, *Aging reduces adaptive capacity and stress protein expression in the liver after heat stress*. J Appl Physiol (1985), 2000. **89**(2): p. 749-59.
72. Fargnoli, J., T. Kunisada, A.J. Fornace, Jr., E.L. Schneider, and N.J. Holbrook, *Decreased expression of heat shock protein 70 mRNA and protein after heat treatment in cells of aged rats*. Proc Natl Acad Sci U S A, 1990. **87**(2): p. 846-50.
73. Ambra, R., 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*. Exp Gerontol, 2004. **39**(10): p. 1475-84.
74. Hsu, A.L., C.T. Murphy, and C. Kenyon, *Regulation of aging and age-related disease by DAF-16 and heat-shock factor*. Science, 2003. **300**(5622): p. 1142-5.
75. Balch, W.E., R.I. Morimoto, A. Dillin, and J.W. Kelly, *Adapting proteostasis for disease intervention*. Science, 2008. **319**(5865): p. 916-9.
76. Chen, G., Y. Andrade-Talavera, S. Tambaro, A. Leppert, H.E. Nilsson, X. Zhong, M. Landreh, P. Nilsson, H. Hebert, H. Biverstål, A. Fisahn, A. Abelein, and J. Johansson, *Augmentation of Bri2 molecular chaperone activity against amyloid- β reduces neurotoxicity in mouse hippocampus in vitro*. Commun Biol, 2020. **3**(1): p. 32.
77. Cohen, E., J. Bieschke, R.M. Perciavalle, J.W. Kelly, and A. Dillin, *Opposing activities protect against age-onset proteotoxicity*. Science, 2006. **313**(5793): p. 1604-10.
78. Santra, M., K.A. Dill, and A.M.R. de Graff, *Proteostasis collapse is a driver of cell aging and death*. Proceedings of the National Academy of Sciences, 2019. **116**(44): p. 22173-22178.
79. Vernace, V.A., L. Arnaud, T. Schmidt-Glenewinkel, and M.E. Figueiredo-Pereira, *Aging perturbs 26S proteasome assembly in Drosophila melanogaster*. Faseb j, 2007. **21**(11): p. 2672-82.
80. Chondrogianni, N., F.L. Stratford, I.P. Trougakos, B. Friguet, A.J. Rivett, and E.S. Gonos, *Central role of the proteasome in senescence and survival of human fibroblasts: induction of a senescence-like phenotype upon its inhibition and resistance to stress upon its activation*. J Biol Chem, 2003. **278**(30): p. 28026-37.
81. Torres, C., L. Lewis, and V.J. Cristofalo, *Proteasome inhibitors shorten replicative life span and induce a senescent-like phenotype of human fibroblasts*. J Cell Physiol, 2006. **207**(3): p. 845-53.
82. Perez, V.I., R. Buffenstein, V. Masamsetti, S. Leonard, A.B. Salmon, J. Mele, B. Andziak, T. Yang, Y. Edrey, B. Friguet, W. Ward, A. Richardson, and A. Chaudhuri, *Protein stability and resistance to oxidative stress are determinants of longevity in the longest-living rodent, the naked mole-rat*. Proc Natl Acad Sci U S A, 2009. **106**(9): p. 3059-64.
83. Rodriguez, K.A., Y.H. Edrey, P. Osmulski, M. Gaczynska, and R. Buffenstein, *Altered composition of liver proteasome assemblies contributes to enhanced proteasome activity in the exceptionally long-lived naked mole-rat*. PLoS One, 2012. **7**(5): p. e35890.
84. Chondrogianni, N., I. Petropoulos, C. Franceschi, B. Friguet, and E.S. Gonos, *Fibroblast cultures from healthy centenarians have an active proteasome*. Exp Gerontol, 2000. **35**(6-7): p. 721-8.

85. Barbosa, M.C., R.A. Grosso, and C.M. Fader, *Hallmarks of Aging: An Autophagic Perspective*. Front Endocrinol (Lausanne), 2018. **9**: p. 790.
86. Yan, Y. and T. Finkel, *Autophagy as a regulator of cardiovascular redox homeostasis*. Free Radic Biol Med, 2017. **109**: p. 108-113.
87. Cuervo, A.M. and J.F. Dice, *Age-related decline in chaperone-mediated autophagy*. J Biol Chem, 2000. **275**(40): p. 31505-13.
88. Pyo, J.O., S.M. Yoo, H.H. Ahn, J. Nah, S.H. Hong, T.I. Kam, S. Jung, and Y.K. Jung, *Overexpression of Atg5 in mice activates autophagy and extends lifespan*. Nat Commun, 2013. **4**: p. 2300.
89. Emanuele, E., P. Minoretti, F. Sanchis-Gomar, H. Pareja-Galeano, Y. Yilmaz, N. Garatachea, and A. Lucia, *Can enhanced autophagy be associated with human longevity? Serum levels of the autophagy biomarker beclin-1 are increased in healthy centenarians*. Rejuvenation Res, 2014. **17**(6): p. 518-24.
90. Dange, T., D. Smith, T. Noy, P.C. Rommel, L. Jurzitza, R.J. Cordero, A. Legendre, D. Finley, A.L. Goldberg, and M. Schmidt, *Bim10 protein promotes proteasomal substrate turnover by an active gating mechanism*. J Biol Chem, 2011. **286**(50): p. 42830-9.
91. Saez, I. and D. Vilchez, *The Mechanistic Links Between Proteasome Activity, Aging and Age-related Diseases*. Curr Genomics, 2014. **15**(1): p. 38-51.
92. Keck, S., R. Nitsch, T. Grune, and O. Ullrich, *Proteasome inhibition by paired helical filament-tau in brains of patients with Alzheimer's disease*. J Neurochem, 2003. **85**(1): p. 115-22.
93. Hipp, M.S., C.N. Patel, K. Bersuker, B.E. Riley, S.E. Kaiser, T.A. Shaler, M. Brandeis, and R.R. Kopito, *Indirect inhibition of 26S proteasome activity in a cellular model of Huntington's disease*. J Cell Biol, 2012. **196**(5): p. 573-87.
94. van Deursen, J.M., *The role of senescent cells in ageing*. Nature, 2014. **509**(7501): p. 439-46.
95. Carrard, G., M. Dieu, M. Raes, O. Toussaint, and B. Friguet, *Impact of ageing on proteasome structure and function in human lymphocytes*. Int J Biochem Cell Biol, 2003. **35**(5): p. 728-39.
96. Ogrodnik, M., S. Miwa, T. Tchkonja, D. Tiniakos, C.L. Wilson, A. Lahat, C.P. Day, A. Burt, A. Palmer, Q.M. Anstee, S.N. Grellscheid, J.H.J. Hoeijmakers, S. Barnhoorn, D.A. Mann, T.G. Bird, W.P. Vermeij, J.L. Kirkland, J.F. Passos, T. von Zglinicki, and D. Jurk, *Cellular senescence drives age-dependent hepatic steatosis*. Nat Commun, 2017. **8**: p. 15691.
97. Petersen, K.F., D. Befroy, S. Dufour, J. Dziura, C. Ariyan, D.L. Rothman, L. DiPietro, G.W. Cline, and G.I. Shulman, *Mitochondrial dysfunction in the elderly: possible role in insulin resistance*. Science, 2003. **300**(5622): p. 1140-2.
98. Passos, J.F., G. Saretzki, S. Ahmed, G. Nelson, T. Richter, H. Peters, I. Wappler, M.J. Birket, G. Harold, K. Schaeuble, M.A. Birch-Machin, T.B. Kirkwood, and T. von Zglinicki, *Mitochondrial dysfunction accounts for the stochastic heterogeneity in telomere-dependent senescence*. PLoS Biol, 2007. **5**(5): p. e110.
99. López-Otín, C., M.A. Blasco, L. Partridge, M. Serrano, and G. Kroemer, *The hallmarks of aging*. Cell, 2013. **153**(6): p. 1194-217.
100. Phipps, S.M., J.B. Berletch, L.G. Andrews, and T.O. Tollefsbol, *Aging cell culture: methods and observations*. Methods Mol Biol, 2007. **371**: p. 9-19.

