MASS SPECTROMETRY APPROACHES TOWARDS THE CHARACTERISATION OF PROTEIN POST-TRANSLATIONAL MODIFICATION CROSSTALK

KISH R. ADONI

A thesis submitted to the University of
Birmingham for the degree of
DOCTOR OF PHILOSOPHY

School of Biosciences
College of Life and Environmental Sciences
University of Birmingham
December 2022

Abstract

Post-translational modifications (PTM) enable a given protein to access different proteoform variants in response to their environment. The modification of a protein with a PTM can facilitate the occurrence of other PTMs within the same protein, allowing it to access a specific proteoform via a mechanism known as positive PTM crosstalk. PTMs are dynamic, thus they provide the infrastructure for real-time protein regulation as the cell adapts to its surroundings. As such, to truly understand the behaviour of a given protein, all possible proteoforms that it can access must be characterised. To this end, mass spectrometry represents a powerful analytical technique for the site-localised identification of PTM decorated proteins. In this thesis, High Field Asymmetric Waveform Ion Mobility Spectrometry (FAIMS) was incorporated into a bottom-up proteomics workflow for the purpose of improving the identification of positive PTM crosstalk sites. A 6-fold improvement in multiple PTM containing peptides, as candidates for positive PTM crosstalk, was demonstrated in HeLa S3 cervical cancer cells, with 40% of these sites not previously reported. The technique was then leveraged against SUM52 triple negative breast carcinoma cancer cells for the investigation of positive PTM crosstalk that facilitates the cancer phenotype. 498 positive PTM crosstalk sites were suggested to play a role in generating the cancer phenotype. Further, novel positive PTM crosstalk mechanisms were proposed via the identification of specific PTM containing peptide isoforms, such as the N-terminal acetylation of cofilin, leading to its sequestration via serine-3 phosphorylation to facilitate actin polymerisation for cancer cell migration. Following this, a multimodal mass spectrometry investigation of a novel S-glutathionylation PTM site within the evolutionarily relevant light harvesting phycobilisome protein complex of cyanobacteria was performed.

Native MS, native top-down MS, LC-MS/MS and LC-FAIMS-MS/MS were applied to demonstrate the relevance of this novel S-glutathionylation PTM in protecting the phycobilisome from the reactive oxygen species that are generated as by-products of photosynthesis. Further, 84 % of the S-glutathionylated peptides that were identified with LC-FAIMS-MS/MS occurred with a concurrent phosphorylation PTM, suggesting positive PTM crosstalk between S-glutathionylation and phosphorylation within cyanobacteria. Together, these findings demonstrate the relevance of mass spectrometry approaches for the aim of characterising proteoform variability within proteins, to truly understand the role of proteins in biochemistry.

Acknowledgments

Firstly, I would like to thank my primary supervisor Dr Aneika Leney for her unwavering support throughout my PhD. When I started this journey, I was completely new (and innocent) to the world of mass spectrometry. Aneika trained me up and gave me the tools that I needed to immerse myself in the field, for which I am incredibly grateful. I would also like to thank my secondary supervisors: Dr Debbie Cunningham and Prof John Heath for their guidance on the biological aspects of my work. To Jas, Dannielle, Matt, Tom, Jedd, and the Cooper group; thank you for making our office such a friendly place to discuss and debate everything from the best mass analysers to the virtues of the fig roll (heavily overrated). Particularly to Jedd, your borderline encyclopaedic understanding of the nuances of biochemistry helped me think about my work in a more analytical way. To Dr Todd Mize and Dr Jinglei Yu, thank you for always being on hand to ensure the smooth running of the instruments.

To my flatmate Max, who's joy of listening to me try to wake up to my alarm every morning was only surpassed by his love of former Tottenham Hotspur attacking midfielder Rafael Van der Vaart ("he just got the club"). I couldn't have asked for a better comrade with which to do battle against the ups and downs of PhD life. And to my older brother Pavan, who has always been there to support me in and out of academia.

To my father, who is sadly no longer with us. His hard work and sacrifice throughout my childhood ensured that we had a roof over our head and food on the table, as well as giving us opportunities that he never had. Thank you for nurturing the scientific curiosity within me that I have had for as long as I can remember. To my mother, whose courage during those dark and uncertain days ensured that I stayed on the straight and narrow when things could have gone very differently. No amount of words could describe how grateful I am for everything you have done for me, and my achievements will always be more yours than mine. Don't be fooled though, like any Indian mother worth their salt, there was no carrot, it was just a series of large sticks and I now find myself in the awkward position of having to thank her for that.

And finally, to my partner Fran, who's kindness and humility was the perfect antidote to those long days in the lab. I am so grateful to have had you by my side during this journey. You have that happy knack of giving me perspective, making me forget about the polyethylene glycol that had found its way into my samples with those special words: "grow up".

Outputs

Peer Reviewed Published Articles

- Adoni, K. R; Cunningham, D. L; Heath, J. K; Leney, A. C. FAIMS Enhances the Detection of PTM Crosstalk Sites, Journal of Proteome Research, 2022, 21, 4, 930–939.

Oral Presentations

- Adoni, K. R; Cunningham, D. L; Heath, J. K; Leney, A. C. FAIMS Enhances the Detection of PTM Crosstalk Sites. British Mass Spectrometry Society, 2021.
- Adoni, K. R; Cunningham, D. L; Heath, J. K; Leney, A. C. FAIMS Enhances the Detection of PTM Crosstalk Sites. Birmingham University Biosciences Post Graduate Research Symposium, 2022 (1st Prize).
- Adoni, K. R; Leney, A. C. A Multi-Modal Mass Spectrometry Investigation of S-Glutathionylation PTM in Cyanobacteria. British Mass Spectrometry Society, 2022 (1st Prize).

Poster Presentations

- Adoni, K. R; Cunningham, D. L; Heath, J. K; Leney, A. C. FAIMS Enhances the Detection of PTM Crosstalk Sites. Isolated Biomolecules and Biomolecular Interactions, 2022 (Poster Prize).
- Adoni, K. R; Cunningham, D. L; Heath, J. K; Leney, A. C. FAIMS Enhances the Detection of PTM Crosstalk Sites. International Mass Spectrometry Conference, 2022.

Contents

List of Figures

List of Tables

Acronyms and Abbreviations

ΔM	Mass difference
AACT	Amino Acid-Coded Mass Tag
AC	Alternating Current
ac-Nt	N-terminal acetylation
ATP	Adenosine Triphosphate
CAD	Collision Activated Dissociation
c-FAIMS	Cylindrical FAIMS
CID	Collision Induced Dissociation
C-phycocyanin	Cyanobacterial phycocyanin
CV	Compensation Voltage
DC	Direct Current
DDA	Data Dependent Acquisition
DIA	Data Independent Acquisition
diGly	Di-glycine
DNA	Deoxyribonucleic Acid
DTIMS	Drift Tube Ion Mobility Spectrometry
DV	Dispersion Voltage
ECD	Electron Capture Dissociation
ERK	Extracellular Signal Regulated Kinase
ESI	Electrospray Ionisation
ETcaD	ETD with supplemental CAD
ETD	Electron Transfer Dissociation
EThcD	ETD with supplemental HCD
F-actin	Filamentous actin
FAIMS	High Field Asymmetric Waveform Ion Mobility Spectrometry
FDR	False Discovery Rate
FGF	Fibroblast Growth Factor
FGFR	Fibroblast Growth Factor Receptor
FLR	False Localisation Rate
FRS2	Fibroblast Growth Factor Receptor Substrate 2

FT-ICR	Fourier Transform-Ion Cyclotron Resonance
G1	Growth phase 1 of cell cycle
G-actin	Globular actin
GAP	GTPase Activating Protein
GEF	Guanine Nucleotide Exchange Factor
GI_{50}	Half maximal growth inhibitory concentration
GO	Gene Ontology
GS	Oxidised glutathione
GSH	Reduced glutathione
GTP	Guanosine Triphosphate
HCD	Higher Energy Collision Induced Dissociation
HILIC	Hydrophobic Interaction Liquid Chromatography
HiLoC	High Low Photodissociation
IC_{50}	Half maximal inhibitory concentration
Ig	Immunoglobulin
IMAC	Immobilised Metal Affinity Chromatography
IMS	Ion Mobility Spectrometry
IRMPD	Infrared Multiphoton Dissociation
LC	Liquid Chromatography
LC-MS/MS	Liquid Chromatography- Tandem Mass Spectrometry
LFQ	Label Free Quantitation
LUCA	Last Universal Common Ancestor
m/z	Mass to charge ratio
MALDI	Matrix-Assisted Laser Desorption/Ionisation
MD-LC	Multidimensional-Liquid Chromatography
MOAC	Metal Oxide Affinity Chromatography
MRM	Multiple Reaction Monitoring
mRNA	Messenger RNA
MS/MS	Tandem mass spectrometry
mTOR	mammalian target of rapamycin
MudPIT	Multidimensional Protein Identification
Multi-PTM peptide	Peptide that contains two or more PTMs

nESI	Nano ESI
NeuCode	Neutron-encoding amino acid labelling
PASEF	Parallel Accumulation Serial Fragmentation
PD	Proteome Discoverer
Peptidoform	Peptide isoform
p-FAIMS	Planar FAIMS
ppm	Parts per million
PRIDE	Proteomics Identification Database
PRM	Parallel Reaction Monitoring
PSM	Peptide Spectral Match
PTM	Post-Translational Modification
QqQ	triple quadrupole
QTOF	Quadrupole-Time of Flight
R	Resolution
RF	Radio Frequency
RNA	Ribonucleic Acid
ROS	Reactive Oxygen Species
RP	Reverse Phase
S	DNA synthesis phase of cell cycle
SAX	Strong Anion Exchange
SCX	Strong Cation Exchange
SEPTM	Serial Enrichment of Post Translational Modifications
SID	Surface Induced Dissociation
SILAC	Stable Isotope Labelling of Amino Acids in Cell Culture
TACC	Transforming Acidic Coiled-Coil Gene Family proteins
TM	Transmembrane
TMT	Tandem Mass Tag
ToF	Time of Flight
tRNA	Transfer RNA
TWIMS	Travelling Wave Ion Mobility Spectrometry
UV	Ultraviolet
UVPD	Ultraviolet Photodissociation

1. Introduction

1.1. Overview

Protein proteoform characterisation represents the next great challenge in biochemistry. The Human Genome Project spearheaded our desire to understand how proteins orchestrate the biochemistry of life, mapping and sequencing every gene within the human genome[1]. But to contextualise the role of each gene coded protein, its associated abundance, distribution, subcellular localisation and interaction with other biomolecules must be realised, in a collaborative effort termed The Human Proteome Project[2]. A major regulator of these protein features is Post-Translational Modification (PTM), whereby a protein is chemically modified after its ribosomal synthesis to modulate its spaciotemporal behaviour in response to environmental change. As such, fulfilling the Human Proteome Project will involve mapping every possible PTM on every protein at any given time, to realise a comprehensive library of each PTM permutation of each protein, termed a proteoform. As well, the mechanism by which a proteoform is generated, whereby one PTM impacts the protein's modification by another PTM, termed PTM-crosstalk, must be characterised. Accomplishing this feat will involve the development of novel biochemical and computational techniques, of which mass spectrometry has been described as one of the three working pillars, together with antibody capture and bioinformatics tools[2]. Progress will facilitate a better understanding of how dysfunctional protein regulation can prime disease states such as cancer, as well as informing the evolutionary development of complex protein systems for such processes as photosynthesis. The work presented in this thesis aims to develop a mass spectrometry technique for the identification of multiple co-existing PTMs on a protein, for better characterised proteoform variability. Following this, the methodology is applied across two distinct biologically relevant contexts. Firstly, for the characterisation

of novel candidates for protein PTM crosstalk in the context of cancer development. Secondly, in a multimodal investigation (with other mass spectrometry techniques) of the PTM status of the photosynthetic protein machinery within cyanobacteria, as the first oxygenic photosynthetic organisms from which eukaryotic photosynthesis evolved[3].

1.2. Post-Translational Modifications

1.2.1. Background

In the context of the central dogma of biology, RNA polymerases use the genetic code of DNA within the nucleus to transcribe a complementary mRNA sequence from a gene. After mRNA splicing removes intronic content, the mature mRNA strand is translated into a polypeptide chain within the ribosome, with chaperones and other accessory proteins aiding the process of protein tertiary conformational folding, co-factor recruitment and oligomerisation, to ultimately form a functional protein[4]. Thus, the protein's behaviour is dictated primarily by its structure in the context of a specific biological environment[4]. However, the final fate of the protein is not sealed. Within a dynamic environment, proteins must have the capacity to modulate their functionality in response to stimuli. As such, the post-translational modification of a protein can reversibly regulate its function post-synthesis. PTMs primarily occur on amino acid side chains, but can also be found at the N-terminal amine group or C-terminal carboxylic acid group of the protein[5]. Added to this, protein cleavage is also considered a PTM[6], and has been demonstrated to be regulated by other PTMs within the protein[7]. Each modification imparts a specific change in the biochemical structure of a protein. The covalent, electrostatic, and steric interactions of the modification finetune the behaviour of a protein, allowing for real-time regulation in response to environmental change. As such, an infrastructure for the dynamic regulation of proteins and their specialised response to environmental stimulation is laid[5].

1.2.2. PTM Regulation in Cell Biology

PTMs play a key role in the regulation of protein driven biological processes, changing the biochemistry of proteins to modulate their behaviour[8]. Further, these modifications can be reversible, allowing for dynamic responses in protein activity that are stimulated by changes in the external environment[9]. Multiple proteins, whose amino acid side chains are susceptible to PTM regulation, can together form a cell signal transduction pathway from external stimulation of a membrane bound receptor, through to activation of transcription factors that translocate to the nucleus and facilitate changes in global proteome abundances, leading to modified cell behaviour[4]. For example, animal cell membrane bound Fibroblast Growth Factor Receptors are stimulated by Fibroblast Growth Factor glycoproteins in the extracellular environment, leading to global changes in the proteome abundance profile of the cell (Chapter 4). Cell biology is strewn with such mechanisms that govern a vast range of processes including proliferation[10], differentiation[11], cytoskeletal reorganisation[12], immune response[13] and apoptosis[14]. Epigenetic modification of histone proteins[15], whereby PTM's lace histone-chromatin structures that DNA is packaged on, has been cited as a key regulator of gene expression. Dysregulation of PTMs that orchestrate the above processes can lead to many diseases including neurodegenerative diseases[16], cancer[17] and diabetes[18,19].

1.2.3. Types of PTM

More than 400 unique PTMs have been reported[20], across a plethora of biological processes. Of these PTMs, according to dbPTM[21] (a database for the curation of PTMs), 24 are considered as the major PTMs used by biology to regulate protein behaviour[5]. Based on experimental evidence within humans; phosphorylation, acetylation and

ubiquitination represent more than 90 % of all reported PTMs[21]. Further, the average amino acid can undergo three dynamic modifications, with lysine capable of 15 separate PTMs and serine/cysteine reported to undergo 10[21]. Serine-phosphorylation is the most reported PTM according to dbPTM[21]. Together, this equates to a cellular framework for the highly orchestrated regulation of proteins, with a vast array of PTMs finetuning the protein's behaviour as it adapts to environmental change (Figure 1.1). Importantly, it has been reported that many types of PTM have an ancient origin[22], with serine- and threonine-phosphorylation identified across three domains of life[23–25]. Further, phylostratigraphic analysis of canonical kinase sequences suggests a monophyletic origin of serine- and threonine-kinases in the last universal common ancestor (LUCA)[26]. This places protein PTM-regulation as an evolutionarily conserved mechanism in biology, with its roots being traced back to early lifeforms. Consequently, an understanding of ancient PTM regulation of proteins would inform an evolutionary understanding of protein biochemistry. Table 1.1 describes ten common protein PTMs, together with their corresponding canonical amino acid targets and associated enzymes.

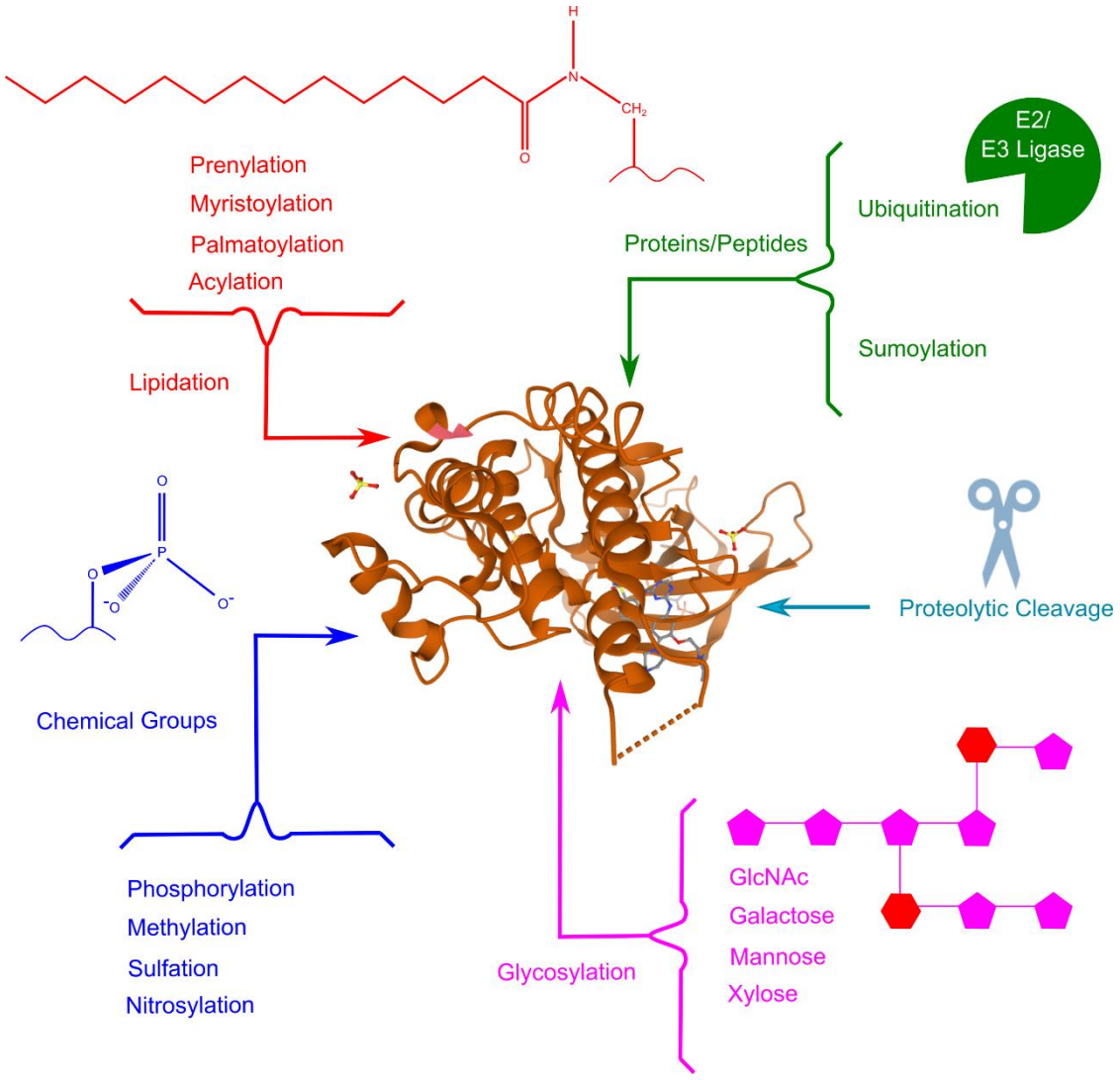


Figure 1.1 | Diagram representing some types of PTM that can occur on a protein. *Protein structure taken from The Protein Data Bank in Europe*[27]*, ID: 6lvk.*

Table 1.1 | Common PTMs and corresponding common amino acid targets and associated enzymes[21]. Nt = N-terminal modification site.