101. Hoang, D.M., P.T. Pham, T.Q. Bach, A.T.L. Ngo, Q.T. Nguyen, T.T.K. Phan, G.H. Nguyen, P.T.T. Le, V.T. Hoang, N.R. Forsyth, M. Heke, and L.T. Nguyen, *Stem cell-based therapy for human diseases*. Signal Transduct Target Ther, 2022. **7**(1): p. 272.
102. Choi, J.R., K.W. Yong, and J.Y. Choi, *Effects of mechanical loading on human mesenchymal stem cells for cartilage tissue engineering*. J Cell Physiol, 2018. **233**(3): p. 1913-1928.
103. Bornes, T.D., A.B. Adesida, and N.M. Jomha, *Mesenchymal stem cells in the treatment of traumatic articular cartilage defects: a comprehensive review*. Arthritis Res Ther, 2014. **16**(5): p. 432.
104. Yamasaki, S., H. Mera, M. Itokazu, Y. Hashimoto, and S. Wakitani, *Cartilage Repair With Autologous Bone Marrow Mesenchymal Stem Cell Transplantation: Review of Preclinical and Clinical Studies*. Cartilage, 2014. **5**(4): p. 196-202.
105. Zhang, Y.T., K.J. He, J.B. Zhang, Q.H. Ma, F. Wang, and C.F. Liu, *Advances in intranasal application of stem cells in the treatment of central nervous system diseases*. Stem Cell Res Ther, 2021. **12**(1): p. 210.
106. Soria, B., A. Martin-Montalvo, Y. Aguilera, N. Mellado-Damas, J. López-Beas, I. Herrera-Herrera, E. López, J.A. Barcia, M. Alvarez-Dolado, A. Hmadcha, and V. Capilla-González, *Human Mesenchymal Stem Cells Prevent Neurological Complications of Radiotherapy*. Front Cell Neurosci, 2019. **13**: p. 204.
107. Mansourabadi, A.H., L. Mohamed Khosroshahi, F. Noorbakhsh, and A. Amirzargar, *Cell therapy in transplantation: A comprehensive review of the current applications of cell therapy in transplant patients with the focus on Tregs, CAR Tregs, and Mesenchymal stem cells*. International immunopharmacology, 2021. **97**: p. 107669-107669.
108. Pan, G.H., Z. Chen, L. Xu, J.H. Zhu, P. Xiang, J.J. Ma, Y.W. Peng, G.H. Li, X.Y. Chen, J.L. Fang, Y.H. Guo, L. Zhang, and L.S. Liu, *Low-dose tacrolimus combined with donor-derived mesenchymal stem cells after renal transplantation: a prospective, non-randomized study*. Oncotarget, 2016. **7**(11): p. 12089-101.
109. Yang, Y.K., C.R. Ogando, C. Wang See, T.Y. Chang, and G.A. Barabino, *Changes in phenotype and differentiation potential of human mesenchymal stem cells aging in vitro*. Stem Cell Res Ther, 2018. **9**(1): p. 131.
110. Bonab, M.M., M.A. Sahraian, A. Aghsaie, S.A. Karvigh, S.M. Hosseinian, B. Nikbin, J. Lotfi, S. Khorramnia, M.R. Motamed, M. Togha, M.H. Harirchian, N.B. Moghadam, K. Alikhani, S. Yadegari, S. Jafarian, and M.R. Gheini, *Autologous mesenchymal stem cell therapy in progressive multiple sclerosis: an open label study*. Curr Stem Cell Res Ther, 2012. **7**(6): p. 407-14.
111. Jarnuczak, A.F., M.G. Albornoz, C.E. Eyers, C.M. Grant, and S.J. Hubbard, *A quantitative and temporal map of proteostasis during heat shock in Saccharomyces cerevisiae*. Mol Omics, 2018. **14**(1): p. 37-52.
112. Brunner, A.D., M. Thielert, C. Vasilopoulou, C. Ammar, F. Coscia, A. Mund, O.B. Hoerning, N. Bache, A. Apalategui, M. Lubeck, S. Richter, D.S. Fischer, O. Raether, M.A. Park, F. Meier, F.J. Theis, and M. Mann, *Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation*. Mol Syst Biol, 2022. **18**(3): p. e10798.
113. Kelly, R.T., *Single-cell Proteomics: Progress and Prospects*. Mol Cell Proteomics, 2020. **19**(11): p. 1739-1748.
114. Thul, P.J. and C. Lindskog, *The human protein atlas: A spatial map of the human proteome*. Protein Sci, 2018. **27**(1): p. 233-244.

115. Kendrew, J.C., G. Bodo, H.M. Dintzis, R.G. Parrish, H. Wyckoff, and D.C. Phillips, *A three-dimensional model of the myoglobin molecule obtained by x-ray analysis*. Nature, 1958. **181**(4610): p. 662-6.
116. Williamson, M.P., T.F. Havel, and K. Wuthrich, *Solution conformation of proteinase inhibitor IIA from bull seminal plasma by ¹H nuclear magnetic resonance and distance geometry*. J Mol Biol, 1985. **182**(2): p. 295-315.
117. Callaway, E., *Revolutionary cryo-EM is taking over structural biology*. Nature (London), 2020. **578**(7794): p. 201-201.
118. Dubrovsky, A., S. Sorrentino, J. Harapin, K.T. Sapra, and O. Medalia, *Developments in cryo-electron tomography for in situ structural analysis*. Arch Biochem Biophys, 2015. **581**: p. 78-85.
119. Artigues, A., O.W. Nadeau, M.A. Rimmer, M.T. Villar, X. Du, A.W. Fenton, and G.M. Carlson, *Protein Structural Analysis via Mass Spectrometry-Based Proteomics*. Adv Exp Med Biol, 2016. **919**: p. 397-431.
120. Vandermarliere, E., E. Stes, K. Gevaert, and L. Martens, *Resolution of protein structure by mass spectrometry*. Mass Spectrom Rev, 2016. **35**(6): p. 653-665.
121. Britt, H.M., T. Cragnolini, and K. Thalassinou, *Integration of Mass Spectrometry Data for Structural Biology*. Chemical reviews, 2022. **122**(8): p. 7952-7986.
122. Katta, V. and B.T. Chait, *Conformational changes in proteins probed by hydrogen-exchange electrospray-ionization mass spectrometry*. Rapid Commun Mass Spectrom, 1991. **5**(4): p. 214-7.
123. Marcsisin, S.R. and J.R. Engen, *Hydrogen exchange mass spectrometry: what is it and what can it tell us?* Anal Bioanal Chem, 2010. **397**(3): p. 967-72.
124. Lanman, J., T.T. Lam, M.R. Emmett, A.G. Marshall, M. Sakalian, and P.E. Prevelige, Jr., *Key interactions in HIV-1 maturation identified by hydrogen-deuterium exchange*. Nat Struct Mol Biol, 2004. **11**(7): p. 676-7.
125. Weis, D.D., T.E. Wales, J.R. Engen, M. Hotchkiss, and L.F. Ten Eyck, *Identification and characterization of EX1 kinetics in H/D exchange mass spectrometry by peak width analysis*. J Am Soc Mass Spectrom, 2006. **17**(11): p. 1498-1509.
126. Mehmood, S., C. Domene, E. Forest, and J.M. Jault, *Dynamics of a bacterial multidrug ABC transporter in the inward- and outward-facing conformations*. Proc Natl Acad Sci U S A, 2012. **109**(27): p. 10832-6.
127. West, G.M., B.D. Pascal, L.M. Ng, F.F. Soon, K. Melcher, H.E. Xu, M.J. Chalmers, and P.R. Griffin, *Protein conformation ensembles monitored by HDX reveal a structural rationale for abscisic acid signaling protein affinities and activities*. Structure, 2013. **21**(2): p. 229-35.
128. Zhang, Q., J. Chen, K. Kuwajima, H.M. Zhang, F. Xian, N.L. Young, and A.G. Marshall, *Nucleotide-induced conformational changes of tetradecameric GroEL mapped by H/D exchange monitored by FT-ICR mass spectrometry*. Sci Rep, 2013. **3**: p. 1247.
129. Kielkopf, C.S., M. Ghosh, G.S. Anand, and S.H.J. Brown, *HDX-MS reveals orthosteric and allosteric changes in apolipoprotein-D structural dynamics upon binding of progesterone*. Protein Sci, 2019. **28**(2): p. 365-374.
130. Khanal, A., Y. Pan, L.S. Brown, and L. Konermann, *Pulsed hydrogen/deuterium exchange mass spectrometry for time-resolved membrane protein folding studies*. J Mass Spectrom, 2012. **47**(12): p. 1620-6.