Modification	**Common Amino Acid Targets**	**Some Examples of Associated Enzymes**
Phosphorylation[28]	S, T, Y	Kinases, Phosphatases
Acetylation[29]	Nt, K, R,	Acetylases, Deacetylases
Ubiquitination[30]	K	E1 (ubiquitin activating), E2 (ubiquitin conjugating), E3 (ubiquitin ligase)

Methylation[31]	K, R	Methyltransferases
N-GlcNAc[32]	N	N-Glycosylation Enzyme C, Endo-β-N-acetylglucosaminidase
O-GlcNAc[7]	S, T	O-GlcNAcase, O-GlcNAc transferase
SUMOylation[33]	K	E1 (activating), E2 (conjugating), E3 (ligase)
S-Palmitoylation[34]	C	Palmitoyltransferases and Acyl Protein Thioesterases
Myristoylation[35]	G	N-myristoyltransferases
Prenylation[36]	C	Farnesyltransferases, Geranylgeranyltransferase
Sulfation[37]	Y	Tyrosylprotein sulfotransferases 1 & 2

1.2.4. PTM Crosstalk

As described above, the number and specific biochemical nature of PTMs that can modify a protein provides the infrastructure for dynamic, real-time regulation of proteins in response to their biological environment. Further, many proteins contain multiple PTMs that differentially contribute to protein function[38]. The concept of PTM crosstalk, whereby one PTM interacts with another PTM, was discovered in 2008 by Hart and co-workers[39]. Here, a reciprocal relationship between phosphorylation and O-GlcNAcylation was reported with O-GlcNAcylation being linked to the downregulation of 280 phosphosites and the upregulation of 148 phosphosites[39]. As such, the occurrence of a phosphosite was found to regulate the occurrence of an O-GlcNAc site and vice-versa. Hart and co-workers described the importance of steric competition for the modification of the same or proximal amino acid sites leading to reciprocal and negative crosstalk, respectively[39,40] . As such, the field of PTM crosstalk was seeded, with reciprocal crosstalk (whereby one modification competes with the other on the same

site), positive crosstalk (whereby one modification enhances the occurrence of another modification site), and negative crosstalk (whereby one modification inhibits the modification of another site) as the underlying mechanisms of action[39,41] (Figure 1.2). Each PTM within a protein would be considered as a contributor to the overall combinatorial impact of all the PTMs on the protein at a specific time, with the proteoform dictating the functionality of the protein. For example, glycogen synthase kinase-3β (GSK3β) preferentially phosphorylates substrates upon pre-phosphorylation, four amino acids upstream of its target phosphosite. Thus, the first phosphorylation is required to prime the second in a positive PTM crosstalking manner[41,42]. Since these discoveries, PTM crosstalk between acetylation and phosphorylation[43], ubiquitination and phosphorylation[44,45], acetylation and methylation[46] and methylation and phosphorylation[47] have all been observed, together with many other PTM-combinations. One of the more prominent examples of PTM crosstalk involves histones, whereby phosphorylation, acetylation, ubiquitination and methylation combinatorically modify the histone proteins, resulting in a complex histone-PTM language known as the 'histone code'[46,48]. The combinatorial impact of these PTMs can modify chromatin structure, enabling interaction of histone-bound genes to transcription factors with consequent mRNA synthesis. For example, CREB Binding Protein (CBP)/p300 acetylase and the Coactivated Associated Arginine Methyltransferase 1 (CARM1) methyltransferase have been reported to collaboratively-positively regulate the expression of oestrogen responsive genes[46]. Thus, to truly understand the nuanced behaviour of a protein at a specific time, its complete proteoform must be characterised.

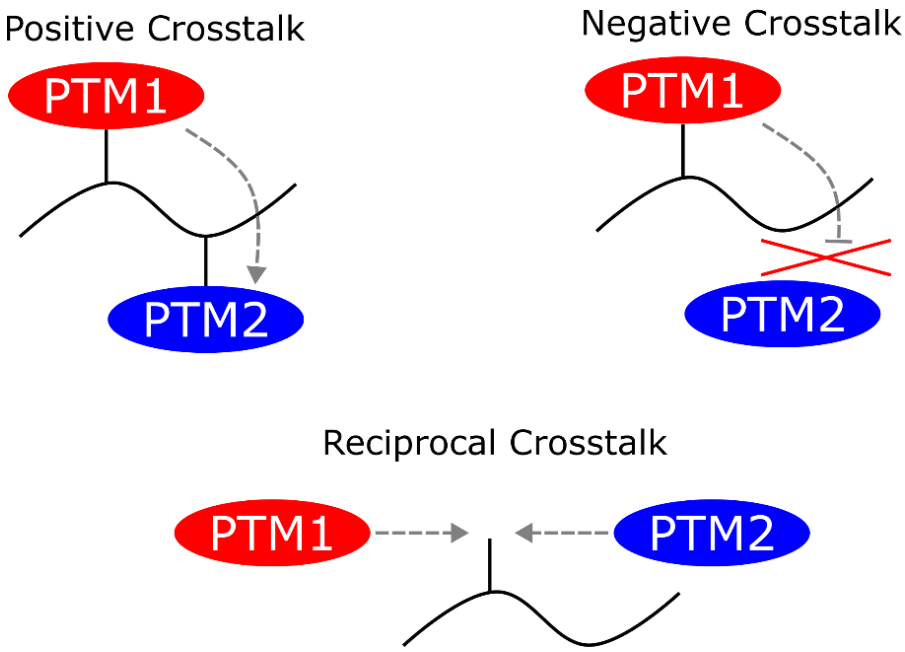


Figure 1.2 | Diagrammatic representation of PTM crosstalk including positive, negative, and reciprocal PTM crosstalk.

1.2.5. Protein Regulation via Proteoform Variability

Before the Human Genome Project came to existence, it was estimated that around 100,000 genes governed protein behaviour in the human cell[49,50]. The elucidation of just ~20,300 genes (from the Human Genome Project) translating to the vast array of protein complexity within the cell could only be reconciled by the consideration of protein variation[51]. Such variation of a protein across cell and tissue types and subcellular location was explainable by diversity at the protein level, rather than exclusively by the distinct number of genes within the human genome[49,51]. For example, allelic variation at the DNA level, mRNA alternative splicing at the RNA level, and post-translational modification at the protein level would be necessary to generate the complexity of protein behaviour across biological processes within the cell[49]. Thus, identifying these

spatiotemporally distinct proteoforms of a protein, governed by PTM-crosstalk to generate the final proteoform, has become a priority in understanding protein behaviour.

1.3. Techniques for the Identification of PTM Crosstalk

To identify the role of PTM crosstalk in biological mechanisms, each PTM must first be identified before the relationship between the PTMs can be investigated. To this end, upregulation of one PTM can be stimulated on a global scale, with the monitoring of another PTM using such methods as Western blotting[52] and quantitative proteomics[38]. There are drawbacks to these approaches, as the relationship between the PTMs is not clear and the site-specific localisation of each PTM cannot be retrieved from Western blotting against a non-specific antibody. Other work has involved the investigation of co-occurring PTMs on synthetic peptides as mimics of regions of a protein[53–55]. The systematic analysis of different synthetic peptides can build a library of consensus PTM crosstalk motifs, and those that inhibit the occurrence of PTM crosstalk[53–55]. However, synthetic peptides do not mimic the protein's natural environment well, so the physiological relevance of such findings is compromised. With PTM-crosstalk now appreciated as a vital regulatory mechanism, improved methods towards its identification and characterisation must be developed. Mass spectrometry, as a major tool for such purposes, has undergone rapid development over the last ~20 years, generating renewed confidence that this knowledge gap can soon be tackled.

1.4. Mass Spectrometry

1.4.1. Ionisation

Initially, the biological analyte of interest must be ionised before analysis can be performed. Numerous ionisation techniques exist within the field of mass spectrometry, with different techniques yielding different information. Hard ionisation techniques often lead to analyte fragmentation, due to the nature of the high residual energy absorbed by the species as it is ionised. Examples include chemical ionisation, electron ionisation and inductively coupled plasma ionisation[56]. Soft ionisation techniques have become more popular in recent years due to the preservation of intact biomolecular structures during the ionisation process, as the low residual energy that is transferred tends to maintain covalent and non-covalent interactions. Examples of soft ionisation techniques include Electrospray Ionisation (ESI) and Matrix-Assisted Laser Desorption/Ionisation (MALDI)[56].

For the mass spectrometry experiments carried out in this thesis, ESI was the ionisation technique used. For a given analyte within an electrically conductive solvent, flowing at atmospheric pressure within a capillary; an electric field is applied between the capillary and an inlet electrode. Thus, as charge accumulates at the liquid-surface interface and the magnitude of force exerted on the solvent droplet approaches that of the surface tension; a Taylor cone is formed (as first described by Taylor *et al.* in 1964)[57,58]. Subsequently, the force exerted by the increasing charge density of the droplet exceeds that of the surface tension, leading to fission of the droplet into smaller droplets. The process is repeated as evaporation again facilitates an increased droplet charge density, such that the droplets undergo repeated cycles of evaporation and fission until ions are eventually released from the charged solvent droplet. Three models describe this

process: the charge residue model (CRM), the chain ejection model (CEM) and the ion evaporation model (IEM)[59]. Globular protein ions preferentially undergo release from the droplet via CRM, whereby the protein ion is trapped within the charged droplet before repeated cycles of solvent evaporation and Coulombic fission leads to the release of this solvent shell with the remaining charges being transferred onto the protein ion. For denatured proteins and proteins with intrinsically disordered regions, the CEM model is preferred. The high coverage of hydrophobic sections within these structures facilitate migration towards the surface of the droplet before the protein terminus undergoes a 'chain ejection', leading to the eventual ejection of the whole protein ion from the charged solvent droplet. Smaller solvent embedded analyte ions tend to migrate towards the edge of the charged droplet before being ejected out via a salt bridge (IEM). More recently, Beveridge and co-workers described an intermediate model between CRM and CEM, whereby the more extended hydrophobic section of a protein ion undergoes chain ejection, consequently dragging the globular hydrophilic protein section through the droplet or stalling the chain ejection process as the globular section undergoes CRM[60].

1.4.2. Mass Spectrometry for PTM Identification and Localisation

For a given protein, its amino acid sequence governs its overall monomeric mass. Protein mass spectrometry involves the analysis of ionised proteins and peptides via elucidation of their mass to charge (*m/z*) ratios. From this, the deconvoluted mass of the ion can be acquired and cross-referenced against a list of known masses for all protein/peptides for a given species by data processing software. Faster and more effective ionisation techniques, improved sensitivity, resolution and dynamic range (detection range of the

instrument) of mass analysers, higher resolution chromatographic methods for the separation of ions before mass analysis, and increased computational power for data processing have revolutionised the utility of mass spectrometry for the identification of proteins/peptides[61]. Further, the mass difference of the protein/peptide ion compared to its expected mass (from its amino acid sequence) can be cross-referenced against a list of known PTM masses to assign the specific PTM (or multiple PTMs) on the protein/peptide. For the accurate characterisation of protein proteoform variability (facilitated by PTM-crosstalk), it is necessary to characterise the specific amino acid site(s) that is modified. A large contributor to the explosion of site localised PTM identifications from mass spectrometry experiments has been the development of a versatile array of ion activation methods[62]. With appropriate ion activation, the protein/peptide ion can be fragmented to form product ions for *m/z* analysis to generate an MS/MS spectrum corresponding to the parent precursor ion, known as tandem mass spectrometry (also referred to as MS/MS or MS^n). This MS/MS spectrum can then be computed to decipher the amino acid sequence of the protein/peptide and its associated PTM(s). Thus, for a given experimental condition, a protein's preference for a specific combinatorial PTM-decorated proteoform can be elucidated. The protein/peptide fragmentation technique determines the type of product ions that are generated, with different techniques giving rise to different product ion types, and thus different chemical information. As such, the choice of fragmentation technique is an important factor in the information that the experiment generates, particularly in the context of PTM site localisation.

1.4.3. Fragmentation Techniques for PTM Characterisation

1.4.3.1. Collision Induced Dissociation

Developed by Jennings, McLafferty and Bryce, Collision Induced Dissociation (CID), also referred to as Collision Activated Dissociation (CAD) is one of the most used fragmentation techniques for the identification and site localisation of PTMs[63]. Higher Energy Collision Induced Dissociation (HCD) is a vendor specific offshoot CID[63]. The differentiating factor for HCD relates to the higher energy applied for this fragmentation technique relative to CID, leading to increased fragmentation of the precursor ion. In this section, both techniques are referred to as CID. In CID, ions are introduced into a chamber of low concentration inert gas and an electric potential accelerates the precursor ions, increasing its kinetic energy. Each ion then collides with an inert gas (often Helium, Nitrogen (N_2), or Argon). This collision stimulates a transfer of the precursor ion kinetic energy into vibrational energy, before fragmentation of the ion at the weakest bond (which is often the amide backbone)[64]. As described by the 'Mobile Proton Model', multiple fragmentation pathways compete, with the final fragmentation pathway driven by the most energetically favourable proton transfer within the precursor ion during the fragmentation process[64,65]. Upon CID, a proton within the precursor ion switches between multiple sites, favouring a basic site such as the N-terminal amine group or a lysine/arginine ε-amine group[64,65]. In the most common fragmentation route, the protonated amine-Nitrogen then undergoes nucleophilic attack by the carbon at the adjacent carbonyl group, driving cleavage of the backbone peptide bond to produce a *b* and *y* ion (Figure 1.3). Despite the relative efficiency of protein/peptide ion cleavage, CID dissipates the influx of internal energy generated upon collision across the precursor ion, often resulting in cleavage of labile bonds. This can translate to the loss of weak

covalently bound PTMs or misassignment of PTM site localisation[66]. When considering that most phosphosites are thermodynamically more stable in basic conditions, yet common proteomics sample preparation steps involve acidic buffers and high temperatures, the identification of novel phosphosites may be compromised by suboptimal fragmentation[67,68]. Further, structurally uninformative neutral losses such as water, ammonia and carbon dioxide are often observed with CID, complicating fragmentation spectra and thus reducing the probability of the data processing software to compute the MS/MS spectrum to a peptide/protein[69].

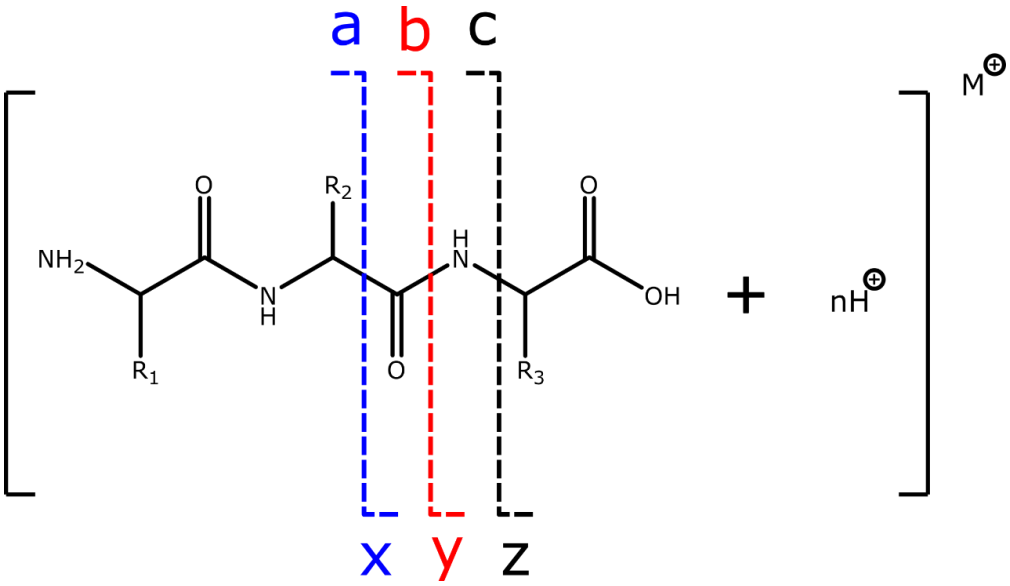


Figure 1.3 | Different fragmentation techniques of a peptide/protein precursor ion produce unique product ions due to differential fragmentation pathways. CID/HCD typically generates *b*/*y* ions (red), and ECD/ETD typically generates *c*/*z* ions or *a*/*x* ions (blue/black).

1.4.3.2. Electron Capture Dissociation & Electron Transfer Dissociation

Electron Capture Dissociation (ECD) was first described by Zuberev and co-workers[70], and has since become a common fragmentation technique for bottom-up and top-down proteomics (discussed in Section 1.6 and Section 1.7.2 respectively)[71]. During ECD, a

heated filament ejects electrons which are captured by the analyte of interest to generate charge-reduced species[71]. These intermediate ions then undergo electronic rearrangement to generate fragment ions primarily by cleavage of the N-C peptide bond. This generates *c* and *z* ions as characteristic fragment products of ECD (Figure 1.3). The precise mechanism of electron driven fragmentation is still debated, with two models proposed: the Cornell- and Utah-Washington-mechanisms[72]. Other minor ECD fragmentation pathways generate $a^{\bullet}/y$ ions, b/y ions and amino acid side chain cleavage[71]. Because of the preservation of labile PTMs with ECD (compared to CID), the technique has proved useful for the identification and localisation of PTMs including phosphorylation[73,74], acetylation[75], ubiquitination[76] and SUMOylation[77]. Initially, ECD was solely compatible with FT-ICR instrumentation, due to the need to trap the ions and electrons (although this is no longer the case with improved hardware). In response to this, Syka and co-workers[73] developed Electron Transfer Dissociation (ETD), as an analogous fragmentation technique to ECD, with the exception that electrons are transferred to the analyte from a radical anion such as anthracene or fluoranthene[73,78]. The major drawback of ECD/ETD based fragmentation relates to its inefficient cleavage of 2^{+} precursor ions relative to CID. The predominant proteomics methodology for high-throughput protein/PTM identification, bottom-up proteomics (discussed in Section 1.6), generates precursor ions that are primarily in the 2^{+} charge state[71], compromising the complementarity of ETD with bottom-up proteomics. However, ETD shows greater protein and peptide sequence coverage for higher charge state ions when compared with CID[79].

1.4.3.3. Ultraviolet Photo-Dissociation

CID and ECD/ETD fragmentation techniques have proved vital for the assignment and characterisation of proteins as well as the PTMs that decorate them. However, alternative fragmentation methodologies with different fragment product ions have been sought to provide additional structural information relating to PTM site localisation. Photodissociation techniques, that is, the emission of photons towards the analyte for absorption and energy deposition followed by fragmentation, have been developed in response[80]. The energy can be deposited by one or many photons, ranging from 0.1 eV using a CO_2 laser, to 3.5 eV from a xenon-fluoride excimer laser. Photodissociation methods range in the energies of the emitted photons, from infrared (780 nm - 1 mm), termed Infrared Multiphoton Dissociation (IRMPD) to high energy ultraviolet (40 nm - 80 nm) photodissociation (UVPD). In photodissociation based fragmentation, the photon is absorbed by the analyte and reaches an excited electronic state that is determined by the energy of the photon, before relaxing back down to its ground state via one of many possible higher energy fragmentation pathways that are not accessible with more common dissociation techniques such as CID and ECD/ETD. UVPD can itself be further categorised according to the wavelength used, including vacuum-UV (10 nm – 200 nm), middle-UV (200 nm – 300 nm) and near-UV (300 nm – 400 nm), with the wavelength inversely proportional to the energy provided by the photon[81]. Girod and co-workers combined 10.6 μm IRMPD and 213 nm UVPD, termed High Low Photodissociation (HiLoC) on a selection of phosphorylated, sulfonated and GalNAc modified peptide ions to report peptide sequence coverage up to 100 % and localised PTM sites due to the production of diverse peptide backbone fragmentation producing *a/x, b/y* and *c/z* ions[82]. Recently, 193 nm UVPD and ETD was applied to identify over 300 histone

modifications, with fragment ion types that were uniquely generated by UVPD proving essential for extensive characterisation of the most heavily modified histone peptides[83]. In a separate experiment of histone PTM regulation, UVPD contributed to sequence coverage gain of 15 % relative to just HCD[84].

1.4.3.4. Hybrid fragmentation techniques

Each fragmentation method generates its specific product ion types, with its own relative advantages/disadvantages relating to sequence coverage, cleavage of labile PTMs and PTM site localisations. Thus, the application of sequential or combinatorial fragmentation techniques for the generation of more informative product ions (by virtue of more product ion types from multiple fragmentation pathways) is known as hybrid fragmentation[85]. For example, Coon and co-workers developed a supplemental activation method for high-efficiency ETD of doubly protonated peptide precursor ions[86]. Here, non-dissociated electron transfer products ($[M+2H]^{+\bullet}$) of 2^+ precursor ions were targeted for supplemental CID to improve ETD cleavage efficiency in a methodology referred to as ETcaD[86]. Mohammed and co-workers built on this concept to describe a method that performed CID on not just unreacted precursor ions but all ions after ETD. Thus, the presence of both *b*/*y* and *c*/*z* ions generated richer fragmentation spectra and substantially increased peptide sequence coverage and thus PTM identification[87]. The same group then characterised this technique for phospho-proteomics to demonstrate superior phosphosite localisation scores compared to ETD and HCD[88]. Further, HCD often cleaves the labile phosphoester linkage of phosphate bonds, with the signature neutral loss of Δ98/Δ80 Da. Thus, so called neutral loss triggered EThcD, which performs ETD on peptides that generate these signature

phosphate-loss ions upon HCD (MS^3), can be performed to improve phosphosite localisation[89]. More recently, Zubarev and co-workers have developed a segmented linear ion trap for multidimensional multiple stage tandem MS, known as the Omnitrap. This device includes fragmentation capabilities via external injection of reagent ions, radical neutral species, photons, electrons and collisions with neutral species to generate information rich spectra[90]. Ultimately, hybrid fragmentation techniques can generate more informative spectra to increase the confident localisation of PTMs. In the context of pursuing the combinatorial PTM status of proteins (proteoforms), such techniques provide added value.

The development of the above precursor ion fragmentation techniques has vastly improved our ability to distinguish different types of site localised PTMs for the improved characterisation of proteoform complexity. However, the accurate, sensitive, and rapid identification of the precursor, as well as the product ions, is determined by the mass spectrometry instrumentation that detects these informative species. As such, a diverse range of mass analysers with differential methods of analysis and associated mass resolution (for mass accuracy), speed of analysis (for high-throughput analysis of complex samples), and sensitivity (for the identification of biologically relevant low stoichiometry PTM containing ions) have been developed.

1.4.4. Mass Analysers

1.4.4.1. Orbitrap

Developed by Alexander Makarov, the Orbitrap has become one of the most relevant mass analysers for the investigation of biological samples. Its design was inspired by the

FT-ICR mass analyser[91]. Briefly, the Orbitrap generates an electromagnetic field to trap ions into an oscillatory motion around a central spindle-shaped electrode inside an outer barrel-shaped electrode, with different ion packets oscillating with different radii (Figure 1.4). As well, the ion packets undergo oscillations in the axial dimension (x axis) as they move from one end of the outer electrode to the other. The two ends of the outer electrode, separated by a gap between them, detect the ions as they oscillate between them. This axial oscillation frequency relates to the ions' mass to charge ratio (*m/z*). The axial oscillatory motion of each ion packet, corresponding to different ions, is measured simultaneously with the output represented by an image current. These data are then converted from the time domain into the frequency domain using Fourier Transformation, which is then deconvoluted into the *m/z* spectral data that is visualised by the user[91].

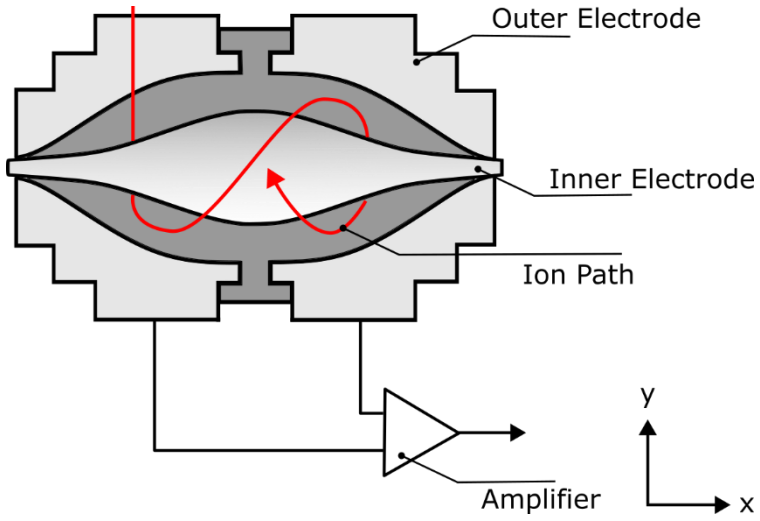


Figure 1.4 | Schematic representation of the Orbitrap mass analyser. Ions are injected into the analyser perpendicular to the radial oscillation path before orbiting the inner electrode. The outer electrode is physically separated, with each sub-electrode detecting the ion packet as it reaches its end during axial oscillation to generate an image current before amplification and Fourier Transformation into a mass spectrum.

1.4.4.2. Linear ion trap

Developed from the 3-dimensional quadrupole ion trap of Wolfgang Paul, the linear ion trap is typically comprised of three segments, each containing four electrodes to which both Alternating Current (AC) and Direct Current (DC) are applied to confine analyte ions in space (Figure 1.5). A DC trapping voltage is applied across the three segments to generate a potential well, thus trapping the ions that have entered from upstream ion optics. Briefly, a radio frequency (RF), termed 'main RF' is applied from the electrodes to radially confine the analyte ions, as well as activating them towards increased motility as they undergo 'secular motion' within the ion trap. The relative mobility of the ion for a given main RF amplitude is inversely proportional to its *m/z*. Further, ramping of another RF voltage, termed 'exit rod AC', is applied to a pair of electrode rods that contain a slit for ion ejection, such that when the ion's secular motion reaches a given frequency that corresponds to that of the exit rod AC, the ion enters resonance. At this point, the ion's secular motion is increased beyond the radial trapping capacity of the main RF. Consequently, these ions are ejected through the exit rod slits, out of the ion trap and towards a detector. With the main RF, the exit rod AC and the intensity of detection all known, the *m/z* of the ions can be derived and plotted against relative abundance to generate a mass spectrum.

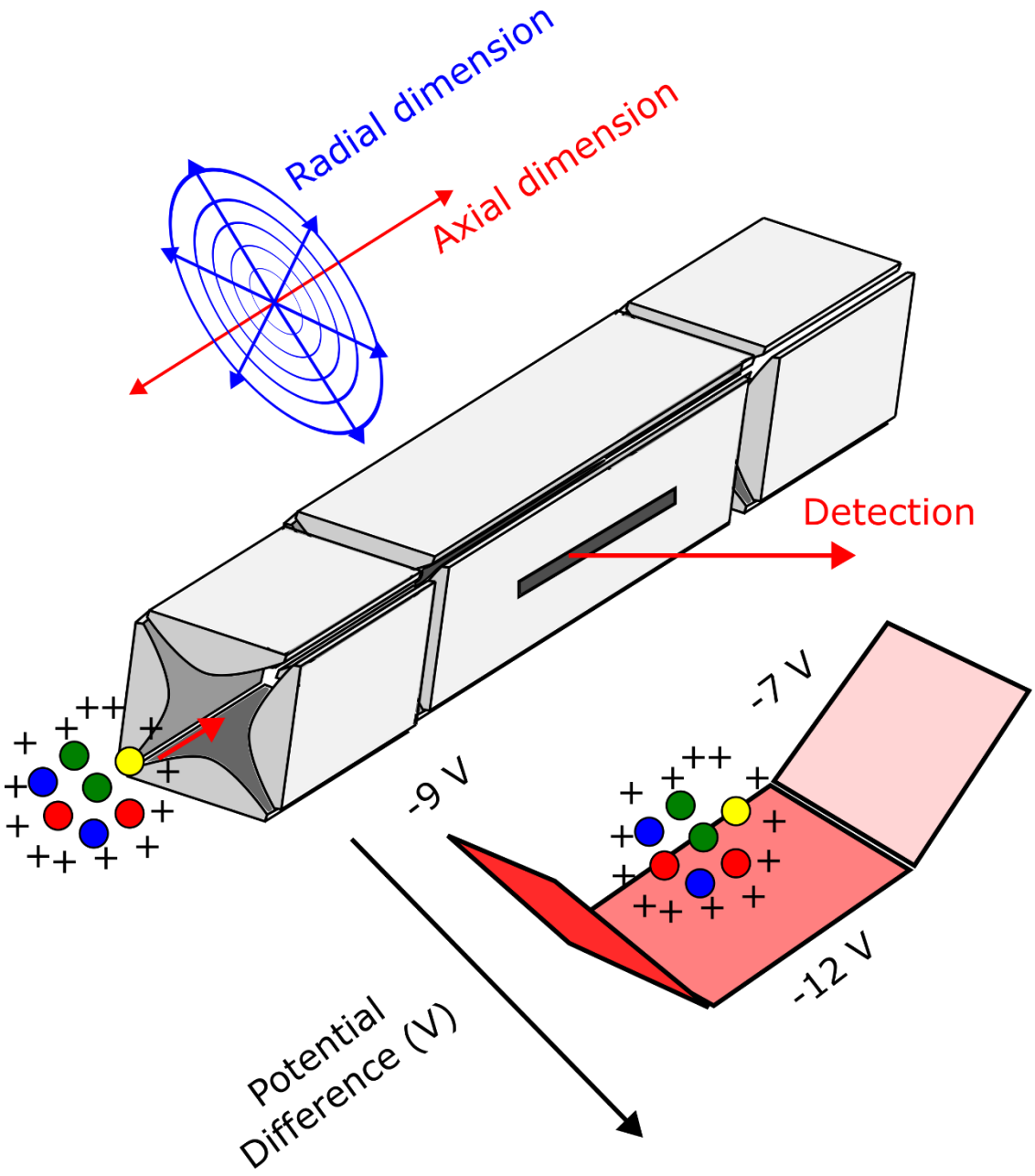


Figure 1.5 | Schematic representation of linear ion trap mass analyser. Ions are guided into the first segment of the trap. The potential energy well, that is generated by axial DC offsets, facilitates the trapping of analyte ions in the main segment. Radially, a main RF alternating current is applied to trap the ions in the radial dimension as well as activating their secular mobility. Application of an exit rod AC current to stimulate ion resonance (in the radial dimension) results in ejection of the ions, via the exit lens slit, towards detection. *Figure based on J. P Savaryn et al Figure 4*[92].

1.4.4.3. Quadrupole

Also known as a 2D quadrupole trap, the quadrupole has similarities with the linear ion trap and was also designed by Wolfgang Paul, alongside his student Helmut Steinwedel[93]. Whilst the ion trap applies an RF voltage in one plain (xy), the 2D quadrupole subjects its ions to an additional quadrupolar Direct Current (DC) in another plain (xz). A positive RF voltage is applied to two opposing rods to radially confine ions and a negative DC voltage is applied to the other two opposing rods (of opposite charge) (Figure 1.6). This produces a complex effect on the electric field experienced by the ions. The RF voltage stimulates secular motion of the ions, with overexcited lower *m/z* ions either colliding with an electrode (and neutralising) or escaping out of the mass analyser through the space between the rods. This represents the 'high mass pass filter'. Further, upon ion excitation as they undergo secular motion, these positively charged ions (in positive ion mode MS) are attracted towards the negatively charged DC rods. However, the radial confinement of the ions from the RF rods counteracts this electrostatic attraction. Larger ions have too much inertia to be radially confined and so escape via neutralisation upon collision with the negative rod or into the space between the rods. This represents the 'low mass pass filter'. By finetuning the RF voltage of the positive electrode pair and the DC voltage of the negative electrode pair, a specific *m/z* range of ions can be filtered. The incorporation of a downstream DC offset across the quadrupole enables transmission of ions before further analysis/activation[92].

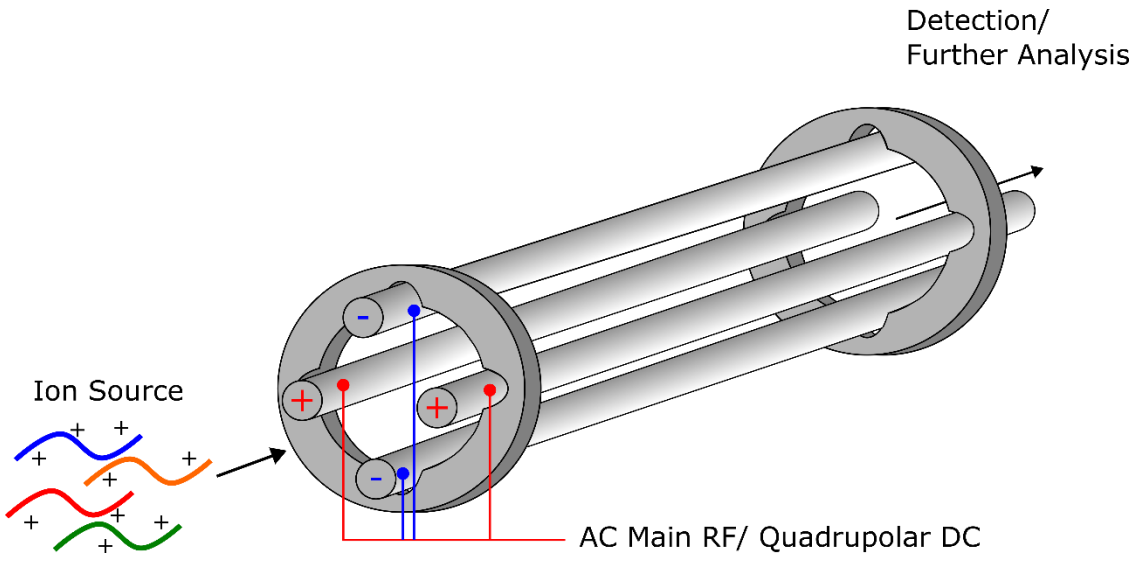


Figure 1.6 | Schematic representation of quadrupole mass analyser/filter. A combination of positive RF voltage and negative DC voltage are used to filter ions as they are transmitted through the quadrupole mass filter for downstream analysis via a DC offset voltage.

1.4.4.4. Comparative analysis of mass analysers

Whilst each mass analyser has its specific advantages and disadvantages relative to its competitors, the choice of analysis is guided by the biological question of interest (and perhaps more commonly, instrument accessibility). The quadrupole represents a relatively low-cost analyser of compact design with high scan speed and good reproducibility[94]. However, limitations include its relatively poor resolution (>4000 at *m/z* 40)[95], poor mass accuracy (>100 ppm)[96] and poor mass range (typically up to 3000 *m/z*)[97]. *Resolution (R) is defined as the ability to distinguish two peaks of slightly different mass-to-charge ratio* ΔM[96]*, in a mass spectrum, whereby at a given m/z:*

$$R = \frac{M}{\Delta M}$$

Linear ion traps represent a popular alternative due to their high storage capacity, relatively low cost, high scan rate and their ability to fragment ions and analyse these

ions with tandem mass spectrometry capabilites[98]. However, the inherently low resolution of linear ion trap spectra (60,000 at *m/z* 400) and limited scan range (2000 *m/z*) represent drawbacks of this mass analyser. The Orbitrap mass analyser, with its high mass resolving power (1,000,000 at *m/z* 200), high mass accuracy (1-2 ppm) and high dynamic range (~6000 *m/z*) is perhaps the most powerful of the analysers described in this section[99]. However, limitations of the Orbitrap include its relatively high cost and low scan speed[100]. Another important mass analyser, the Time of Flight (ToF) provides an alternative to the above instrumentation, with its high scan speed (~1000 full spectra per second), high mass accuracy (4 ppm at ~1000 *m/z*), theoretically unlimited mass range and moderate to high resolving power (>10,000 across all *m/z* ranges)[100,101]. *All values are representative of the mass analyser in general, with the specific values corresponding to specific models across multiple manufacturers*.

1.4.4.5. Hybrid instrumentation

Whilst each mass analyser has its own corresponding advantages and disadvantages, many instruments now incorporate multiple mass analyser modules to improve analyte characterisation for samples ranging from purified protein to high throughput analytes such as cell lysate. Perhaps the most common hybrid instrumentation due to its relatively low cost and broad application range, in the fields of drug testing[102], pharmacokinetics[103], proteomics[104] and extra-terrestrial atmospheric analysis of Mars[105] is the triple quadrupole mass spectrometer (QqQ). Here, the combination of three quadrupoles in a linear QqQ formation increases instrument sensitivity and throughput whilst providing a cheap alternative to other hybrid instruments[106]. The first and third quadrupoles function as mass filters, with the middle quadrupole functioning as a

collision cell for the generation of fragment ions. This design lends well to Multiple Reaction Monitoring (MRM) experiments, whereby an ion of a particular *m/z* is selected in the first quadrupole, before being fragmented in the second quadrupole and pre-selected product ions of the precursor are detected in the third quadrupole[104]. Another important possibility with hybrid instrumentation involves increasing the speed of analysis of each precursor ion within a sample, which is beneficial for high-throughput experiments. This is often measured by the duty cycle associated with specific instrument settings, which refers to the time it takes to acquire all necessary mass spectra for an ion before the instrument moves onto the next ion[107]. The duty cycle can be improved by simultaneous analysis of a given ion by multiple mass analysers, for example where one analyser acquires the precursor mass spectrum whilst another acquires the tandem fragmentation mass spectra for that precursor ion. The Orbitrap Eclipse Tribrid mass spectrometer, used throughout the work presented in this thesis, is another example of hybrid instrumentation. The complete duty cycle of the instrument, including all three mass analyser/filters has been shown to improve proteome coverage and PTM identification rates[108].

In the case of the Orbitrap Eclipse Tribrid mass spectrometer, one workflow involves ions entering the instrument from the ion source, before being guided via the advanced ion beam guide, upon which mass filtering is applied in the quadrupole and ions are stored in the C-trap. Then, precursor ions are directed into the Orbitrap for high resolution *m/z* analysis whilst simultaneously being selected by the quadrupole for fragmentation into product ions in the linear ion trap[109] before fragment MS/MS acquisition (Figure 1.7). As such, a high-resolution precursor peptide ion mass spectrum (complete with isotopic distribution) and multiple tandem MS spectra are acquired for

each peptide. The use of all three modules within the instrument vastly improves throughput, ultimately leading to greater protein identifications[108]. In this way, the mass filtering capacity of the quadrupole, high resolution of the Orbitrap and superior scan speed of the ion trap are coordinated to improve instrument productivity. However, with MS/MS spectra acquisition occurring in the ion trap, the inferior resolving power of this mass analyser (relative to the Orbitrap) compromises MS/MS mass accuracy[66]. Consequently, incorrect assignment of peptides/proteins can occur as well as misassigned PTM localisation[66]. In summary, throughput is prioritised (with improved duty cycle) over accuracy of PTM site localisation, and so the decision of speed vs resolution must be considered before the MS workflow is chosen.

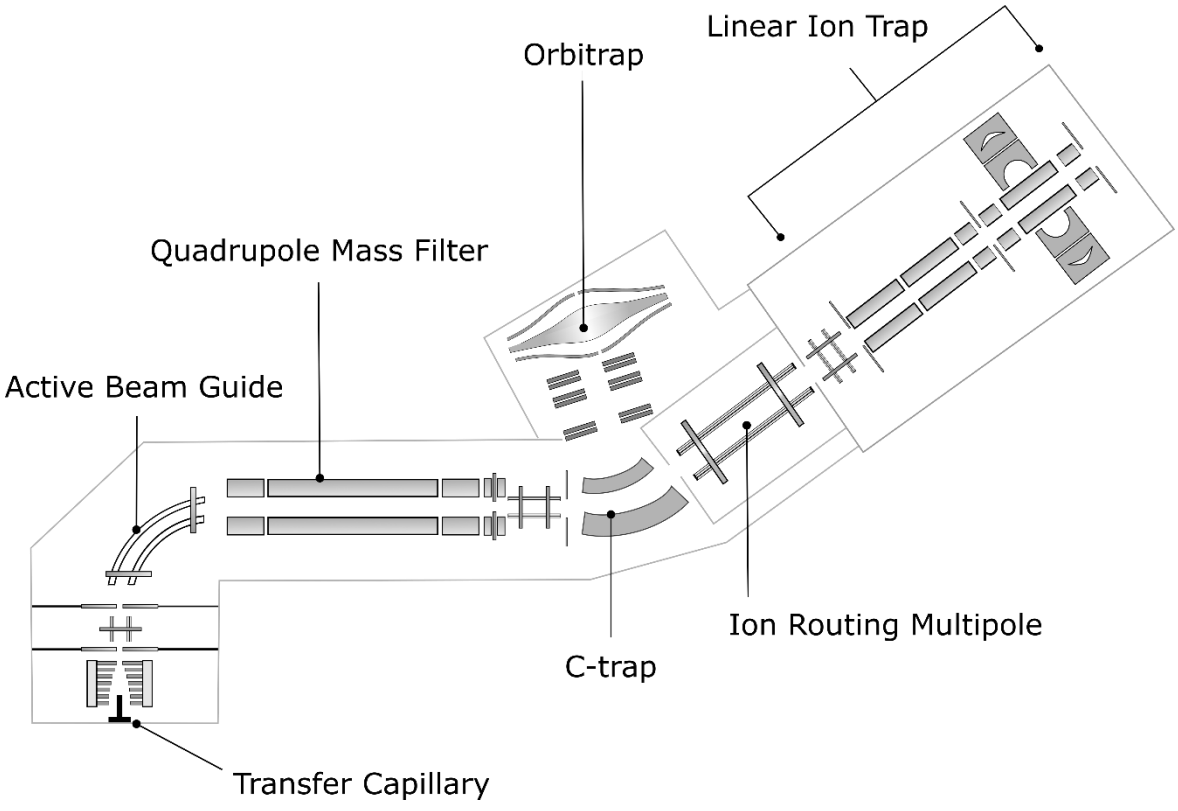


Figure 1.7 | Schematic representation of the Orbitrap Eclipse Tribrid mass spectrometer, with mass analysers/filters and ion optics. The hybridisation of multiple mass analysers within an individual instrument vastly improves throughput and sensitivity for optimised performance.