131. Engen, J.R., T.E. Wales, S. Chen, E.M. Marzluff, K.M. Hassell, D.D. Weis, and T.E. Smithgall, *Partial cooperative unfolding in proteins as observed by hydrogen exchange mass spectrometry*. Int Rev Phys Chem, 2013. **32**(1): p. 96-127.
132. Wang, L., L.C. Lane, and D.L. Smith, *Detecting structural changes in viral capsids by hydrogen exchange and mass spectrometry*. Protein Sci, 2001. **10**(6): p. 1234-43.
133. Goswami, D., S. Devarakonda, M.J. Chalmers, B.D. Pascal, B.M. Spiegelman, and P.R. Griffin, *Time window expansion for HDX analysis of an intrinsically disordered protein*. J Am Soc Mass Spectrom, 2013. **24**(10): p. 1584-92.
134. Ben-Nissan, G. and M. Sharon, *The application of ion-mobility mass spectrometry for structure/function investigation of protein complexes*. Curr Opin Chem Biol, 2018. **42**: p. 25-33.
135. Lanucara, F., S.W. Holman, C.J. Gray, and C.E. Eyers, *The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics*. Nat Chem, 2014. **6**(4): p. 281-94.
136. Fye, J.L., J. Woenckhaus, and M.F. Jarrold, *Hydration of Folded and Unfolded Gas-Phase Proteins: Saturation of Cytochrome c and Apomyoglobin*. Journal of the American Chemical Society, 1998. **120**(6): p. 1327-1328.
137. Ruotolo, B.T. and C.V. Robinson, *Aspects of native proteins are retained in vacuum*. Curr Opin Chem Biol, 2006. **10**(5): p. 402-8.
138. Jenner, M., J. Ellis, W.C. Huang, E. Lloyd Raven, G.C. Roberts, and N.J. Oldham, *Detection of a protein conformational equilibrium by electrospray ionisation-ion mobility-mass spectrometry*. Angew Chem Int Ed Engl, 2011. **50**(36): p. 8291-4.
139. McCabe, J.W., M.J. Hebert, M. Shirzadeh, C.S. Mallis, J.K. Denton, T.E. Walker, and D.H. Russell, *THE IMS PARADOX: A PERSPECTIVE ON STRUCTURAL ION MOBILITY-MASS SPECTROMETRY*. Mass Spectrom Rev, 2021. **40**(3): p. 280-305.
140. Grabenauer, M., S.L. Bernstein, J.C. Lee, T. Wyttenbach, N.F. Dupuis, H.B. Gray, J.R. Winkler, and M.T. Bowers, *Spermine binding to Parkinson's protein alpha-synuclein and its disease-related A30P and A53T mutants*. J Phys Chem B, 2008. **112**(35): p. 11147-54.
141. Smith, D.P., S.E. Radford, and A.E. Ashcroft, *Elongated oligomers in beta2-microglobulin amyloid assembly revealed by ion mobility spectrometry-mass spectrometry*. Proc Natl Acad Sci U S A, 2010. **107**(15): p. 6794-8.
142. Bernstein, S.L., N.F. Dupuis, N.D. Lazo, T. Wyttenbach, M.M. Condron, G. Bitan, D.B. Teplow, J.E. Shea, B.T. Ruotolo, C.V. Robinson, and M.T. Bowers, *Amyloid-beta protein oligomerization and the importance of tetramers and dodecamers in the aetiology of Alzheimer's disease*. Nat Chem, 2009. **1**(4): p. 326-31.
143. Gan, J., G. Ben-Nissan, G. Arkind, M. Tarnavsky, D. Trudeau, L. Noda Garcia, D.S. Tawfik, and M. Sharon, *Native Mass Spectrometry of Recombinant Proteins from Crude Cell Lysates*. Anal Chem, 2017. **89**(8): p. 4398-4404.
144. Jurneczko, E. and P.E. Barran, *How useful is ion mobility mass spectrometry for structural biology? The relationship between protein crystal structures and their collision cross sections in the gas phase*. Analyst, 2011. **136**(1): p. 20-8.
145. Chavez, J.D. and J.E. Bruce, *Chemical cross-linking with mass spectrometry: a tool for systems structural biology*. Curr Opin Chem Biol, 2019. **48**: p. 8-18.
146. Hauri, S., H. Khakzad, L. Happonen, J. Teلمان, J. Malmstrom, and L. Malmstrom, *Rapid determination of quaternary protein structures in complex biological samples*. Nat Commun, 2019. **10**(1): p. 192.

147. Schmidt, C., 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*. Nat Commun, 2013. **4**: p. 1985.
148. Belsom, A., M. Schneider, L. Fischer, M. Mabrouk, K. Stahl, O. Brock, and J. Rappsilber, *Blind testing cross-linking/mass spectrometry under the auspices of the 11(th) critical assessment of methods of protein structure prediction (CASP11)*. Wellcome Open Res, 2016. **1**: p. 24.
149. Rinner, O., J. Seebacher, T. Walzthoeni, L.N. Mueller, M. Beck, A. Schmidt, M. Mueller, and R. Aebersold, *Identification of cross-linked peptides from large sequence databases*. Nat Methods, 2008. **5**(4): p. 315-8.
150. Yang, B., 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*. Nat Methods, 2012. **9**(9): p. 904-6.
151. Liu, F., P. Lossel, R. Scheltema, R. Viner, and A.J.R. Heck, *Optimized fragmentation schemes and data analysis strategies for proteome-wide cross-link identification*. Nat Commun, 2017. **8**: p. 15473.
152. Kaake, R.M., X. Wang, A. Burke, C. Yu, W. Kandur, Y. Yang, E.J. Novtisky, T. Second, J. Duan, A. Kao, S. Guan, D. Vellucci, S.D. Rychnovsky, and L. Huang, *A new in vivo cross-linking mass spectrometry platform to define protein-protein interactions in living cells*. Mol Cell Proteomics, 2014. **13**(12): p. 3533-43.
153. Jayaraj, G.G., M.S. Hipp, and F.U. Hartl, *Functional Modules of the Proteostasis Network*. Cold Spring Harb Perspect Biol, 2020. **12**(1).
154. Fontana, A., G. Fassina, C. Vita, D. Dalzoppo, M. Zamai, and M. Zambonin, *Correlation between sites of limited proteolysis and segmental mobility in thermolysin*. Biochemistry, 1986. **25**(8): p. 1847-51.
155. Fontana, A., P.P. Delaureto, and V. Defilippis, *Molecular Aspects of Proteolysis of Globular-Proteins. Stability and Stabilization of Enzymes*, 1993. **47**: p. 101-110.
156. Hubbard, S.J., *The structural aspects of limited proteolysis of native proteins*. Biochim Biophys Acta, 1998. **1382**(2): p. 191-206.
157. Hubbard, S.J., S.F. Campbell, and J.M. Thornton, *Molecular recognition. Conformational analysis of limited proteolytic sites and serine proteinase protein inhibitors*. J Mol Biol, 1991. **220**(2): p. 507-30.
158. Hubbard, S.J., F. Eisenmenger, and J.M. Thornton, *Modeling studies of the change in conformation required for cleavage of limited proteolytic sites*. Protein Sci, 1994. **3**(5): p. 757-68.
159. Novotny, J. and R.E. Brucoleri, *Correlation among sites of limited proteolysis, enzyme accessibility and segmental mobility*. FEBS Lett, 1987. **211**(2): p. 185-9.
160. Kazanov, M.D., Y. Igarashi, A.M. Eroshkin, P. Cieplak, B. Ratnikov, Y. Zhang, Z. Li, A. Godzik, A.L. Osterman, and J.W. Smith, *Structural determinants of limited proteolysis*. J Proteome Res, 2011. **10**(8): p. 3642-51.
161. Fontana, A., M. Zambonin, P. Polverino de Laureto, V. De Filippis, A. Clementi, and E. Scaramella, *Probing the conformational state of apomyoglobin by limited proteolysis*. J Mol Biol, 1997. **266**(2): p. 223-30.
162. Picotti, P., A. Marabotti, A. Negro, V. Musi, B. Spolaore, M. Zambonin, and A. Fontana, *Modulation of the structural integrity of helix F in apomyoglobin by single amino acid replacements*. Protein Sci, 2004. **13**(6): p. 1572-85.