Thus far, fragmentation techniques and mass analysers have been considered in the context of improving the ability of mass spectrometry to identify and characterise combinatorial PTM modified proteins as individual proteoforms. However, to truly characterise the proteoform variability of a protein across biological contexts, all proteoforms that the protein can occupy must be identified and characterised. This represents a challenge for mass spectrometry, largely due to the nature of abundance distributions of protein proteoform variants. In many cases, some biologically relevant proteoforms occur in low stoichiometry relative to less modified proteoforms of the protein[66]. The identification of these species is challenging for the mass spectrometer due to limits of its dynamic range, as these ions are drowned out by more abundant ions. As such, low abundance ions are often not detected and selected for tandem MS analysis. Despite the development of online, pre-MS separation by liquid chromatography (LC) to increase species separation before mass analysis, the detection of low stoichiometry ions has continued to prove challenging. In response, orthogonal separation techniques to liquid chromatography for the separation of these ions to increase the probability of their detection are required. Ion mobility spectrometry, as an online pre-MS separation technique, has recently gained considerable traction[110–113].

1.5. Ion Mobility for PTM Characterisation

As discussed in Section 1.2, whilst identifying PTMs on proteins is a crucial step in improving our understanding of protein regulation, the site-specific nature of these modifications is equally important in the quest to characterise proteoform level regulation of a protein. Distinguishing the different PTM sites on a protein/peptide helps us to better characterise the multiple proteoforms that the protein can occupy. Whilst online liquid chromatography aids the separation of different peptides according to their differential physiochemical properties, ion mobility has been increasingly used in tandem with LC to further separate isobaric peptidoforms, boosting identification of multiple isobaric proteoforms with differential PTM sites[114]. Further, these techniques improve the sensitivity of mass spectrometry-based workflows to identify low stoichiometry PTM containing proteins/peptides that would otherwise be drowned-out by more abundant co-eluting peptides[115]. Ion mobility spectrometry (IMS) can be defined as a technique for the separation of ionised molecules in the gas phase based on their carrier gas driven mobility in an electric field[116]. IMS exploits the differential ion mobility of molecules in an electric field that is a product of their mass, collisional cross section and charge. From such experiments, molecule collision cross section values can be derived to probe protein conformational changes and unfolding pathways[117].

1.5.1. Ion Mobility Methodologies

Multiple IMS methods have been developed, including Drift Tube Ion Mobility Spectrometry (DTIMS)[118], Trapped Ion Mobility Spectrometry (TIMS), Travelling Wave Ion Mobility (TWIMS)[119], High Field Asymmetric Waveform Ion Mobility Spectrometry (FAIMS)[120] and more recently Cyclic IMS[118,121]. DTIMS involves the

injection of the ions into a series of stacked ring electrodes with a potential difference across the device axis to accelerate the ions down the device. As they travel, ions collide with counterflowing drift gas molecules, differentiating their arrival time at the end of the device by their ion mobility[118]. TWIMS uses similar principles, with the stacked rings instead propagating a 'travelling wave' across the device's axis. Ion arrival times are differentiated as higher ion mobility ions are able to 'roll over' the travelling waves more frequently than larger ions and so take longer to traverse the device[122] This technology has been coupled to a quadrupole-ToF (QTOF) instrument in a duty cycle known as Parallel Accumulation-Serial Fragmentation (PASEF) to increase sensitivity for multiple isobaric ions, with a reported 10-fold gain in peptide sequencing speed for bottom-up proteomics[123]. . For TIMS, ions are subjected to a parallel gas flow (unlike DIMS and TWIMS which generally use a stationary buffer gas) and an opposing electric field to differentiate ion arrival time based on ion mobility. More recently, cyclic IMS has been developed. Here, ions enter an array where they pass through a circular ion guide and are separated according to their ion mobility. In 'Multi Pass Acquisition', cyclic IMS enables these ions to undergo multiple passes around the circular array, increasing path length and thus the resolution of their separation. Further, with IMS^n slicing, these separated ions of interest can be sent to the array pre-store whilst the remaining ions are ejected to the mass analyser, the ions are then reinjected from the array pre-store into the array for further separation and analysis. With 'Activated Multi Pass Acquisition', the ion of interested can be energetically activated (e.g. for protein complex dissociation) after being sent to the array pre-store, and then reinjected into the array for further IMS analysis of its product ions[124]. In the context of mass spectrometry, these ion mobility strategies separate out mixtures of ions, such that for multiple

proteoforms corresponding to a protein, each species is independently injected into the mass analyser for detection. Ultimately, more proteoforms are detected, giving a better representation of the proteoform variation that can be occupied by a protein.

1.5.2. High Field Asymmetric Waveform Ion Mobility Spectrometry

1.5.2.1. Principles of FAIMS

High Field Asymmetric Waveform Ion Mobility Spectrometry (FAIMS), first developed in 1982[125], can be described as the atmospheric separation of ions on the basis of their mobility in high and low electric fields. Ions are passed through two parallel electrodes with a carrier gas (usually Helium or Nitrogen (N_2)) and subjected to alternating high and low electric fields of opposing charge. This is generated by an asymmetric RF waveform (with an amplitude defined by the Dispersion Voltage (DV)) that alternates between positive and negative polarity. The RF waveform is manipulated such that an equal product is generated from high and low electric fields (the product of the dispersion voltage and its duration is equal for both positive and negative polarity) (Figure 1.8a). FAIMS exploits the differential mobility of ions in high and low electric fields for their separation[126]. Ions can be stratified according to their behaviour in these conditions. As the electric field strength increases, ions either increase in mobility (type A), decreasing in mobility (type C) or increase then decrease mobility (type B)[127]. As the ions enter the FAIMS device post-ESI (electrospray ionisation), they drift towards the electrode due to its negative polarity's electrostatic interaction with the positive ions. Then the RF waveform switches polarity, causing the ion the repel the now positive polarity of the electrode. This ion behaviour is repeated as each cycle of changing waveform occurs until the ion trajectory reaches the electrode and is neutralised. For a small packet of ions

of specific ion mobility range in this asymmetric field, the trajectory of their mobility towards the electrode in the negative field is offset by their mobility in the positive field. These ions therefore pass through the device and out towards the mass analyser (Figure 1.8a). By superimposing a small direct current onto the electrode known as a Compensation Voltage (CV), ions of differing ion mobility in this asymmetric field can be directed through the device. Thus, different ions can be filtered towards the mass spectrometer by manipulating the CV. Further, scanning across multiple CV's can filter different ions according to their differential ion mobility in an asymmetric field.

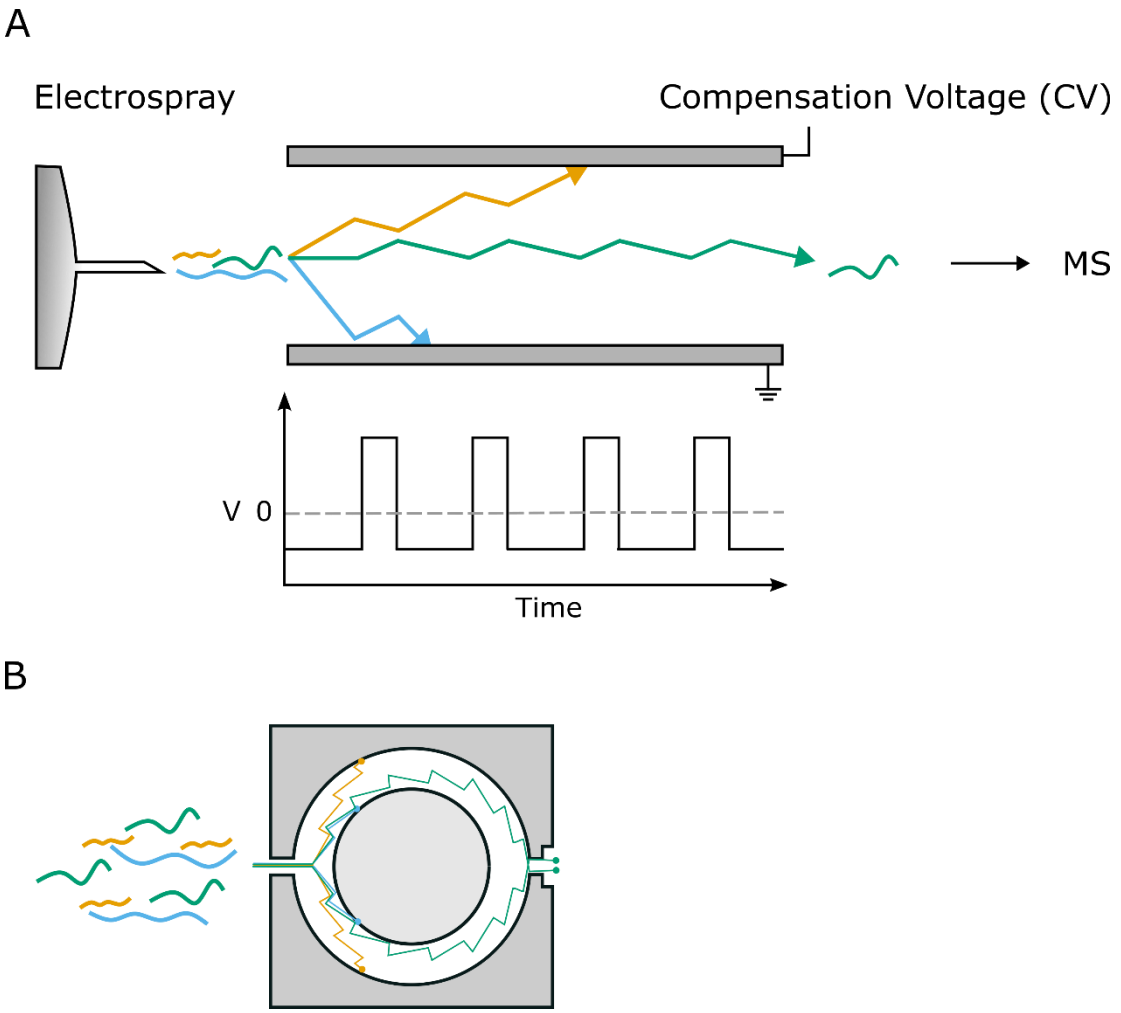


Figure 1.8 | Schematic representation of FAIMS device. (A) Following ESI, the ion path is dictated by the ion's behaviour in high and low electric fields. As the ion traverses across the device, it eventually collides with the electrode and is neutralised. However, for a group of ions

of specific ion mobility range, their trajectory in the negative field is offset by their trajectory in the positive field, such that they pass through the device and into the mass spectrometer for analysis. (B) Schematic representation of the cylindrical FAIMS device. The cylindrical design has an inhomogeneous ion focussing affect, improving the efficiency of focussing ions of high field dependence through the device, whilst increasing the neutralisation of ions of low field dependence.

1.5.2.2. Types of FAIMS device

Planar FAIMS (p-FAIMS) involves two parallel planar electrodes. Without a necessary deviation in ion trajectory for access through the device, p-FAIMS can be used in non-FAIMS mode with the electrodes grounded[120]. The interlinking of multiple p-FAIMS ion guides into parallel stacks of plates within a chip is known as UltraFAIMS[128]. The reduced electrode gap (35-100 μm) allows for higher dispersion voltages, which corresponds to increased separation of ions of differing mobility[128]. Cylindrical FAIMS (c-FAIMS) was initially developed by Carnahan and co-workers and showed superior sensitivity to p-FAIMS on account of its ion focussing effect[129]. Here, two coaxially aligned cylinders representing parallel electrodes are used as illustrated in Figure 1.8b. This design creates an inhomogeneous ion-focussing effect, whereby ions of high field dependence are focussed towards the mass analyser whilst ions of low field dependence are less effected and collide with the electrode to be neutralised[130]. Structural improvements in the ion separating capacity of c-FAIMS have included the reduction of the electrode gap from 5 mm to 2.5 mm, and its compatibility with ESI, elevating its utility in mass spectrometry-based analysis[131]. More recently, the commercialisation of the FAIMS Pro Interface[132] by Thermo Fisher Scientific, with a further improved electrode gap width of 1.5 mm, has fast-tracked its use in protein analysis. This technology is compatible with ten different Thermo Fisher Scientific mass spectrometers, and has demonstrated improvements in ion sampling, electric field

strength intensity and the use of helium-free carrier gas[132,133]. In the context of PTM proteoform characterisation, FAIMS' ability to resolve multiple protein/peptide ions of differing PTM status translates to the improved detection of these ions by mass spectrometry, informing proteoform variability for a given protein (as discussed in Section 1.6.4), as the biochemistry community strives for complete proteoform characterisation.

Thus, improvements in ion mobility-based species separation pre-MS, fragmentation techniques and mass analysers have together aided the mass spectrometrist's toolbox for the targeted detection of multiple proteoforms of a given protein. Another important feature of mass spectrometry experiments for the characterisation of combinatorial PTM analysis is the choice of proteomics workflow. Multiple workflows have been developed, with bottom-up proteomics being most widely applied on account of its high-throughput nature for the analysis of complex samples together with well characterised data processing software[8,134]. As such, this methodology has identified thousands of site localised PTMs, providing valuable information for PTM crosstalk related studies and overall proteoform analysis.

1.6. Bottom-Up Proteomics

1.6.1. Overview

In bottom-up proteomics (also known as shotgun proteomics or LC-MS/MS), a purified protein or complex mixture is subjected to proteolytic digestion to form peptides (typically using trypsin). These peptides are then separated by on-line liquid chromatography (typically reverse-phase LC) before being eluted and ionised prior to entering the mass spectrometer for detection (Figure 1.9). For a selected peptide ion, a precursor mass spectrum is acquired together with tandem mass spectra corresponding to its fragmentation products (typically using HCD). The tandem mass spectrum, containing a series of *b* and *y* ions that represent the cleavage products of the peptide at multiple carbonyl C-N bonds (with associated neutral loss) and often the precursor ion itself (Figure 1.3), is then matched to a theoretical mass spectrum from the *in-silico* digestion of the protein[135,136]. Provided the probability of the match between the tandem MS spectrum and the theoretical spectrum satisfies the preselected statistical criteria, the amino acid sequence of the peptide can be inferred (computational approaches to peptide/PTM identification in bottom-up proteomics are discussed in Section 1.6.5). Thus, the presence of a protein in the sample can be derived from the identification of multiple tryptic peptides within the amino acid sequence. Rapid advances in mass spectrum resolution, mass accuracy, scan rate, dynamic range and fragmentation techniques, supplemented with the advent of hybrid instrumentation and data processing computational efficiency has enabled the routine detection of thousands of proteins within a complex sample[137].

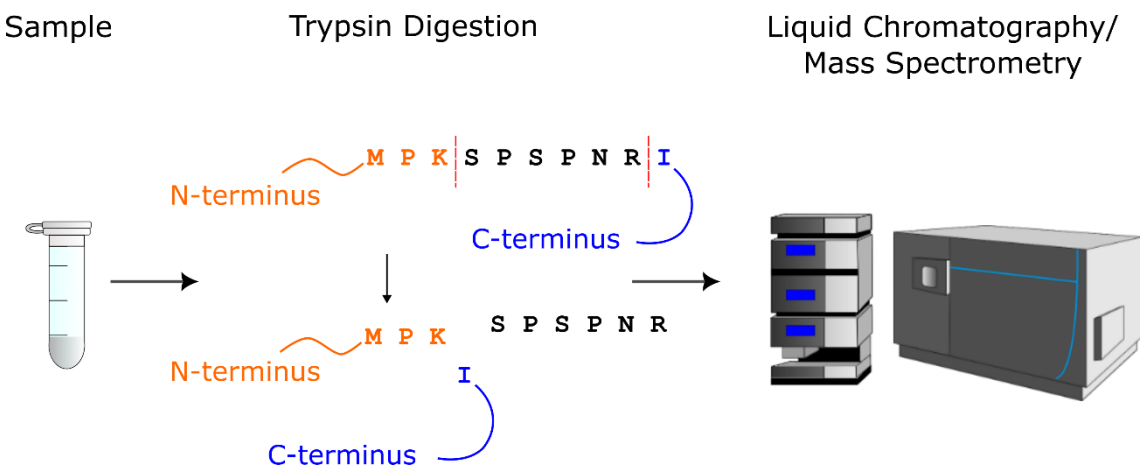


Figure 1.9 | Schematic representation of bottom-up proteomics workflow. Sample is enzymatically (or less often, chemically) digested before the resulting peptides are subjected to LC-MS/MS.

Currently, the gold-standard for proteomics analysis is bottom-up proteomics owing to its high sensitivity and throughput[134]. Typically, such experiments involve mass spectrometry workflows that can be categorised as Data Dependant Acquisition (DDA) or Data Independent Acquisition (DIA). For DDA, a preselected number of the most abundant precursor ion peaks in the MS^1 scan are sequentially selected and fragmented for tandem MS analysis. This avoids the repeated analysis of the same peptide, increasing the number of identified peptides from each ion packet that is eluted from the LC. In DIA, for each LC eluted ion packet; precursor ion fragmentation is performed sequentially across *m/z* windows or 'bins'. Therefore, less abundant peptides have an increased likelihood of tandem MS selection and characterisation, as the instrument is not biased towards selecting the most abundant peptides[138]. This approach has been reported to promote PTM identification on account of improving the likelihood of detecting low stoichiometry peptide ions[139]. For example, Mann and co-workers combined DIA with glycine-glycine (diGly) antibody peptide enrichment (whereby

diGly stubs represent remnants of ubiquitination after trypsin digestion) to identify 35,000 diGly peptides and uncover dynamic ubiquitination sites in the context of circadian regulation. They reported that DIA doubled the number and quantitative accuracy of diGly identifications relative to DDA[140]. An important aspect of the utility of bottom-up proteomics derives from the high-throughput analysis of thousands of proteins in a single experiment. For each tandem MS spectrum, the series of *b* and *y* ions (with HCD fragmentation) can localise a PTM to a specific amino acid via detection of the additional mass of the PTM. The identification and curation of thousands of site localised PTMs onto databases such as PhosphositePlus[141] and iPTMnet[142] demonstrates the huge impact of this technique to our understanding of protein biochemistry. For the goal of complete proteoform characterisation, such site localised PTM curation is invaluable. As is the case with all proteomics techniques, aspects of the above workflow can be modified for the specific biological question of interest (and availability of the technology). Thus, multiple modified workflows for the analysis of PTMs exist.

1.6.2. PTM Enrichment Strategies

Despite bottom-up proteomics' capacity to identify PTMs from tandem MS spectra, identification rates for cell lysates tend to be in the order of tens, with variability to its success depending on data processing search strategies. Many PTM containing peptides occur in low stoichiometry, relative to their corresponding non-modified peptide isoforms[66]. As well, some more abundant multiple-PTM peptides do not ionise well. These issues hinder identification of such peptides, that are derived from biologically relevant modified proteins. Consequently, these PTMs are missed at the expense of more abundant co-eluted peptides. Chemical or affinity enrichment strategies in the sample

preparation workflow have been developed to counter such issues, vastly improving PTM identification rates. Affinity enrichment, whereby antibodies (often immobilised onto a resin) specifically bind PTMs before bottom-up proteomics analysis, is routinely used for the enrichment of such modifications as phosphorylation[143], acetylation[144], methylation[145], and ubiquitination[146]. Further, phosphosite identification has been improved using Immobilised Metal Affinity Chromatography (IMAC) and Metal Oxide Affinity Chromatography (MOAC) for the chelation of negatively charged phosphate-moieties. Another PTM enrichment strategy involves the chemical labelling (e.g., avidin) of a PTM via click chemistry, followed by immunoprecipitation against the tags' cognate binding partner (e.g., streptavidin). Also, separation of complex peptide samples into multiple fractions, by the exploitation of the different chemical properties of modified-peptides, increases the concentration of PTM-peptides in the sample. As such, the probability of modified peptide precursor ion identification followed by tandem MS analysis is improved. Such techniques include strong cation exchange (SCX) and strong anion exchange (SAX) which separate peptides into fractions of differing ionic strength, and hydrophilic interaction liquid chromatography (HILIC), which separate peptides according to hydrophobicity. Ultimately, the fractionation method of choice is dependent on the PTM being investigated, with charge inducing PTMs (e.g. PO_4^{2-}) being more suited to ion exchange methods and charge removing PTMs (e.g. acetylation, methylation)[147] more compatible with non-ionic separation techniques.

In the context of combinatorial PTM analysis for PTM crosstalk, serial enrichment of different PTMs, termed SEPTM, has been developed. For example, Carr and co-workers combined IMAC for phosphoenrichment, antibody immunoprecipitation for the tryptic diGly remnant of lysine bound ubiquitin and antibody immunoprecipitation for lysine-

acetylation to identify multiple-PTM containing peptides as candidates for positive crosstalk[148]. SEPTM of the same PTM, with orthogonal enrichment techniques, can also be used to investigate individual PTMs. For example, IMAC and MOAC enrichment have been combined to improve phosphosite identification in proteomics on account of the orthogonal enrichment chemistry of the two techniques[149]. Despite the recent advances in PTM identification that have been realised from such enrichment strategies, PTM enrichment has drawbacks. For example, PTM enrichment strategies are often dependent upon consensus motifs that surround the PTM, thus biasing results based on the PTM's affinity for the antibody, which is partially dictated by the amino acid sequence that the PTM modifies. Further, with over 400 different PTMs now identified, a cognate enrichment technique is not available for many of these biologically relevant modifications. From an application standpoint, sample enrichment can reduce yields for downstream analysis and are time consuming additions to the overall experimental workflow. Crucially, as the combinatorial impact of multiple PTMs on a protein is now considered to be more relevant than isolated PTM regulation[150], the issues are exacerbated with SEPTM analysis. As such, development of online separation methods, bespoke mass spectrometry workflows and data processing strategies represent an opportunity to take MS-based PTM analysis to the next level. An alternative strategy for the enrichment of PTM containing peptides involves modification of the online pre-MS liquid chromatography step within the bottom-up workflow. The combination of multiple orthogonal LC techniques, termed Multidimensional LC (MD-LC) has shown promise in improving LC resolution for superior peptide identification, improving the likelihood of low stoichiometry PTM-containing peptide precursor ions being selected and fragmented for site localisation[151,152]. For example, Multidimensional Protein

Identification (MudPIT) involves the integration of multiple LC techniques into one workflow such as reverse phase (RP), SCX and a second RP separation before ESI[153,154]. Thus, more biologically relevant PTMs are detected. At the next stage of PTM characterisation, upon identification of a given PTM by the above enrichment techniques, biological characterisation of the PTM must be probed to improve understanding of the relevance of such spatiotemporal modifications in the context of cell behaviour. The ability to identify a protein's differential occupancy of specific proteoforms under different biological conditions is required for a more informed understanding of how the protein responds to environmental cues. Workflows for the quantitative characterisation of changes in the abundance of different peptidoforms of a peptide in response to different experimental conditions has proved incredibly powerful for this purpose. For such analyses, multiple quantitative proteomics techniques have been developed.

1.6.3. Quantitation

The ever-improving performance of modern mass spectrometers over the past two decades has driven the field of global proteomics towards accurate protein and PTM quantitation, rather than just identification[155]. As such, many techniques for the elucidation of protein/peptide quantitation have been developed, facilitated by the high mass resolution and scan speed of modern instrumentation. This has enabled the tracking of phenotype correlated protein changes upon different independent variables such as genetic modification[156], drug administration[157] and the manipulation of environmental factors[155,158]. For such experiments to provide biological value, tracking of protein and PTM abundance changes across conditions is vital. Quantitative techniques can be

stratified according to the level of protein/peptide quantitation investigated. Absolute quantitation, the investigation of actual abundances of proteins/peptides, involves the comparison of quantitative data from an analyte of known concentration, that has been spiked into the sample, to that of other proteins within the analyte. Relative quantitation studies the differential response of the protein/peptide in different conditions, with its upregulation/downregulation considered against independent variables. This often involves the use of stable isotope labelling of the sample and can be further stratified into categories depending on the method of the isotope's incorporation into the protein/peptide.

1.6.3.1. Label free quantitation

Label Free Quantitation (LFQ) can provide an alternative strategy to proteomics quantitation without the need for labelling techniques. Compared to labelling techniques, LFQ allows for comparison between a theoretically unlimited number of conditions, as it is not limited by the number of different metabolic labels that are available (also known as quantitative channels)[159]. LFQ compares protein abundances across multiple LC-MS/MS runs. The two major approaches towards LFQ include spectral counting and extracted ion chromatogram analysis. With spectral counting, relative quantitation is calculated by comparing the number of MS/MS spectra produced by the same protein across multiple LC-MS/MS samples[160]. However, the DDA MS set-up is not particularly compatible with the spectral counting methodology, as the correlation between the number of MS/MS spectra and the peptides abundance is compromised by the instruments' pre-programmed sequential selection of the most abundant ions to increase the number of precursor peptide ions that are sequenced[161]. Another issue relates to the

inherent positive relationship between protein amino acid sequence size and number of MS/MS spectra per protein, making quantitation of lower molecular weight proteins (<20 kDa) less accurate[162,163]. Whilst normalisation of such discrepancies is offset by data processing software packages, these issues reduce the overall accuracy and reliability of spectral counting[155]. Alternatively, quantitation based on precursor ion signal intensity from the extracted ion chromatogram and LC retention time can be performed. Here, the abundance from a given precursor ion (derived from the area under the peak curve or the peak height), together with its retention time is compared across multiple LC-MS/MS experiments to calculate relative abundance. A drawback of this methodology involves its poor compatibility with the bottom-up strategy of sequentially identifying precursor ions and acquiring MS/MS spectra. Whilst this quantitation technique specifically requires precursor ion identification, acquisition of tandem MS spectra reduces the time spent on acquiring precursor ion spectra to favour fragmentation and tandem MS acquisition. Thus, a compromise between sequence validation (via tandem MS acquisition) and quantitation (via precursor ion acquisition) must be sought as MS^1 and MS^2 compete for scan time[155]. Specifically in the context of PTM analysis, MS^2 spectra provide the data for PTM localisation to be deciphered. Also, extracted ion chromatogram based quantitation requires highly robust LC systems for retention time alignment between runs and marginally more computational processing power than spectral counting for downstream analysis[164].

1.6.3.2. Post-harvest peptide labelling with isobaric tags

Random variation in ion selection, ionisation efficiency and retention time drift across LC-MS/MS runs curtail accuracy and reproducibility for LFQ experiments[155]. Thus,

sample labelling, whether chemically or metabolically, has vastly improved the field of relative and absolute quantitation. Added to this, such techniques involve the pooling of samples from multiple conditions into one multiplexed sample for ease of data analysis and reduction in instrument time requirements. Initially, Isotope Coded Affinity Tagging was used to tag proteins of two separate experimental conditions with a biotin tag coupled to a stable isotope labelled linker containing 8^1H or 8^2H via a reactive group that targets cysteine sulfhydryl groups[165]. Since then, isotope-tagging was developed for peptides, from which relative quantitation of peptides across conditions could be compared using MS/MS fragmentation in bottom-up proteomics. For example, Tandem Mass Tag (TMT) quantitation involves the relative quantitation of peptide ions based on varying reporter ion abundances corresponding to each condition following collisional activation[166]. Figure 1.10 illustrates the chemical structure for 6-plex (six tags for six quantitative channels) reagents with ^{13}C and ^{15}N based stable isotopes. Asterisks represent atomic sites for the possible substitution of C^{12} and N^{14} for isotopic labels. The final tag ensures that the increased mass of the reporter ion segment is counterbalanced by the balance group segment such that the tagged precursor ion mass is the same across all quantitative channels. MS/MS fragmentation initially breaks the reporter ion off the precursor tag, and the abundance of the different reporter ions corresponding to different experimental condition peptides is used to derive relative quantitation. Following this, MS^3 can be performed to characterise peptide sequence and PTM site localisation (comparable to MS^2 in a normal bottom-up experiment). Coupling of this technique to high-resolution mass spectrometers facilitates accurate and sensitive quantitative proteomics analysis. Further, spiking a tagged peptide of known concentration can derive absolute concentrations of the same peptide across different conditions. Recently,

16-plex TMT has been used to identify >10,000 proteins and >10,000 PTM sites[167]. However, issues around coelution of peptides and contaminants within an isolation window can result in complex spectra that lead to an underestimated protein abundance[168,169]. Other peptide-tag based techniques such as isobaric peptide termini labelling[170] and dimethyl leucine labelling[171] work in a similar manner and have been employed for global quantitative proteomics experiments.

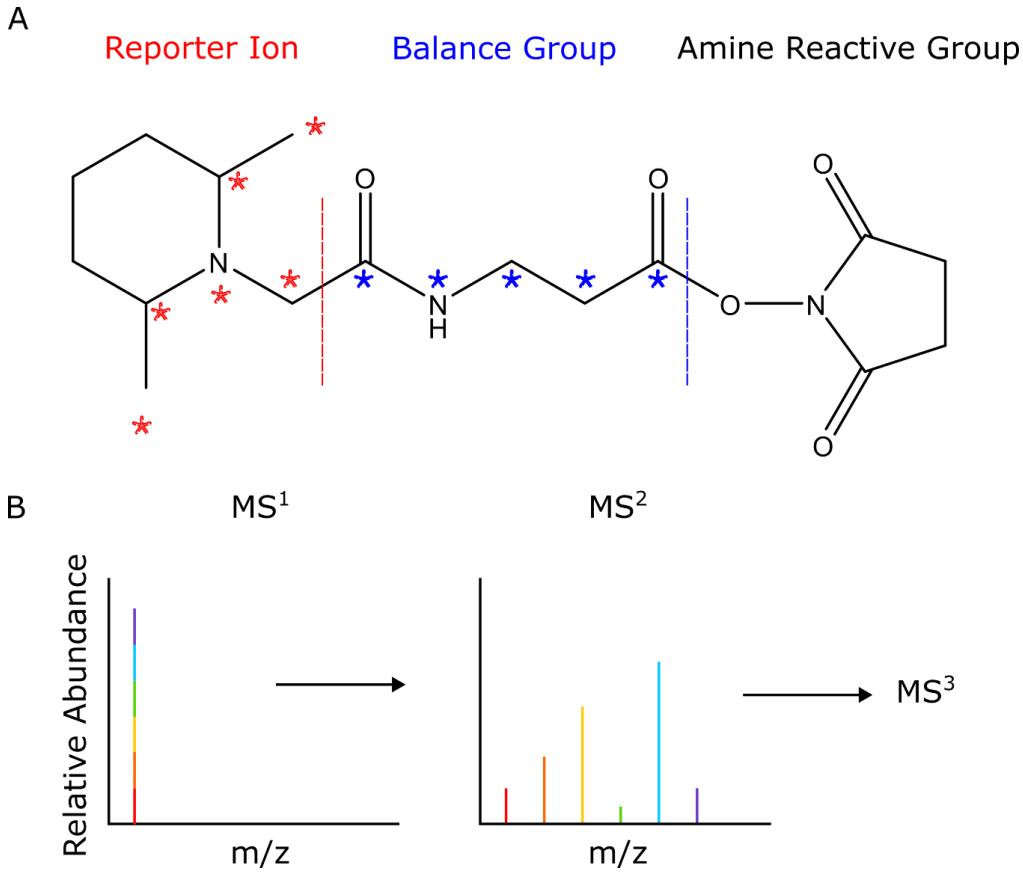


Figure 1.10 | Tandem Mass Tagging (TMT) quantitation methodology. (A) Molecular tag for 6-plex TMT labelling[166]. * represents isotopically labelable atom, Dashed lines represent labile bonds for tandem MS fragmentation. Each radiolabelled atom on the reporter ion is counterbalanced by a radiolabelled atom in the balance group such that for each precursor ion across all labelled quantitative channels, that correspond to the same peptide, their *m/z* is the same. (B) Precursor ion scan shows peptides from all six conditions have the same MS^1 *m/z*. MS/MS fragmentation breaks the reporter ion labile bond to release the reporter ion. Reporter ion abundance corresponds to peptide abundance. MS^3 fragmentation can then be performed to characterise amino acid sequence and possible PTMs.

1.6.3.3. Stable Isotope Labelling of Amino Acids in Cell Culture

Peptide-tagging quantitation requires the post-harvest incorporation of the tag, resulting in issues of sample loss and variability in tag-peptide reactivity. However, the metabolic incorporation of a mass tag into the proteins during cell growth, first described in 2000, eradicates such issues[172]. Amino Acid-Coded Mass Tags (AACT) are the most common metabolic tags, whereby AACTs are incorporated into the cellular nutrient media during cell growth. Thus, the cell's protein synthesis machinery ensures that for multiple experimental conditions, detectable mass differences between peptides of the same protein are used to derive changes in relative abundance[155]. Of the AACT methodologies, Stable Isotope Labelling of Amino Acids in Cell Culture (SILAC) is the most well-known. For a SILAC experiment involving three conditions (including a control), each sample is grown in a unique nutrient media, that typically contains alternatively labelled arginine and lysine amino acids. Whilst the control condition is grown in isotopically natural lysine and arginine (known as 'Light' media) , one experimental condition is grown in 'Medium' labelled media (containing lysine-4 and arginine-6 stable amino acid isotopes of lysine and arginine), and the other experimental condition is grown in 'Heavy' media (containing lysine-8 and arginine-10 stable amino acid isotopes of lysine and arginine) (Figure 1.11). As such, after trypsin digestion, all peptides contain at least one lysine or arginine (as trypsin cleaves at the C-terminus of lysine/arginine, except when there is an adjacent C-terminal proline) and so all peptides can be quantified across conditions. More recently, the development of neutron-encoding amino acid labelling (NeuCode) has doubled the number of quantitative channels to six[173], and its combination with other chemical labelling techniques has enabled up to 30-plex conditions in one quantitative Parallel Reaction Monitoring

(PRM) experiment[174] (PRM is similar to MRM, except all fragment ions are detected with high mass accuracy, rather than preselecting specific product ions for detection, as in MRM). Further, metabolic incorporation of isotopically labelled tyrosine has been used to probe phospho-tyrosine changes in the context of growth factor signalling cascades. For example, an MRM experiment tracked the phospho-status of 222 tyrosine residues across pre-defined peptides, at different times post-growth factor stimulation[175]. The accurate quantitative capacity of SILAC makes it a popular choice for the detection of differentially abundant PTMs across biological conditions, in combination with PTM-enrichment. For example, it has been used for accurate quantitation of histone PTM profiling across multiple biological conditions[176], phosphoproteomics analysis of breast cancer[177], redox dependent post-translational modification of cysteines[178], the differential nucleosolic and cytosolic global protein methylation status of mammalian cells[179] and the quantitative investigation of unconventional ubiquitination in the endoplasmic reticulum degradation pathway[180]. For improved identification of multi-PTM containing peptides in bottom-up proteomics workflows to better characterise protein proteoform variability (with the added layer of PTM quantitation across biological conditions), better separation of the complex mixture of peptide ions would aid in the detection of lower abundance multi-PTM containing peptides. As such, the incorporation of FAIMS into bottom-up proteomics workflows has shown promise in increasing peptidoform detection[133,181].

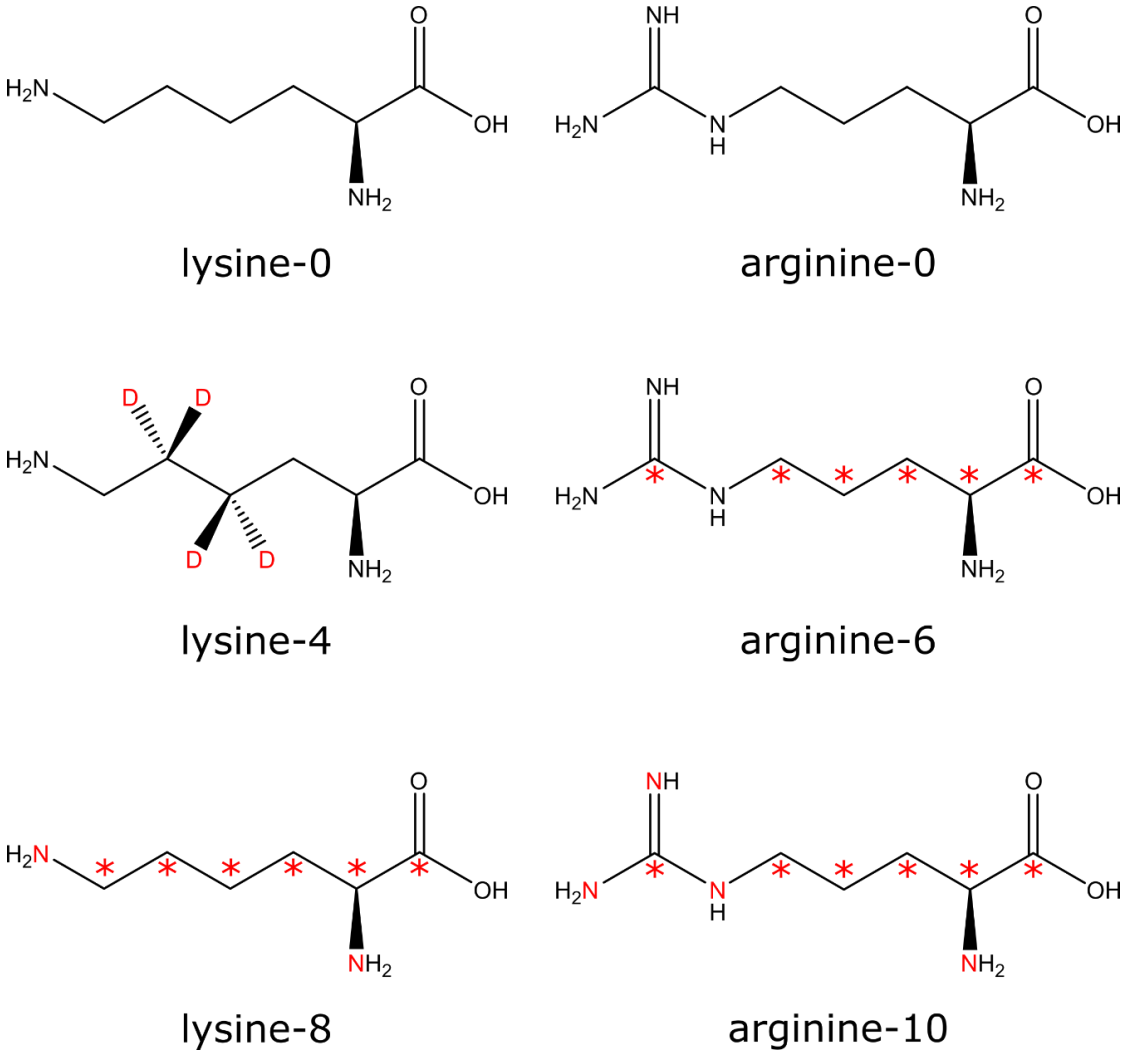


Figure 1.11 | Amino acid structures of lysine and arginine in SILAC. Red colour represents unnatural atomic isotope, where D is Deuterium, N is Nitrogen-15 and * is Carbon-13. Natural isotopes of lysine and arginine are incorporated into 'Light' labelled media, lysine-4/arginine-6 are incorporated into 'Medium' labelled media and lysine-8/arginine-10 are incorporated into 'Heavy' labelled media during a cell growth stage of a SILAC proteomics experiment. Following incorporation of the different SILAC isotopes for different conditions, the independent variables can be applied before cell lysis and protein extraction, digestion, cleaning, and LC-MS/MS analysis (Figure 1.12).