163. Feng, Y., G. De Franceschi, A. Kahraman, M. Soste, A. Melnik, P.J. Boersema, P.P. de Laureto, Y. Nikolaev, A.P. Oliveira, and P. Picotti, *Global analysis of protein structural changes in complex proteomes*. Nat Biotechnol, 2014. **32**(10): p. 1036-44.
164. Bantscheff, M., V. Weiss, and M.O. Glocker, *Identification of linker regions and domain borders of the transcription activator protein NtrC from Escherichia coli by limited proteolysis, in-gel digestion, and mass spectrometry*. Biochemistry, 1999. **38**(34): p. 11012-20.
165. Acquasaliente, L., L.A. Pelc, and E. Di Cera, *Probing prothrombin structure by limited proteolysis*. Sci Rep, 2019. **9**(1): p. 6125.
166. Leuenberger, P., S. Gansch, A. Kahraman, V. Cappelletti, P.J. Boersema, C. von Mering, M. Claassen, and P. Picotti, *Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability*. Science, 2017. **355**(6327).
167. Cappelletti, V., T. Hauser, I. Piazza, M. Pepelnjak, L. Malinowska, T. Fuhrer, Y. Li, C. Dörig, P. Boersema, L. Gillet, J. Grossbach, A. Dugourd, J. Saez-Rodriguez, A. Beyer, N. Zamboni, A. Caflisch, N. de Souza, and P. Picotti, *Dynamic 3D proteomes reveal protein functional alterations at high resolution in situ*. Cell, 2021. **184**(2): p. 545-559.e22.
168. Gates, S.N., A.L. Yokom, J. Lin, M.E. Jackrel, A.N. Rizo, N.M. Kendersky, C.E. Buell, E.A. Sweeny, K.L. Mack, E. Chuang, M.P. Torrente, M. Su, J. Shorter, and D.R. Southworth, *Ratchet-like polypeptide translocation mechanism of the AAA+ disaggregase Hsp104*. Science, 2017. **357**(6348): p. 273-279.
169. Piazza, I., K. Kochanowski, V. Cappelletti, T. Fuhrer, E. Noor, U. Sauer, and P. Picotti, *A Map of Protein-Metabolite Interactions Reveals Principles of Chemical Communication*. Cell, 2018. **172**(1-2): p. 358-372.e23.
170. Hubbard, S.J., R.J. Beynon, and J.M. Thornton, *Assessment of conformational parameters as predictors of limited proteolytic sites in native protein structures*. Protein Eng, 1998. **11**(5): p. 349-59.
171. Quast, J.P., Schuster, D., Picotti, P., *protti: an R package for comprehensive data analysis of peptide- and protein-centric bottom-up proteomics data*. Bioinformatics Advances,, 2022. **2**.
172. Sormanni, P., F.A. Aprile, and M. Vendruscolo, *The CamSol method of rational design of protein mutants with enhanced solubility*. J Mol Biol, 2015. **427**(2): p. 478-90.
173. Alexander Miguel, M., B. Paolo, N. Marco, C.E.T. Silvio, and P. Damiano, *FLIPPER: Predicting and Characterizing Linear Interacting Peptides in the Protein Data Bank*. Journal of Molecular Biology, 2021. **433**(9): p. 166900.
174. Piovesan, D., A. Del Conte, D. Clementel, A.M. Monzon, M. Bevilacqua, M.C. Aspromonte, J.A. Iserle, F.E. Orti, C. Marino-Buslje, and S.C.E. Tosatto, *MobiDB: 10 years of intrinsically disordered proteins*. Nucleic Acids Res, 2023. **51**(D1): p. D438-d444.
175. Piovesan, D., M. Necci, N. Escobedo, A.M. Monzon, A. Hatos, I. Mičetić, F. Quaglia, L. Paladin, P. Ramasamy, Z. Dosztányi, W.F. Vranken, N.E. Davey, G. Parisi, M. Fuxreiter, and S.C.E. Tosatto, *MobiDB: intrinsically disordered proteins in 2021*. Nucleic Acids Res, 2021. **49**(D1): p. D361-d367.
176. Meldal, B.H.M., A.J.H. Bye, L. Gajdoš, Z. Hammerová, A. Horácková, F. Melicher, L. Perfetto, D. Pokorný, M.R. Lopez, A. Türková, E.D. Wong, Z. Xie, E.B. Casanova, N. Del-Toro, M. Koch, P. Porras, H. Hermjakob, and S. Orchard, *Complex Portal 2018*:

- extended content and enhanced visualization tools for macromolecular complexes.* Nucleic Acids Res, 2019. **47**(D1): p. D550-d558.
177. Hornbeck, P.V., B. Zhang, B. Murray, J.M. Kornhauser, V. Latham, and E. Skrzypek, *PhosphoSitePlus, 2014: mutations, PTMs and recalibrations.* Nucleic Acids Res, 2015. **43**(Database issue): p. D512-20.
 178. Strassburg, S., S.M. Richardson, A.J. Freemont, and J.A. Hoyland, *Co-culture induces mesenchymal stem cell differentiation and modulation of the degenerate human nucleus pulposus cell phenotype.* Regen Med, 2010. **5**(5): p. 701-11.
 179. James, S., J. Fox, F. Afsari, J. Lee, S. Clough, C. Knight, J. Ashmore, P. Ashton, O. Preham, M. Hoogduijn, A. Ponzoni Rde, Y. Hancock, M. Coles, and P. Genever, *Multiparameter Analysis of Human Bone Marrow Stromal Cells Identifies Distinct Immunomodulatory and Differentiation-Competent Subtypes.* Stem Cell Reports, 2015. **4**(6): p. 1004-15.
 180. Galarza Torre, A., J.E. Shaw, A. Wood, H.T.J. Gilbert, O. Dobre, P. Genever, K. Brennan, S.M. Richardson, and J. Swift, *An immortalised mesenchymal stem cell line maintains mechano-responsive behaviour and can be used as a reporter of substrate stiffness.* Sci Rep, 2018. **8**(1): p. 8981.
 181. Leu, J.I., J. Pimkina, A. Frank, M.E. Murphy, and D.L. George, *A small molecule inhibitor of inducible heat shock protein 70.* Mol Cell, 2009. **36**(1): p. 15-27.
 182. Borner, G.H.H., *Organellar Maps Through Proteomic Profiling - A Conceptual Guide.* Mol Cell Proteomics, 2020. **19**(7): p. 1076-1087.
 183. Chang-Hui, S., *Chapter 4 - Gene Expression: Translation of the Genetic Code.* 2019: p. 87-116.
 184. Lee, B. and F.M. Richards, *The interpretation of protein structures: estimation of static accessibility.* J Mol Biol, 1971. **55**(3): p. 379-400.
 185. Taylor WR, T.J., Turnell WG, *An ellipsoidal approximation of protein shape.* Journal of Molecular Graphics, 1983. **1**(2): p. 30-38.
 186. Kabsch, W. and C. Sander, *Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features.* Biopolymers, 1983. **22**(12): p. 2577-637.
 187. NISHIKAWA, K. and T. OOI, *Radial Locations of Amino Acid Residues in a Globular Protein: Correlation with the Sequence1.* The Journal of Biochemistry, 1986. **100**(4): p. 1043-1047.
 188. Jumper, J., R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S.A.A. Kohl, A.J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A.W. Senior, K. Kavukcuoglu, P. Kohli, and D. Hassabis, *Highly accurate protein structure prediction with AlphaFold.* Nature, 2021. **596**(7873): p. 583-589.
 189. Varadi, M., S. Anyango, M. Deshpande, S. Nair, C. Natassia, G. Yordanova, D. Yuan, O. Stroe, G. Wood, A. Laydon, A. Žídek, T. Green, K. Tunyasuvunakool, S. Petersen, J. Jumper, E. Clancy, R. Green, A. Vora, M. Lutfi, M. Figurnov, A. Cowie, N. Hobbs, P. Kohli, G. Kleywegt, E. Birney, D. Hassabis, and S. Velankar, *AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models.* Nucleic Acids Research, 2021. **50**(D1): p. D439-D444.

190. Perrakis, A. and T.K. Sixma, *AI revolutions in biology: The joys and perils of AlphaFold*. EMBO Rep, 2021. **22**(11): p. e54046.
191. Daisuke, Y., N. Tomoyuki, K. Nobuo, T. Kei, and H. Tasuku, *Calumenin, a Ca²⁺-binding Protein Retained in the Endoplasmic Reticulum with a Novel Carboxyl-terminal Sequence, HDEF**. Journal of Biological Chemistry, 1997. **272**(29): p. 18232-18239.
192. Duchesne, L., S. Deschamps, I. Pellerin, V. Lagree, A. Froger, D. Thomas, P. Bron, C. Delamarche, and J.F. Hubert, *Oligomerization of water and solute channels of the major intrinsic protein (MIP) family*. Kidney Int, 2001. **60**(2): p. 422-6.
193. Fontana, A., P.P. de Laureto, B. Spolaore, and E. Frare, *Identifying disordered regions in proteins by limited proteolysis*. Methods Mol Biol, 2012. **896**: p. 297-318.
194. Fontana, A., P.P. de Laureto, B. Spolaore, E. Frare, P. Picotti, and M. Zambonin, *Probing protein structure by limited proteolysis*. Acta Biochim Pol, 2004. **51**(2): p. 299-321.
195. Ghosh, K. and K. Dill, *Cellular proteomes have broad distributions of protein stability*. Biophys J, 2010. **99**(12): p. 3996-4002.
196. Mahat, D.B., H.H. Salamanca, F.M. Duarte, C.G. Danko, and J.T. Lis, *Mammalian Heat Shock Response and Mechanisms Underlying Its Genome-wide Transcriptional Regulation*. Mol Cell, 2016. **62**(1): p. 63-78.
197. Alves, P., R.J. Arnold, D.E. Clemmer, Y. Li, J.P. Reilly, Q. Sheng, H. Tang, Z. Xun, R. Zeng, and P. Radivojac, *Fast and accurate identification of semi-tryptic peptides in shotgun proteomics*. Bioinformatics, 2008. **24**(1): p. 102-9.
198. Strader, M.B., D.L. Tabb, W.J. Hervey, C. Pan, and G.B. Hurst, *Efficient and specific trypsin digestion of microgram to nanogram quantities of proteins in organic-aqueous solvent systems*. Anal Chem, 2006. **78**(1): p. 125-34.
199. Olsen, J.V., S.E. Ong, and M. Mann, *Trypsin cleaves exclusively C-terminal to arginine and lysine residues*. Mol Cell Proteomics, 2004. **3**(6): p. 608-14.
200. Pirovich, D.B., A.A. Da'dara, and P.J. Skelly, *Multifunctional Fructose 1,6-Bisphosphate Aldolase as a Therapeutic Target*. Frontiers in Molecular Biosciences, 2021. **8**.
201. Gamblin, S.J., G.J. Davies, J.M. Grimes, R.M. Jackson, J.A. Littlechild, and H.C. Watson, *Activity and specificity of human aldolases*. J Mol Biol, 1991. **219**(4): p. 573-6.
202. Adams, D.R., D. Ron, and P.A. Kiely, *RACK1, A multifaceted scaffolding protein: Structure and function*. Cell Commun Signal, 2011. **9**: p. 22.
203. Ruiz Carrillo, D., R. Chandrasekaran, M. Nilsson, T. Cornvik, C.W. Liew, S.M. Tan, and J. Lescar, *Structure of human Rack1 protein at a resolution of 2.45 Å*. Acta Crystallogr Sect F Struct Biol Cryst Commun, 2012. **68**(Pt 8): p. 867-72.
204. Cox, J., N. Neuhauser, A. Michalski, R.A. Scheltema, J.V. Olsen, and M. Mann, *Andromeda: a peptide search engine integrated into the MaxQuant environment*. J Proteome Res, 2011. **10**(4): p. 1794-805.
205. de Souza, N. and P. Picotti, *Mass spectrometry analysis of the structural proteome*. Curr Opin Struct Biol, 2020. **60**: p. 57-65.
206. Mwangangi, D.M., E. Manser, and R.C. Robinson, *The structure of the actin filament uncapping complex mediated by twinfilin*. Sci Adv, 2021. **7**(5).
207. Palmgren, S., M. Vartiainen, and P. Lappalainen, *Twinfilin, a molecular mailman for actin monomers*. Journal of Cell Science, 2002. **115**(5): p. 881-886.