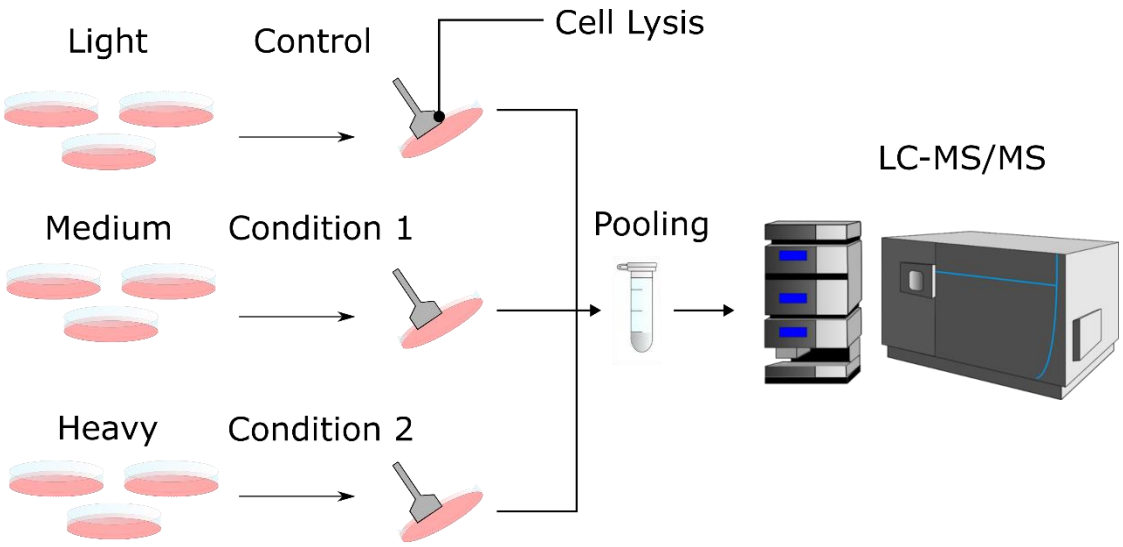


Figure 1.12 | Schematic representation of SILAC workflow. Cells are grown in labelled media such that protein biosynthesis incorporates the metabolic tag. After the cells are subjected to the independent variables, cells are lysed, and protein is extracted before sample from all quantitative channels are pooled together. Standard bottom-up digestion followed by LC-MS/MS is then carried-out.

1.6.4. FAIMS and Bottom-Up Proteomics

The integration of FAIMS with electrospray-MS in 1999 by Guevremont and co-workers, illustrated the potential for its incorporation into proteomics experiments[182]. Following this, FAIMS was coupled to peptide analysis to show that smaller peptides exhibit type-A ion mobility behaviour and larger peptides relate more to type-C ion mobility[182,183]. Further, FAIMS was shown to reduce signal to noise for the tryptic digest of porcine heamoglobin[184]. Next, FAIMS was used in a peptide tandem MS experiment to illustrate its ability to reduce background fragment ions which translated to more confident peptide identification[185]. The major development of FAIMS' incorporation with liquid chromatography for the analysis of tryptic peptides from multiple proteins presented opportunities for LC-FAIMS-MS/MS (Figure 1.13). In this work, the authors discussed the benefit of scanning across discrete peptide ions (of a specific range of ion

mobilities) by ‘stepping’ across multiple CVs to ensure analysis of different ions from the same LC eluted ion packet. Since then, Thibault and co-workers applied FAIMS (by stepping across three distinct FAIMS CV values in one LC-MS/MS run) to a nanoLC bottom-up proteomic analysis of human monoblastic U937 lysate, yielding a 6-12 fold improvement in signal to noise, 20 % increase in peptide identification and reduction in peptide signal intensity variation[186]. The removal of background singly charged ions by their neutralisation by FAIMS contributed to these improvements. Further, MacCoss and co-workers stepped across 5 CV values in one LC-MS/MS run to demonstrate a >8-fold improvement in LC peak capacity and >5-fold improvement in dynamic range. Thus, the performance of this workflow was comparable to MudPIT workflows[187]. Since then, Cooper and co-workers have demonstrated the utility of combining multiple LC-FAIMS-MS/MS runs with variation of the FAIMS CV (termed external stepping) as opposed to the previously described stepping across multiple CV values in a single run (termed internal stepping) for the analysis of SUM52 triple negative breast cancer cell lysate. They compared external stepping LC-FAIMS-MS/MS with offline SCX LC-MS/MS to demonstrate the lack of crossover in peptide identifications for the two techniques, thus illustrating their orthogonality[188]. Added to this, LC-FAIMS-MS/MS identified and characterised proportionally more 3^+ precursor peptide ions. In this work, external stepping was shown to vastly improve peptide identification for complex samples, as cycling across CVs for internal stepping was less compatible with LC. The FAIMS device scanned across multiple CVs before returning to the original CV, yet the small window within which a specific peptide would elute from the LC system often did not match its preferential ion mobility dictated CV. Consequently, many peptides would not successfully traverse the FAIMS interface for mass spectrometry analysis[188]. As

FAIMS technology and mass analyser development have improved over the last few years, their integration in bottom-up proteomics has improved performance. Coon and co-workers demonstrated the integration of the FAIMS Pro Interface with a Hybrid Orbitrap mass spectrometer to vastly improve peptide identification. The FAIMS Pro's improved internal compensation voltage stepping (25 ms/transition) together with the increased MS duty cycle of quadrupole isolation, orbitrap high resolution precursor ion analysis and linear ion trap tandem MS analysis identified >8000 proteins in one 6 h single shot experiment, outperforming a prefractionation LC-MS/MS based analysis of the same sample (4 fractions, 1.5 h runtime per fraction)[133].

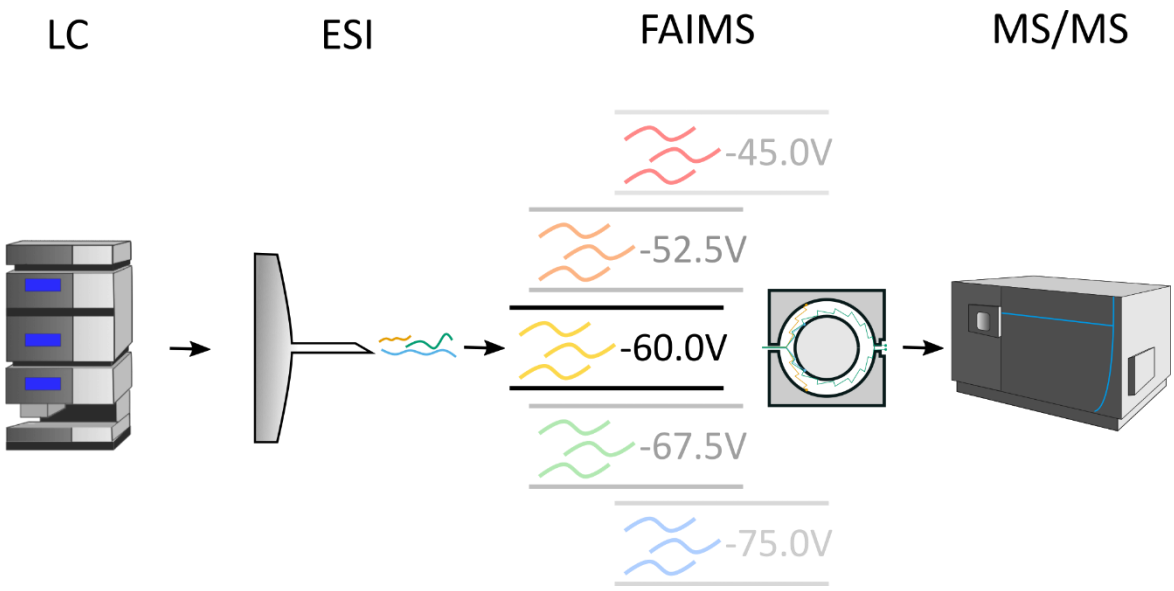


Figure 1.13 | Workflow for LC-FAIMS-MS/MS. After LC separation, peptide ions are electrospray ionised before entering the FAIMS device and onward MS analysis. Different colours of peptides represent their differential ion mobility in the FAIMS device.

The ability of FAIMS to improve signal to noise for lower abundance peptides makes it a viable candidate for improving the identification of PTM containing peptides in bottom-up proteomics workflows, as well as differentiating positional isomers of

modified peptides[181]. Data from such experiments would inform our understanding of PTM crosstalk. Using this rational, Cooper and co-workers applied an external stepping workflow to phosphoenriched SUM52 lysate. Compared to LC-MS/MS, more multiple-phosphosite containing peptides, and more phospho-threonine/phospho-tyrosine sites were identified[188]. FAIMS has also been used to investigate variant methylated histone tails in middle-down proteomics[189], O-linked glycoform variants in mucin glycopeptides[190] and in tandem with ETD fragmentation for the purpose of localisation of PTMs[191].

Thus far, the development of MS instrumentation, together with comprehensive precursor ion fragmentation techniques, sensitive quantitation methodologies, PTM enrichment strategies and increased peptide separation pre-MS analysis using ion mobility have been considered for the improved identification of PTM-containing peptides to inform proteoform variability of proteins via PTM crosstalk. However, the translation of these technological advances and their corresponding information rich spectra is underpinned by the necessity to process such vast amounts of raw data into tangible results, informing our understanding of proteomics (including proteoform variability). As such, significant efforts have been made to improve computational processing software for this purpose.

1.6.5. Proteomics Data Processing

1.6.5.1. Proteomics data processing software for PTM characterisation

The accurate conversion of acquired data from a proteomics experiment into peptide and protein hits, complete with quantitation and PTM site localisation is crucial to the

success of the method. To this end, numerous data processing workflows have been developed, each with their own advantages and disadvantages. As with the choice of proteomics experiment-type, LC, mass analyser(s) and ion mobility device, the data processing software that is selected must be guided by the scientific question. Advances in instrument sensitivity, resolution and duty cycle speed have vastly increased the number of MS/MS spectra acquired from the bottom-up analysis of complex tryptic samples[192]. Thus, computational processing software must keep up to fully exploit these advances. Briefly, two different strategies can be applied for the conversion of MS/MS spectra into peptides: database searching and *de novo* sequencing[193]. For the former, each MS/MS spectrum is compared against theoretical peptides derived from the *in silico* digestion of all proteins within a user inputted search database FASTA file (a list of all gene product proteins of the sample species). A peptide spectrum match (PSM) score is calculated for each MS/MS spectrum against all theoretical spectra corresponding to the *in silico* digested peptides. From this, the highest scoring PSM is then selected as a viable candidate peptide within the sample[135]. Data processing software that use this methodology include Andromeda[194], Mascot[195], SEQUEST[196] and X! Tandem[197]. Alternatively, *de novo* sequencing does not involve a search database of proteins and instead relies on the acquired fragment spectra and fragmentation method properties to elucidate an amino acid sequence[193]. Such search strategies are more compatible for the identification of uncharacterised proteins, or where search databases are not available. Examples of *de novo* sequencing processing software include UniNovo[198], PepNovo[199] and PEAKS[200]. Whilst the choice of processing software affects the output peptide and protein identifications, input parameters such as precursor and fragment ion mass

tolerance, fragment ion types, number of missed cleavages, protease cleavage sites, sequence database and the list of modifications also contribute to the final data output.

1.6.5.2. False Discovery Rate

For the analysis of complex samples that contain thousands of proteins, the searching of hundreds of thousands of MS/MS spectra against all theoretical peptides of a complete proteome requires vast amounts of computation search space. Not only does this increase data processing time and require appropriate computational power, but larger search databases increase the probability of falsely matching spectra to peptides, known as false positive identifications[201]. The introduction of a standardised False Discovery Rate (FDR) as a ratio of falsely discovered PSMs and total number of PSMs somewhat alleviated the issues around false positive identifications[202]. As such, lower FDR correlates to more accurate data and thus greater reliability of results. Two major methods of calculating FDR are used, the target-decoy search strategy and mixture model-based methods[203,204]. Of these, the target-decoy strategy is most practiced. This involves the scrambling of the user inputted search database FASTA file protein sequences to create a decoy database, such that all sequences are factitious. Each MS/MS spectrum is searched against *in-silico* digested peptides of these fabricated proteins just like for the real *in-silico* digested peptides. Matches between MS/MS spectra and scrambled peptides are false positive identifications as the peptide does not theoretically exist in the sample. The number of these false positive identifications against the decoy database divided by the number of peptide identifications from the real search database gives the ratio of probable false positive identifications in the resultant PSM dataset. This dataset is then filtered according to its confidence score by a user defined FDR to

generate a final list of PSM identifications. Typically, an FDR of 0.01 (1 %) is used for proteomics experiments, such that 99% of these PSMs are estimated to be accurate and present in the sample. Generally, for confident protein identification, at least two separate peptides corresponding to the protein must be present with high confidence.

1.6.5.3. Data processing software for PTM analysis

The identification of a dynamic PTM (a PTM that may or may not modify an amino acid site) on a protein adds another layer of complexity to the processing of bottom-up proteomics data. Using the database search strategy, for a given modified trypsin digested peptide to match to an *in silico* digested peptide within the search database, the mass shift caused by the PTM must be considered[135]. This is enabled by the user-defined selection of possible PTMs (and their corresponding amino acid targets) such that these amino acids are considered in their modified and unmodified form at occurrence of that amino acid in a theoretical peptide sequence. Further, when performing the search of a given MS/MS spectrum that corresponds to a modified peptide, the mass shift from the modification(s) is considered for each possible amino acid site of the peptide. Using the *b* ion and *y* ion series that is generated from HCD fragmentation (as described in Section 1.4.3.1), the specific amino acid site(s) of modification can be deciphered (Figure 1.14). The additional searching requirement of variable modifications translates to increased computational search space and therefore longer processing times. Another consequence of this increased search space is the increased number of false positive identifications. Conversely, with around one-third of all unassigned tandem MS spectra (MS/MS spectra from a proteomics run that are not matched to search database peptides) reportedly caused by modifications[205], accurate and increased peptide identification is reliant on the

incorporation of PTMs into the processing software. As such, efforts have been made to solve this problem. For example, so called 'Multipass Search Strategies', whereby one basic search is performed with no modifications before several rounds of iterative searches with a small pool of modifications, can vastly reduce search space[206]. Further, the combination of *de novo* sequencing and database searching methods, termed hybrid search engines, can reduce search space. Examples of hybrid search software include Byonic[207], JUMP[208] and ProteomeGenerator[209]. In the case of Byonic, short peptide sequences (two to four amino acids) are generated by *de novo* sequencing of the MS/MS spectrum, thus creating a list of peptides within the database to search against. This improves the confident identification of multiple localised PTMs on a peptide, which lends well to the investigation of multi-PTM peptides as candidates for positive crosstalk.

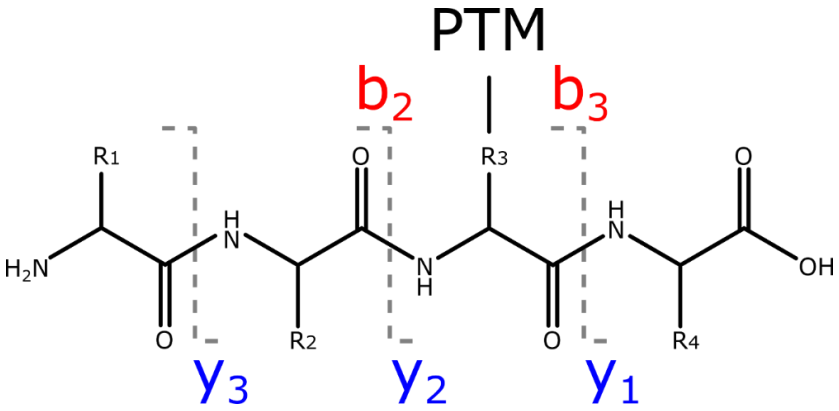


Figure 1.14 | Identification and localisation of a PTM in bottom-up proteomics with HCD fragmentation. For a given peptide, the *b* ion series jump from b_2 to b_3, and *y* ion series jump from y_1 to y_2 provide evidence of the post-translational modification of the R_3 amino acid side chain. Dashed lines represent typical cleavage sites with HCD fragmentation.

Comparative analyses have been performed between different data processing software packages that use this bioinformatic methodology[210,211]. To this end, Drahos and co-

workers investigated the combined searching of up to 10 different PTMs across several thousand proteomics searches using the Byonic hybrid search strategy, and database search strategies MASCOT and MaxQuant[210]. They reported that the inclusion of 'frequent PTMs' (i.e., PTMs that are present with at least 2 % frequency in a sample) improved peptide and thus protein identifications. Further, a linear relationship between number of considered frequent PTMs and number of reliably identified peptides and proteins was evidenced. They also demonstrated comparatively reduced loss of protein and peptide identifications upon including 10 fictitious PTMs, that they knew not to be present in the sample, when using Byonic, compared to MASCOT or MaxQuant, where a significant drop in peptide and protein identifications was observed, as well as an increase in false positive identifications due to the increased search space[210]. Thus, when investigating complex samples where multiple PTMs may be occurring; the hybrid searching strategy of Byonic currently represents one of the most appropriate data processing software available. Despite this, even with hybrid search strategies such as Byonic, the identification of PTM containing peptides is still reliant upon user defined inputting of the PTMs that will be searched against (unless using the Wildcard feature). This prevents the identification of any other PTM, including novel, uncharacterised modifications. In an effort to solve this issue, so called 'open PTM searching' was first developed by Pevzner and co-workers in 2005[212]. Here, PTMs were not inputted into the search parameters and instead, an unrestrictive PTM search algorithm was used to search for all types of PTMs (although this was accompanied by a vast increase in search space). More recently, Paek and co-workers have developed 'MODa', which uses a spectral alignment algorithm that allows for unrestrictive PTM searches with no limitation on the number and type of PTMs per peptide. Nesvizhskii and co-workers have developed

'MSFragger', which uses an open searching strategy together with a fragment ion indexing method to generate a 100-fold improvement in search speed over more conventional database search strategies[213]. This software has since been improved for the incorporation of PTM site localisation[214]. Further, Byonic contains a 'Wildcard' function that can be activated for open-searching[207].

1.6.5.4. PTM False Localisation Rate

Score values to represent the likelihood that a given PSM has been correctly assigned are vital for the robust statistical analysis of proteomics datasets, with FDR based filtering adding confidence that a given dataset is accurate. This has become common practice within the field of bottom-up proteomics, and the publication of such experiments with associated FDR parameters is expected. However, this has not yet been completely translated to the identification and characterisation of PTMs. As a consequence, several papers have expressed concerns over the ability of the proteomics community to accurately localise phosphorylation sites[215–217]. To counter these issues, data processing software packages routinely provide software specific modification site localisation scores that calculate the probability that a PTM at a specific site has been correctly (or incorrectly) assigned, sometimes referred to as False Localisation Rate (FLR)[201]. For example, MaxQuant, Mascot and Sequest include algorithms such as A-score[218], MD-Score[219] and ptmRS[220], respectively, that provide metrics relating to the confidence in a specific PTM at a specific amino acid residue. FLR calculations consider the most likely amino acid sites of modification for a peptide that contains multiple sites that can be modified. As expected, more modifiable sites within a peptide increases complexity in the calculation and leads to a reduced probability of each site being

modified, unless the MS/MS data is highly evidenced towards one specific site. Thus, thresholds can be applied to proteomics datasets to increase confidence in reported PTMs. However, the scoring methodologies for the algorithms vary. For example, estimating *p* values from the fragmentation peaks that are used to identify and localise a PTM, or providing a score difference between the highest scoring PSM and the second highest scoring PSM for a modified peptide. Further, different algorithms differentially interpret the MS/MS signal (such as noise thresholding, fragment selection or neutral loss consideration)[221]. Because of this, different scores from different search software cannot easily be compared[221]. Jensen and co-workers demonstrated this by comparing 22 popular phosphoproteomics pipelines, including four software tools and six search engines, to demonstrate that the same inputted FLR thresholds can yield very different results[221]. In response to these issues, several algorithms have been developed to standardise FLR calculations across search software using decoy amino acids for the estimation of global FLR[201,222,223]. In the context of phosphosite analysis, an amino acid that cannot be phosphorylated (e.g., alanine/leucine) is included as a possible site of phosphorylation. Thus, any reported erroneous modifications to these amino acids are categorised as false positives and used for FLR calculations. However, as with other algorithms for the confident localisation of PTMs, multiply modified peptides represent increased computational complexity and search space, hindering the effectiveness of PTM localisation scoring for such peptides.

Despite the progress of data processing techniques for bottom-up proteomics in identifying site localised and quantifiable PTMs, an important drawback towards its application for the complete proteoform characterisation of a protein, via PTM-crosstalk involves the process of digesting proteins into peptides. This results in the loss of

information regarding the relationship between PTMs that occur on different tryptic peptides within the protein, as only the relationship of PTMs within a peptide can be considered concurrently. To this end, alternative proteomics workflows add value.