208. Vukmirovic, M., Z. Manojlovic, and B. Stefanovic, *Serine-threonine kinase receptor-associated protein (STRAP) regulates translation of type I collagen mRNAs*. Mol Cell Biol, 2013. **33**(19): p. 3893-906.
209. Jung, H., H.A. Seong, R. Manoharan, and H. Ha, *Serine-threonine kinase receptor-associated protein inhibits apoptosis signal-regulating kinase 1 function through direct interaction*. J Biol Chem, 2010. **285**(1): p. 54-70.
210. Sivasubramanian, N., G. Adhikary, P.C. Sil, and S. Sen, *Cardiac myotrophin exhibits rel/NF-kappa B interacting activity in vitro*. J Biol Chem, 1996. **271**(5): p. 2812-6.
211. Takeda, S., R. Koike, T. Nagae, I. Fujiwara, A. Narita, Y. Maéda, and M. Ota, *Crystal structure of human V-1 in the apo form*. Acta Crystallogr F Struct Biol Commun, 2021. **77**(Pt 1): p. 13-21.
212. Saveliev, S., M. Bratz, R. Zubarev, M. Szapacs, H. Budamgunta, and M. Urh, *Trypsin/Lys-C protease mix for enhanced protein mass spectrometry analysis*. Nature methods, 2013. **10**(11): p. i-ii.
213. Caruso Bavisotto, C., G. Alberti, A.M. Vitale, L. Paladino, C. Campanella, F. Rappa, M. Gorska, E. Conway de Macario, F. Cappello, A.J.L. Macario, and A. Marino Gammazza, *Hsp60 Post-translational Modifications: Functional and Pathological Consequences*. Front Mol Biosci, 2020. **7**: p. 95.
214. Zang, Y., M. Jin, H. Wang, Z. Cui, L. Kong, C. Liu, and Y. Cong, *Staggered ATP binding mechanism of eukaryotic chaperonin TRiC (CCT) revealed through high-resolution cryo-EM*. Nat Struct Mol Biol, 2016. **23**(12): p. 1083-1091.
215. Quintana-Gallardo, L., J. Martín-Benito, M. Marcilla, G. Espadas, E. Sabidó, and J.M. Valpuesta, *The cochaperone CHIP marks Hsp70- and Hsp90-bound substrates for degradation through a very flexible mechanism*. Sci Rep, 2019. **9**(1): p. 5102.
216. Koga, H., S. Kaushik, and A.M. Cuervo, *Protein homeostasis and aging: The importance of exquisite quality control*. Ageing Res Rev, 2011. **10**(2): p. 205-15.
217. Kors, S., K. Geijtenbeek, E. Reits, and S. Schipper-Krom, *Regulation of Proteasome Activity by (Post-)transcriptional Mechanisms*. Frontiers in Molecular Biosciences, 2019. **6**.
218. Miyamoto, Y., T. Saiwaki, J. Yamashita, Y. Yasuda, I. Kotera, S. Shibata, M. Shigeta, Y. Hiraoka, T. Haraguchi, and Y. Yoneda, *Cellular stresses induce the nuclear accumulation of importin alpha and cause a conventional nuclear import block*. J Cell Biol, 2004. **165**(5): p. 617-23.
219. Masser, A.E., M. Ciccarelli, and C. Andréasson, *Hsf1 on a leash - controlling the heat shock response by chaperone titration*. Exp Cell Res, 2020. **396**(1): p. 112246.
220. Ochoa, D., A.F. Jarnuczak, C. Viéitez, M. Gehre, M. Soucheray, A. Mateus, A.A. Kleefeldt, A. Hill, L. Garcia-Alonso, F. Stein, N.J. Krogan, M.M. Savitski, D.L. Swaney, J.A. Vizcaíno, K.M. Noh, and P. Beltrao, *The functional landscape of the human phosphoproteome*. Nat Biotechnol, 2020. **38**(3): p. 365-373.
221. Lokireddy, S., N.V. Kukushkin, and A.L. Goldberg, *cAMP-induced phosphorylation of 26S proteasomes on Rpn6/PSMD11 enhances their activity and the degradation of misfolded proteins*. Proc Natl Acad Sci U S A, 2015. **112**(52): p. E7176-85.
222. Tang, H., E. Tian, C. Liu, Q. Wang, and H. Deng, *Oxidative stress induces monocyte necrosis with enrichment of cell-bound albumin and overexpression of endoplasmic reticulum and mitochondrial chaperones*. PLoS One, 2013. **8**(3): p. e59610.
223. Marino Gammazza, A., C. Campanella, R. Barone, C. Caruso Bavisotto, M. Gorska, M. Wozniak, F. Carini, F. Cappello, A. D'Anneo, M. Lauricella, G. Zummo, E. Conway de Macario, A.J. Macario, and V. Di Felice, *Doxorubicin anti-tumor mechanisms*

- include Hsp60 post-translational modifications leading to the Hsp60/p53 complex dissociation and instauration of replicative senescence.* Cancer Lett, 2017. **385**: p. 75-86.
224. Wang, R.Y., C.M. Noddings, E. Kirschke, A.G. Myasnikov, J.L. Johnson, and D.A. Agard, *Structure of Hsp90-Hsp70-Hop-GR reveals the Hsp90 client-loading mechanism.* Nature, 2022. **601**(7893): p. 460-464.
 225. Mackmull, M.T., L. Nagel, F. Sesterhenn, J. Muntel, J. Grossbach, P. Stalder, R. Bruderer, L. Reiter, W.D.J. van de Berg, N. de Souza, A. Beyer, and P. Picotti, *Global, in situ analysis of the structural proteome in individuals with Parkinson's disease to identify a new class of biomarker.* Nat Struct Mol Biol, 2022. **29**(10): p. 978-989.
 226. Hoxha, E., F. Tempia, P. Lippiello, and M.C. Miniaci, *Modulation, Plasticity and Pathophysiology of the Parallel Fiber-Purkinje Cell Synapse.* Front Synaptic Neurosci, 2016. **8**: p. 35.
 227. Wilhelmus, M.M., J.G. Bol, E.S. Van Haastert, A.J. Rozemuller, G. Bu, B. Drukarch, and J.J. Hoozemans, *Apolipoprotein E and LRP1 Increase Early in Parkinson's Disease Pathogenesis.* Am J Pathol, 2011. **179**(5): p. 2152-6.

Appendix

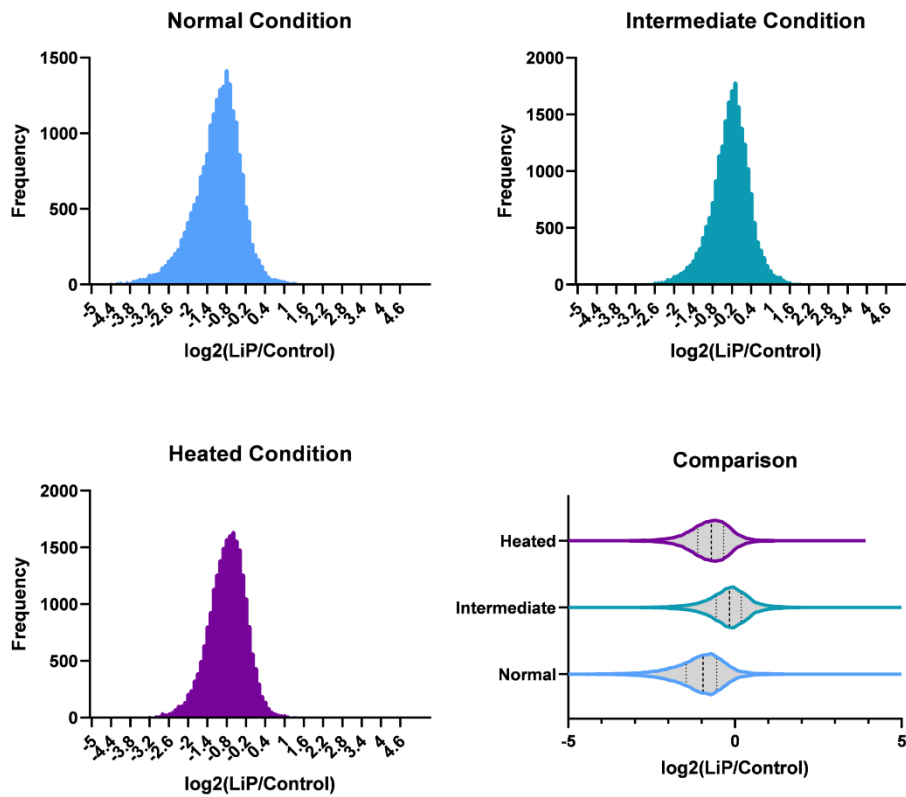


Figure A.1 : Tryptic peptide intensity \log_2 distribution, MaxQuant processed data for Chapter 4. For the normal condition, 53% of X_i values fell between -1 and 1. This number was 87% in the intermediate condition and 68% in the heated condition. 47% of peptides in the normal condition had $X_i < -1$, 10% in the intermediate condition and 31% in the heated condition. The graph on the bottom right shows the X_i values for all conditions plotted together, this time displayed as a violin plot.

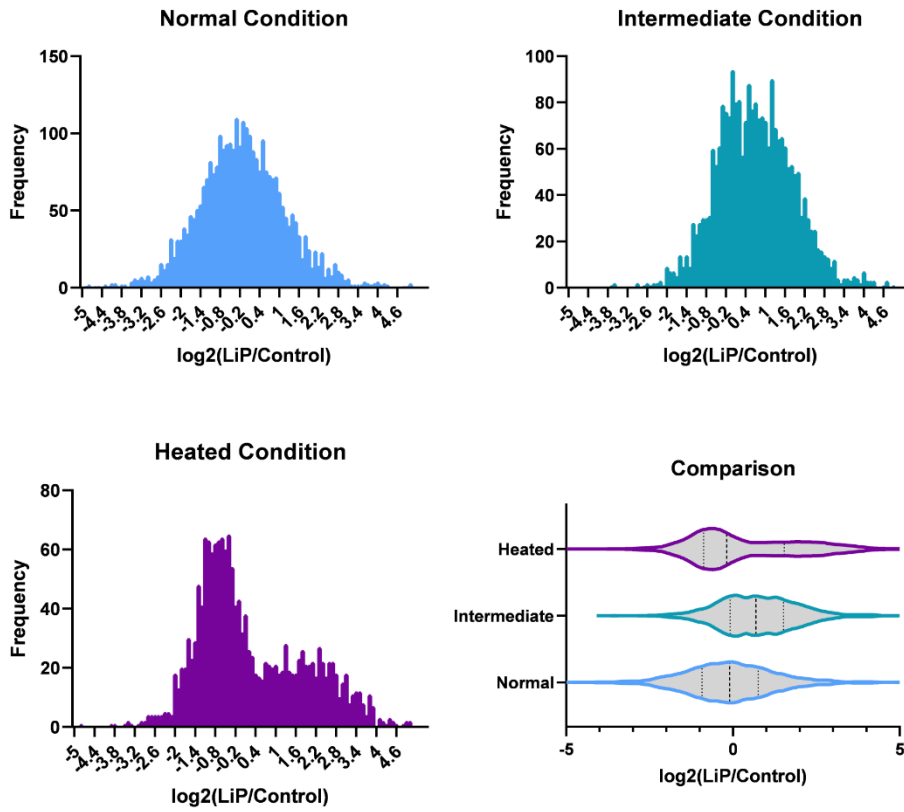


Figure A.2: Semi-tryptic peptide intensity \log_2 distribution, MaxQuant processed data for Chapter 4. For the normal condition, 57% of X_i values fell between -1 and 1. This number was 54% in the intermediate condition and 48% in the heated condition. 19% of peptides in the normal condition had $X_i > 1$, 40% in the intermediate condition and 32% in the heated condition. The graph on the bottom right shows the X_i values for all conditions plotted together, this time displayed as a violin plot.