1.7. Alternative Proteomics Techniques

1.7.1. Native Mass Spectrometry

Native MS involves the analysis of biological ions, whereby their native structure is maintained by non-denaturing volatile buffers before desolvation in the gas phase with soft ionisation techniques[224]. Current opinion suggests that even though these molecules are transferred from an aqueous solution into the gas phase, protein conformation and non-covalent interactions are retained such that native MS data reflects their native structure in solution[225]. Whilst adduction with salts such as Na^+ can broaden MS peaks and supress signal, the use of nESI[226,227], collisional cleaning by Surface Induced Dissociation (SID)[228] and other techniques can somewhat negate such issues. The advantages of native MS for the investigation of protein PTMs stems from their ability to investigate the combinatorial decoration of PTMs on a protein at the proteoform level. As discussed in Section 1.2, the activity status of a protein should be considered in the context of all the PTMs that are present at any given time (via PTM crosstalk), rather than considering individual PTMs in an isolated manner. Thus, to completely characterise a protein's behaviour, this information is highly relevant. Native MS enables such investigation to be performed. For a purified protein of multiple proteoforms that derive from different PTM-statuses, stoichiometric analysis of the different proteoforms can be derived from the relative abundance of the different proteoform peaks. Further, the structural impact of a PTM on a protein can be probed, which is an important factor when considering how the protein's function is modified by the PTM via the structure-function relationship. For example, Native MS, with SID, has been applied to provide insight into the role of PTMs on the Aquaporin-0 tetrameric membrane protein[229] and for the PTM characterisation of therapeutic monoclonal

antibodies[230]. Recently, native MS has been combined in a biopharmaceutical context with online native liquid chromatography techniques (e.g. Size Exclusion Chromatography, Ion Exchange Chromatography, HILIC) to combine the advantages of native MS with increased throughput[231,232].

1.7.2. Top-Down Mass Spectrometry

Top-down proteomics is defined as the analysis of large polypeptide or protein ions by their initial precursor *m/z*, followed by the energetic activation and identification of product ions relating to the protein. Like native MS, the combinatorial PTM status of a protein can be derived from such information, with multiple proteoforms of a protein identified together with their corresponding stoichiometries[233]. Top-down MS can be associated with on-line liquid chromatography to increase the throughput of the experiment, as proteins within a complex sample are separated before top-down analysis. The identification of multiple PTM-proteoforms for multiple proteins in an individual proteomics experiment is highly advantageous, especially in the context of PTM crosstalk[233]. Added to this, numerous fragmentation techniques, which generate different diagnostic fragment ions, can increase sequence coverage and site localised PTM identification. For example, Top-down MS was used to investigate the combinatorial PTM code of histones using a combination of reverse phase and Hydrophilic Interaction Liquid Chromatography, followed by ECD FT-MS top-down analysis. This led to the site localised characterisation of 42 H4 histone proteoforms, with different combinations of lysine and arginine acetylation and methylation[234]. In another top-down MS experiment, stoichiometric analysis of histone H3.1 lysine-4 methylation and lyine-9 dimethyalation were calculated to occur at 5 % and 50 %, respectively, using

comparative analysis of modified and unmodified ECD tandem MS fragment ion peak ratios[235]. Despite the unrivalled advantages of top-down MS in the analysis of combinatorial PTM regulation of proteins, the huge search space required for processing the MS and tandem MS spectra of such experiments remains challenging. Also, the substantial number of ions that are generated from fragmentation can supress signal intensity. To counteract this, increased data acquisition time is required for each precursor ion. This can compromise the performance of LC-Top-Down-MS experiments, as synchronising the LC system with the increased data acquisition time remains a challenge[233]. However, with rapid improvement in MS configuration and computational efficiency, top-down proteomics will continue to grow as a vital proteomics technique for the analysis of PTMs and their crosstalk. Further, Native MS and Top-Down MS can be combined to form Native-Top-Down MS workflows, whereby for a given protein ion, its native mass spectrum is acquired as well as the spectra of product ions from precursor activation[236].

1.7.3. Middle-Down Mass Spectrometry

Middle-down MS can be considered as a compromise between top-down and bottom-up MS. This technique involves the generation of large polypeptides (~3 – 10 kDa) using specialised protein digestion enzymes. Such enzymes include AspN and GluC that cleave N-terminal to asparagine and C-terminal to lysine, respectively. The natural distribution of these amino acids within proteins and the highly specific nature of the enzymes yield large polypeptides that are then analysed by LC-MS/MS[233]. The large size of these polypeptides is associated with several advantages for combinatorial PTM analysis. Firstly, larger peptides have a statistically greater chance of containing multiple

PTMs, and much like with native MS and top-down MS, the identification of multiple PTMs on the same precursor ion indicates that these PTMs occur concomitantly within that proteoform of the protein. Secondly, increased peptide length generally correlates with increased peptide charge[233]. As discussed in Section 1.4.3.2, ECD/ETD is the optimal fragmentation method for PTM identification by virtue of its sparing of labile PTMs from cleavage. The sequence coverage-based performance of ECD/ETD is charge dependent, with it performing poorly on 2^+ precursor ions. The longer peptides generated by middle-down proteomics (and the proteins of top-down MS) are more suited to ETD than smaller peptides, on account of their increased charge. Ultimately, middle-down MS combined with ETD fragmentation can preserve, identify and localise multiple labile PTMs to build a picture of the various PTM-proteoforms that a protein can occupy. As is the case with top-down MS, larger polypeptides generate more convoluted tandem MS spectra, with more permutations for matching PSMs to a theoretical database resulting in increased search requirements. Thus, the computational demands for data processing still pose a challenge for high-throughput middle-down MS experiments.

1.8. Applications in Biological Contexts

1.8.1. Biological Context 1: Investigating PTM Crosstalk in SUM52 Breast Cancer Cells

The advances in proteomics techniques, as described above, have laid the foundations for a new dawn in PTM identification. Further, an appreciation of the importance of PTM crosstalk, leading to biologically relevant proteoforms associated with disease, has opened novel therapeutic opportunities. Cancer, a broad term to describe a disease caused by uncontrolled division of abnormal cells, has been found to undergo many examples of aberrant PTM crosstalk[17]. For example, the anti-cancer tumour suppressor protein p53 undergoes Tip60 (also known as histone acetyltransferase KAT5) mediated acetylation to prime a subsequent phosphorylation leading to activation[237]. However, reciprocal ubiquitination at this acetylation site competes with Tip60 mediated acetylation to target p53 for proteasomal degradation (as an example of reciprocal PTM crosstalk), reducing the cells ability to evade cancer[238]. Further, it has recently been proposed that cancer testis antigen Morn3 protein can bind p53, recruit the Sirt1 deacetylase for p53 deacetylation and Mdm2 ubiquitin ligase for p53 ubiquitination leading to p53 degradation as a cause of colorectal cancer[239]. In a separate example of PTM crosstalk leading to cancer, SRY-box 2 (SOX2) transcription factor is vital for maintaining the self-renewal of embryonic stem cells, with its aberrant expression linked with cancer. Upon extracellular signal regulated kinase 1 (ERK1) catalysed phosphorylation, SOX2 undergoes SUMOylation, resulting in its autophagic degradation. However, X chromosome-linked inhibitor of apoptosis proteins (XIAP) is reported to inhibit ERK1 kinase activity on SOX2, preventing its phospho-SUMO positive PTM crosstalk regulated degradation and leading to nasopharyngeal

carcinoma[240]. As more experimental analysis of PTM crosstalk in cancer is performed, more examples of proteoform specific regulation (via PTM crosstalk) will be identified. As such, better informed drug design will aid therapeutic action against such cancers. For example, for a positive PTM crosstalking mechanism leading to the cancer phenotype, specific targeting of the first PTM (that stimulates the occurrence of the disease state proteoform) would improve drug efficacy. Upon the development of novel mass spectrometry workflows for the characterisation of multiple PTMs, with elucidation of the relationship between these PTMs; its compatibility to biological contexts must be investigated. To this end, the workflow developed in Chapter 3 was applied to the analysis of PTM crosstalk in SUM52 breast cancer cells (Chapter 4). These cells contain a gene amplification of Fibroblast Growth Factor Receptor (FGFR) 2, stimulating hyperactive tyrosine kinase activity and subsequently generating the cancer phenotype.

1.8.1.1. Fibroblast Growth Factor Signalling

1.8.1.1.1. Fibroblast Growth Factors

Fibroblast Growth Factors (FGF), first described by H A Armelin in 1973[241], represent a family of signalling glycoproteins produced by animal macrophage cells that are involved in such biological processes as development, proliferation, differentiation and angiogenesis[242]. Whilst most of the 22 structurally homologous FGFs are produced in the endoplasmic reticulum before extracellular secretion, some (FGF1 and FGF2) have been reported with nucleolar localisation[243,244]. Structurally, FGF's contain a 120 amino acid sequence that forms 12 conserved β-barrel sheets as well as a signal peptide region for extracellular translocation[245]. FGFs also contain a separate Heparin Sulphate

Proteoglycans (HSPG) binding site that stabilises their interactions with FGFR for continuous downstream activation of FGFR kinase activity[245].

1.8.1.1.2. Fibroblast Growth Factor Receptors

The cognate binding partner of FGF glycoproteins, FGFRs, represent a class of four structurally homologous transmembrane receptor tyrosine kinases. Other receptor growth factor tyrosine kinases include Epidermal Growth Factor Receptor, Platelet Derived Growth Factor Receptor, Human Epidermal Growth Factor Receptor and Vascular Endothelial Growth Factor Receptor[246]. FGFRs are comprised of an extracellular FGF-binding region (including three immunoglobulin (Ig) domains and an acid box region), a single pass transmembrane domain, and an intracellular region (comprised of a juxtamembrane domain, two tyrosine kinase domains and a C-terminal tail)[247] (Figure 1.15). The FGFR family includes five FGFR genes (FGFR1-5), with each FGFR incorporating isoform variability in the Ig domains from differential mRNA splicing. FGFR 1-4 exhibit tyrosine kinase activity via their intracellular tyrosine kinase domain, with FGFR5 lacking the tyrosine kinase domain. Each FGF has a preferred splice isoform variant of a specific FGFR binding partner, dictated by its structure, although some members (FGF11, FGF12, FGF13 and FGF14) do not stimulate FGFR activation. Some FGFs (FGF1, FGF2) can interact with at least one isoform of all kinase active FGFRs, whilst others have more specific FGFR binding affinities[248]. Further, differential FGF and FGFR expression profiles are exhibited by different tissues[248].

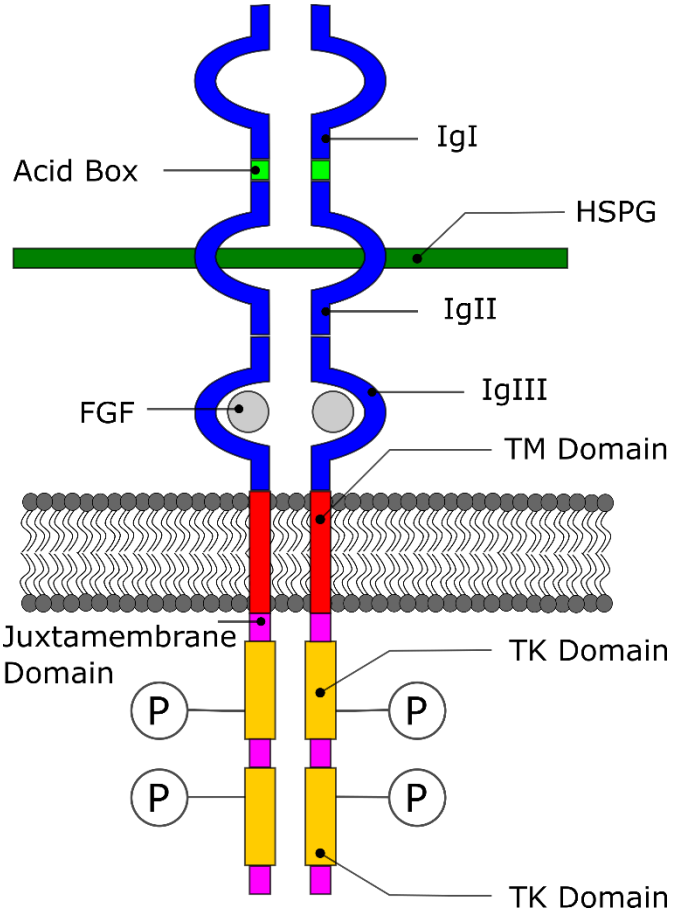


Figure 1.15 | Schematic representation of dimerised Fibroblast Growth Factor Receptor. TM = Transmembrane, FGF = Fibroblast Growth Factor.

1.8.1.1.3. Fibroblast Growth Factor signalling

Upon binding of FGF with its cognate FGFR, structural changes facilitate lateral dimerisation of FGFR in the plasma membrane[249]. The proximity of the two FGFR tyrosine kinase domains upon dimerisation is necessary for their trans-autophosphorylation at the opposing FGFR activation loop region. The now phosphorylated activation loop can stimulate signalling of substrate proteins such as Fibroblast Growth Factor Receptor Substrate 2 (FRS2) via their docking to the autophosphorylation segment at the intracellular juxtamembrane region of FGFR. Phosphorylated FRS2 can then activate proteins further downstream of the signal

cascade[249]. The role of FGF-FGFR activity on cytoplasmic signalling protein phosphorylation status was initially described by Basilico and co-workers upon their observation of elevated tyrosine-phosphorylation levels within 3T3 fibroblasts when stimulated by FGF1/FGF2[250]. The FGFR family control a range of signalling mediated developmental processes including embryonic, foetal and postnatal development[251]. Such processes are specifically driven by activation of the RAS/MAP/MEK/ERK and PI3K/PDK1/AKT intracellular signalling pathways, whereby changes in gene expression, metabolism, survival, cell morphology and cell motility are orchestrated by the terminal targets of these kinase-dependant signalling pathways[177,252–254]. The central role of FGFR signalling in these physiological mechanisms places it as a prominent driver of the cancer phenotype upon aberrant signalling[255]. Firstly, FGF signalling molecules are heavily involved in the regulation of the tumour/stromal microenvironment, including angiogenesis, tumour metastasis and immune evasion[177,256,257]. Added to this, numerous mutations of FGFR can facilitate their hyperphosphorylation, with associated proliferation into tumorous tissue. For example, glioblastoma and haematological malignancies are associated with gene fusions that result in the forced dimerisation of the FGFR kinase domain with such partners as Transforming Acidic Coiled-Coil Gene Family proteins (TACC), leading to potent oncogenic activity[258]. Further, point mutations within the FGFR gene can facilitate forced FGFR dimerisation and ligand independent autophosphorylation, stimulating hypersensitive signalling[259]. FGFR gene amplification represents the most commonly observed cancer lesion causing mutation, including 18 % of breast cancer tumours[260,261]. SUM52, the cancer cell line investigated in Chapter 4, contains an FGFR2 gene amplification and are subsequently FGF addicted, as the prevention of FGF stimulation

results in cell lethality. The heavy involvement of aberrant FGFR signalling in cancer morphogenesis represents a targetable mechanism for cancer therapy. To this end, different strategies for the inhibition of hyperactive FGFR signalling have been developed. Such methods include extracellular 'FGF ligand traps', which bind and sequester peripheral FGF ligands before they can stimulate FGFR dimerisation[262]. However, the most promising strategy involves the synthesis of small molecule inhibitors that competitively bind the ATP-binding cleft of FGFR, inhibiting its tyrosine kinase activity. These molecules can be potent towards multiple receptor tyrosine kinases (e.g., Dovitinib[263], SU5402[177]) or FGFR specific (e.g., AZD4547[264]). Preclinical models of FGFR gene amplifications have demonstrated a high inhibitory effect of small molecule inhibitors on FGFR signalling[265], However, the results of clinical trials have been shown to be dependent on the heterogeneity of hyperactive FGFR stimulated tumour genetics[266]. To complicate things, the acquisition of drug resistance via secondary site gatekeeper mutations[267], and reactivation via other receptor tyrosine kinases (Platelet Derived Growth Factor Receptor-α, Human Epidermal Growth Factor Receptor 2)[268] as well as activating mutations downstream of the receptors within the PI3K/PTEN signalling pathway[269] have compromised the efficacy of these therapeutic strategies. To improve drug efficacy for targeted inhibition of overactive FGFR signalling, a better understanding of the PTM-modulated regulation of downstream signalling pathways from FGFR must be investigated. To this end, Heath and co-workers used sample phosphoenrichment and SILAC quantitative bottom-up proteomics techniques to better characterise the associated phospho-signalling pathways[177]. To supplement these findings, the involvement of other PTMs that may be crosstalking with

the functionally relevant phospho-sites, as well as their independent involvement in generating the cancer phenotype, must also be considered.

1.8.2. Biological Context 2: Investigating S-Glutathionylation of Cyanobacteria Phycobilisome Light Harvesting Complex

1.8.2.1. Cyanobacteria as trailblazers for the biological history of planet earth

Another important consideration of the benefits of improving our understanding of PTM crosstalk and proteoform variability involves evolution within the context of biochemistry. Many PTMs have been reported to have an ancient origin[22], thus our improved understanding of how PTMs regulate proteins would provide a more detailed framework for evolutionary biology. For example, the earliest oxygenic photosynthesisers: cyanobacteria, undergo well characterised modification of their photosynthetic protein machinery. Phycobilin PTMs decorate their light harvesting antennae proteins to absorb and transmit light energy for photosynthesis. To this end, examples of PTM regulatory changes in light harvesting proteins in response to environmental cues have been reported[270,271]. The improved mass spectrometry tool kit for PTM crosstalk (and proteoform characterisation) is well placed to investigate other evolutionarily relevant PTMs. These PTMs, and their crosstalking relationship could be considered as primordial protein regulatory mechanisms that were exploited by evolution towards modern biology.

Cyanobacteria represent the first organisms to perform oxygenic photosynthesis ~2.3 billion years ago[272], using electrons from hydrolysed water to fix carbon and generate biomass. The by-product of this process, molecular oxygen, contributed towards the

Great Oxidation Event, with profound effects on biology[273]. It has been suggested that oxygenic photosynthesis preceded and contributed to the development of aerobic respiration, with the vastly improved efficiency of ATP synthesis fuelling the explosion of eukaryotic multicellular life during the Cambrian period[272]. As such, the importance of cyanobacteria's photosynthetic machinery to the evolutionary history of the planet cannot be overstated. Further, modern analytical techniques may reveal PTM modulation within these protein factories that underpinned these profound changes in our planet's biology. Cyanobacteria remain amongst the world's most efficient light-harvesters, owing to the highly adapted structure of the phycobilisome for channelling excitation energy into their photosystems[274] (Figure 1.16a), as well as the complementary peak absorbance wavelengths of their phycobilisome light harvesting complex and chlorophyll containing thylakoids (Figure 1.16b), but little is known about PTM regulation within these systems.

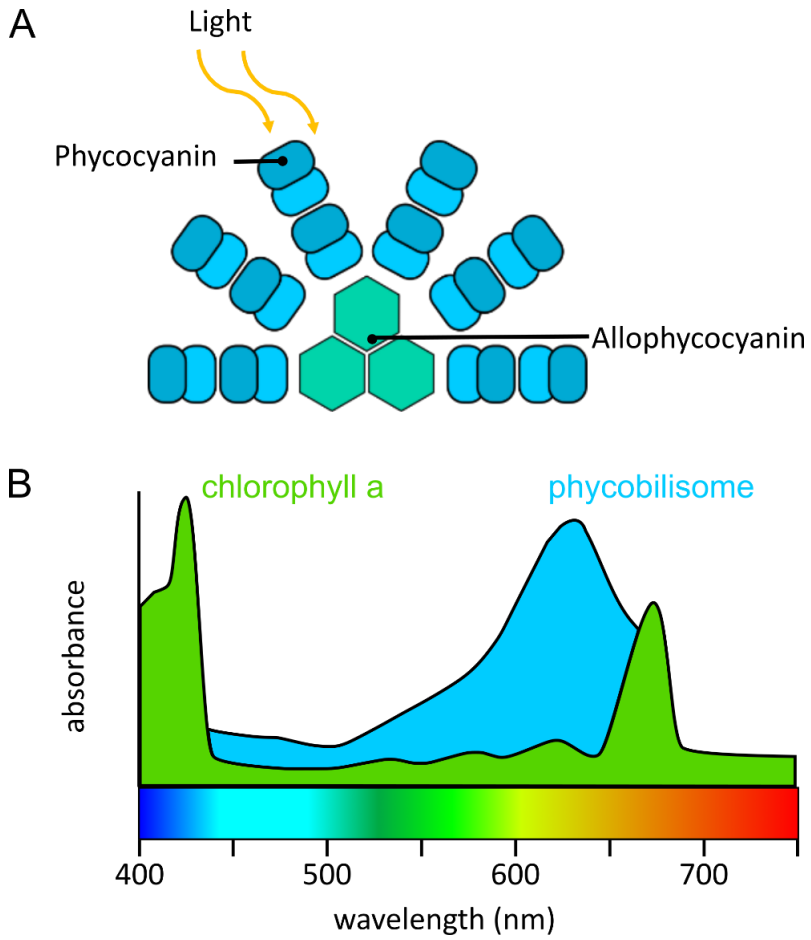


Figure 1.16 | The phycobilisome contributes to the efficient light harvesting capacity of cyanobacteria. (A) The phycobilisome is composed of stacked hexamers of phycocyanin to form dodecamers (phycocyanin, blue), protruding from three core hexamer subunits (allophycocyanin, green). Each allophycocyanin hexamer shape represents hexameric allophycocyanin, Each phycocyanin shape represents a hexamer, with two hexamers forming a dodecamer. (B) Complementarity of chlorophyll a and phycobilisome peak absorbance wavelengths. Chlorophyll a absorbs primarily at ~ 430 – 470 nm and 660 - 680 nm (annotated green) and phycobilisome light harvest protein complex peak absorbance: 620 nm (annotated blue).

1.8.2.2. Cyanobacteria adaptation to reactive oxygen species generated by oxygenic photosynthesis

Briefly, oxygenic photosynthesis involves the stripping of electrons from the hydrogen of water molecules, followed by the ordered flow of these electrons using specialised electron carrier molecules to provide the energy required for carbon fixing[275]. However,

as studied extensively in another electron transport process for the purpose of ATP synthesis, mitochondrial aerobic respiration[276,277], these electrons are capable of generating reactive oxygen species (ROS) such as hydrogen peroxide (H_2O_2), superoxide radical ions ($O_2^{\bullet-}$), hydroxyl radical ions ($HO^{\bullet}$) and more[278]. As such, the generation of ROS, whilst important in regulated concentrations for signalling purposes, may exceed acceptable thresholds and result in damage to DNA, proteins and lipids[279]. Cyanobacteria, with their phycobilisomal light harvesting capacity for photosynthesis, require regulatory mechanisms to prevent overproduction of ROS, especially when light-driven electron transport is not balanced by carbon fixation[280]. Importantly, cyanobacteria can modify their phycobilisome in the context of structural composition in response to environmental conditions[270,271], a key adaptation that has underpinned their ability to survive and thrive. For example, cyanobacteria disassemble their phycobilisome in response to increased ROS from high light intensity to protect these proteins from oxidative damage[281]. In low light environments, phycobilisome rod length has been shown to increase for greater light absorption[282]. Further, macronutrient limitation has been shown to drive phycobilisome disassembly[283], possibly to provide amino acids for the generation of more adapted proteins for the acclimation process[270]. The signalling mechanisms that underpin these changes are yet unknown. Further, PTM regulation may generate as yet unknown proteoform variants to aid such regulatory processes. The glutathione tripeptide (γ-Glu-Cys-Gly, Figure 1.17), is crucial for constraining ROS to manageable levels[284], reducing and consequently soaking up free ROS to manageable levels, that would otherwise damage cellular content. Furthermore, glutathione has been shown to post-translationally modify proteins at their cysteine-sulfhydryl groups, thus protecting these side chains from ROS-facilitated oxidation[285].

Moreover, S-glutathionylation is involved in a host of other signalling contexts including but not limited to: infection[286], inflammation[286] and apoptosis[285]. However, the complexity and lability of S-glutathionylation as a PTM has impeded it's complete global characterisation within evolutionarily relevant organisms such as cyanobacteria[287]. Novel mass spectrometry techniques may unlock the characterisation of specialised PTM-regulatory mechanisms that could inform the above processes.

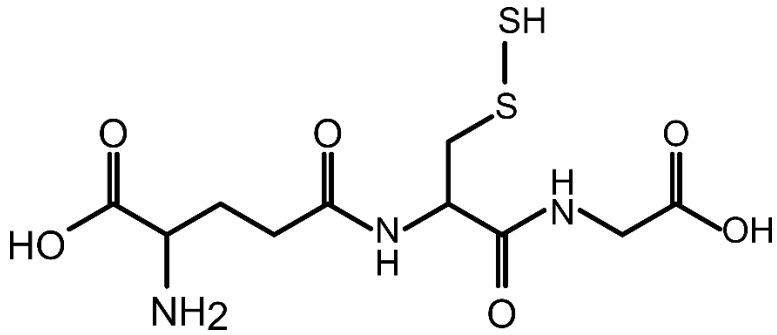


Figure 1.17 | Structure of reduced glutathione. Glutathione is a tripeptide that consists of a glutamate, bonded, via its side chain amide, to a cysteine's backbone amine, which is bonded via its carboxyl group to the backbone amine of a glycine residue (γ-Glu-Cys-Gly).

1.8.2.3. Techniques for the identification of S-glutathionylation PTM

In recent years, the S-glutathionylome has started to be unpicked in both scale and complexity. *In-vitro* radiolabelling methodologies, using ^{3}H-glutathione[288–290] and ^{35}S-glutathione[291–293], coupled with electrophoresis based separation and autoradiography/phosphor imaging techniques[294] pioneered the identification of S-glutathionylated glyceraldehyde-3-phosphate dehydrogenase (GAPDH)[294], actin[295], ubiquitin conjugating enzyme[296], protein kinase C-α[292], pax-8[297], cyclophilin A[298] and thioredoxin[299] amongst other proteins. However, progress using these methods has been hindered by the need to inhibit protein synthesis using cycloheximide (to prevent labelled sulfur from being incorporated into proteins as appose to glutathione) as well as

drawbacks of gel electrophoresis techniques including loading concentration limits and limits of molecular weight[300]. Further, these techniques require cell culture and are dependent upon efficient transfer of cysteine from culture medium- both are issues, particularly for the analysis of photosynthetic organisms such as cyanobacteria[300]. Moreover, the use of anti-glutathione antibodies as a technique that overcomes the above issues is stifled by glutathione's ability to occupy hundreds of conformations resulting in the varied ability of these antibodies to bind glutathione[300]. Mass spectrometry has revolutionised our understanding of post-translational modifications; from the breadth of different PTMs that decorate proteoforms, to the complexity in their signalling contexts via intra- and inter-protein crosstalk. S-glutathionylation enrichment strategies via biotinylated glutathione administration followed by streptavidin-based enrichment coupled to LC-MS/MS have helped overcome the above problems. For example, Chardonnet and co-workers performed a large-scale biotinylated-glutathione LC-MS/MS analysis to identify 383 S-glutathionylated proteins in the model cyanobacterium *Synechocystis* sp. PCC6803[301]. However, biotinylated glutathione requires a bulky biotin tag that likely impedes normal physiological function, and thus results acquired from such techniques may not be representative of the *in-vivo* biochemistry[300]. Thus, it has become increasingly clear that new techniques for the identification of the S-glutathionylation PTM are required. Our understanding of how such PTMs such as S-glutathione modulate protein behaviour via novel proteoform occupancy is still in its infancy. For example, in a photosynthetic context, the exposure of complex protein systems to reactive oxygen species suggests a nuanced role of S-glutathionylation to protect these proteins from ROS-mediated degradation. Further, stoichiometric ratios of different S-glutathionylated proteoforms can inform judgement

on the physiological relevance of these changes. Recently, native, together with top-down and bottom-up mass spectrometry were used to unravel the complex heterogeneity of the cyanobacteria phycobilisome in red microalgae[302], providing structural insight into the complex heterogeneity of the phycobilisome. Hence, native, and top-down mass spectrometry may be useful in analysing assembly dynamics and PTM's of the phycobilisome in response to ROS and other environmental stimuli. Further, more specialised mass spectrometry techniques could allow for multi-modal proteomics to tackle these questions.

1.9. Research Objectives

To truly understand the complexity of biochemistry, protein behaviour must be considered in the context the various proteoforms that a protein can occupy, with PTM-crosstalk mechanisms facilitating access to the different proteoforms. As such, analytical techniques must develop to meet this challenge. Thus, in this thesis, several objectives were targeted.

- Develop a novel LC-FAIMS-MS/MS technique that leverages the FAIMS based ion mobility separation of ions post-LC against mass spectrometry, to improve the identification of low abundance multiple-PTM containing peptides as candidates for positive PTM crosstalk (Chapter 3)
- Investigate the most appropriate data-processing methodologies for the identification of these multi-PTM containing peptides (Chapter 3).
- Apply this LC-FAIMS-MS/MS workflow to the investigation of SUM52 triple negative breast cancer cells via the incorporation of pre-MS phospho-enrichment and SILAC quantitation techniques to improve understanding of PTM regulation within this cancer phenotype, as well as the combinatorial behaviour of multiple PTMs (Chapter 4).
- Apply a multi-model MS approach, including native MS, top-down MS, LC-MS/MS and LC-FAIMS-MS/MS for the investigation of an evolutionarily relevant putative S-glutathionylation PTM within the light harvesting phycobilisome protein complex of cyanobacteria in the context of protection against ROS that are generated by photosynthesis (Chapter 5).

2. Materials and Methods

All figures in this thesis, unless otherwise stated, were created using Inkscape version 1.2.2.

2.1. Methods for Chapter 3: FAIMS Enhances the Identification of PTM crosstalk

Methods for this chapter as described by Adoni, K. R et al[115].

2.1.1. Chemicals and Reagents

All chemicals and materials were purchased from Thermo Fisher Scientific unless otherwise stated. Pierce HeLa Protein Digest Standard, used throughout Chapter 3, was reconstituted in 10% (*v/v*) formic acid at a concentration of 200 ng/µL. Each LC-FAIMS-MS/MS experiment and LC-MS/MS experiment used 1 µg HeLa digest.

2.1.2. LC-FAIMS-MS/MS Instrumentation

2.1.2.1.Online liquid chromatography

The UltiMate™ 3000 RSLCnano liquid chromatography system (Thermo Fisher Scientific), with a 75 µm x 500 mm analytical column (Acclaim PepMap C_{18} 100 Å, 3 µm) and a 0.075 mm x 20 mm trap cartridge (Acclaim PepMap C_{18} 100 Å, 3 µm), was connected to a SilicaTip emitter using a Teflon® union. Column temperature was set at 40 ºC and column flow rate was set to 350 µL/min. Mobile phase A (0.1 % formic acid) and mobile phase B (100 % acetonitrile/0.1 % formic acid) were applied with an elution gradient from 3.2 to 44 % mobile phase B over 155 min. Total run time per sample was 180 minutes.

2.1.2.2. High Field Asymmetric Waveform Ion Mobility Spectrometry

The FAIMS Pro™ was set up in an online LC-FAIMS-MS/MS workflow between the LC and nanoESI system with inner and outer electrode temperatures set to 100 ºC, entrance plate voltage set to 250 V, carrier gas flow rate (N_2) set to 0.00 L/min, dispersion voltage set to -5000 V and the FAIMS electrode ion separation gap at 1.50 mm. Within the external stepping FAIMS condition, separate LC-FAIMS-MS/MS runs at: 0.00, −45.0, −52.5, −60.0, −67.5, −75.0, and −90.0 V were acquired (chosen CVs based on previous work by Hebert *et al.,* 2018[133]). External stepping referred to the pooling of data acquired from all static CV LC-FAIMS-MS/MS runs. In the internal stepping FAIMS condition, the FAIMS device was cycled (3 s) between −45.0, −60.0, −75.0, and −90.0 V, with the duration of each CV in a single cycle set at 0.75 s (chosen CVs based on previous work by Hebert *et al,* 2018[133]). Data acquired from external and internal stepping LC-FAIMS-MS/MS was compared against a standard LC-MS/MS experiment (with the FAIMS Pro™ detached from the workflow). All conditions were run in technical triplicate.

2.1.2.3.Mass spectrometry

The Thermo Fisher Scientific Orbitrap Eclipse Tribrid mass spectrometer was used for this work. The mass spectrometer was externally calibrated using Pierce™ FlexMix™ calibration solution. nanoESI was performed by the application of a voltage to a SilicaTip (MS Wil, FS360-75-15-N) emitter, via a HPLC liquid junction cross (P/N ES269). Spray stability and intensity was optimised by varying the SilicaTip electrospray voltage (1.4 kV – 2.0 kV) using Instrument Control Software version 3.3.2782.34 and varying SilicaTip positioning (in the *x*, *y* and *z* dimensions). Transfer

capillary temperature was set to 300 ºC, RF lens was set to 30 %, precursor ion mass spectrum was acquired at a resolution of 60,000 with a mass range of 380 - 2000 *m/z* and a precursor ion charge state range of 3^+ - 8^+. MS^1 spectra were recorded at 3 s intervals, Automatic Gain Control (AGC) was set to a target of 400,000 (100 %), maximum injection time was set to 50 ms, precursor ions were isolated with a 1.2 *m/z* window using the quadrupole mass filter and Monoisotopic Precursor Selection (MIPS) was enabled. For ion trap mass analysis based MS^2 acquisition (set to Rapid mode), higher energy collision induced dissociation (HCD) was applied with a normalised energy of 35 %, dynamic exclusion was set to 60 s for a single charge state per precursor and precursor filter priority for MS^2 was based on the highest charge state, based on the high frequency of missed cleavages observed in PTM-containing peptides with an associated increase in charge. Ion trap scan range was set to "auto" and maximum injection time was set to "dynamic".

2.1.3. Data Analysis

2.1.3.1.Database searching

A 64 GB Intel(R) Core (TM) i9-10900 CPU @ 2.80 GHz with 12 virtual cores was used with Windows 10 Enterprise N. For data processing, default settings within the Thermo Proteome Discover 2.4 search engine were used unless otherwise stated. All acquired RAW files were searched against *Homo sapiens* (SwissProt TaxID 9606, 20444 proteins) FASTA file, generated September 2019. Dynamic N-terminal methionine loss (−131.040 Da), dynamic N-terminal methionine loss plus acetylation (−89.030 Da), dynamic methionine oxidation (+15.995 Da) and static cysteine carbamidomethylation (+57.021 Da) were included as standard modification in all searches. As well, trypsin

cleavage was set to full, maximum number of modifications was set to 4, precursor mass tolerance was set to 10 ppm and fragment ion mass tolerance was set to 0.5 Da (due to the relatively low ion trap resolution). For FDR calculation, Percolator node was used to filter for PSMs, peptide groups and proteins with a high confidence threshold of 0.01 and a medium confidence threshold of 0.05, and the final filtering set to 0.01 within the final Proteome Discoverer results file. For the initial pairwise PTM search, the database searching strategy of Sequest HT processing node was used with six separate searches pertaining to the different pairwise combinations of S/T/Y phosphorylation, K/R acetylation and K/R methylation. To reduce search space and consequent run time within the broad PTM search (using hybrid search strategy of PMi Byonic[207]); a bespoke FASTA file was generated by including only proteins that were detected in at least two of three replicates of at least one FAIMS condition of internal stepping, external stepping and standard LC-MS/MS, with no PTMs beyond the standard dynamic N-terminal methionine loss (−131.040 Da), dynamic N-terminal methionine loss plus acetylation (−89.030 Da), dynamic methionine oxidation (+15.995 Da) and static cysteine carbamidomethylation (+57.021 Da) included in the search.

2.1.3.1.1. Pairwise PTM searching

For the initial pairwise PTM search, six separate searches were conducted pertaining to the different dual-combinations of phosphorylation (+79.996 Da; S, T, Y), acetylation (+42.011; K, R, N-terminal), and methylation (+14.016; K, R) using the Sequest HT[196] processing node. All these searches included the standard dynamic N-terminal methionine loss (−131.040 Da), dynamic N-terminal methionine loss plus acetylation (−89.030 Da), dynamic methionine oxidation (+15.995 Da) and static cysteine

carbamidomethylation (+57.021 Da). For all acetyl containing searches, N-terminal methionine loss plus acetylation was included as an acetylation PTM within the context of multi-PTM containing peptides. The PTM site probability threshold was set to 75 % using the ptmRS[220] node within the processing workflow of Proteome Discoverer. Processing and consensus pairwise PTM search parameters in Appendix Section 8.1.

2.1.3.1.2. Broad PTM searching

For the broad PTM search, PTMs including: phosphorylation (+79.996 Da; S, T, Y), acetylation (+42.011; K, R), methylation (+14.016; K, R), nitrosylation (+28.990; C, Y), deamidation (+0.984; N), GG corresponding to ubiquitination (+114.042; K), and GlcNAcylation (+203.079; N, S, T), together with the standard PTMs: dynamic N-terminal methionine loss (−131.040 Da), dynamic N-terminal methionine loss plus acetylation (−89.030 Da), dynamic methionine oxidation (+15.995 Da) and static cysteine carbamidomethylation (+57.021 Da) were searched in one run using the hybrid search strategy of PMi Byonic[207] node within the processing workflow. Again, for all acetyl containing searches, N-terminal methionine loss plus acetyl was included as an acetylation PTM within the context of multi-PTM containing peptides. Due to the increased search space of this larger search, potentially facilitating the identification of false positives, the Byonic score threshold (pertaining to the confidence of the PSM match to the theoretical *in silico* digested spectrum) was set to ≥ 300 to minimize false-positive identifications. Again, ptmRS was used for PTM site localisation score analysis, with at least one PTM within the identified multi-PTM peptide corresponding to a ptmRS score of 100 % and another at least scoring 75 %. Further, to ensure that the identifications from this broad search were not entirely false positive detections, 200

scrambled sequences (of amino acid length ranging from 6 – 30) were added to the bespoke FASTA file that the broad PTM search was searched against. No identifications were made against these factitious sequences, adding confidence to the peptide identifications in this search. Processing and consensus broad PTM search parameters in Appendix Section 8.2.

2.1.3.1.3. Open PTM searching

MSFragger[213] (version 3.3) FragPipe (version 16.0) was chosen for open-PTM searching together with Philosopher PTMProphet[303] (version 4.0.0) for its delta mass site localisation capacity within the open-PTM pipeline. The lower precursor mass was set to -150 and higher precursor mass was set to 500. Precursor and fragment mass tolerances were set to 10 ppm and 200 ppm, respectively. All other parameters were kept as default values.

2.1.3.2. Data Analysis

For a multi-PTM peptide to be considered for analysis, its identification in at least two out of the three technical replicates of a specific FAIMS condition (i.e., internal stepping, LC-MS/MS or each individual static CV LC-FAIMS-MS/MS experiment within the external stepping condition) was required. All peptides that contained at least two PTMs of: phosphorylation, acetylation (including methionine loss plus acetylation at N-terminus), methylation, deamidation, nitrosylation, ubiquitination, and glycosylation (GlcNAc) were considered as multi-PTM containing peptides. Multi-PTM peptides that corresponded to the same peptide but differed in sequence length were considered as one

multi-PTM peptide. Multi-PTM peptides that were the same, but with one containing extra 'standard modifications' (including: N-terminal methionine loss (−131.040 Da), N-terminal methionine loss plus acetylation (−89.030 Da), methionine oxidation (+15.995 Da) and cysteine carbamidomethylation (+57.021 Da)) were considered as one multi-PTM peptide.

2.1.3.2.1. Gene Ontology Enrichment Analysis

Proteins selected for gene ontology analysis were grouped and their accession numbers input into g:profiler[304] g:GOSt functional profiling (Ensembl 105, Ensembl Genomes 52 (database built on 2022-01-03)) with organism set to *Homo sapiens* and default advanced options selected.

2.1.3.2.2. Peptide sequence length analysis

For sequence length analysis of non-modified peptides, peptide groups with no modifications that satisfied FDR = 0.01 were grouped according to their amino acid sequence length and plotted against number of identifications for each sequence length. For sequence length analysis of multiple PTM containing peptides (with cysteine carbamidomethylation and methionine oxidation not considered as PTMs), all multiple PTM peptides that satisfied FDR = 0.01 were grouped according to amino acid sequence length and plotted against number of identifications for each sequence length. Graphs were generated in GraphPad Prism 9.

2.2. Methods for Chapter 4: SILAC LC-FAIMS-MS/MS for the Investigation of Novel PTM Crosstalk in SUM52 Breast Cancer

2.2.1. Chemicals and Reagents

All non-cell culture buffers were made using type II ultrapure water. All cell-culture buffers were made using type I ultrapure water. Buffers were pH adjusted using HCl or appropriate hydroxide solution and monitored with a Mettler Toledo benchtop pH Meter. All chemicals and reagents were purchased from Thermo Fischer Scientific unless otherwise stated. All procedures were performed at room temperature unless otherwise stated. All percentage solutions refer to volume of solute in volume of solvent (*v/v)* or mass (g) of solute in volume of solvent (*w/v*) depending on appropriate solution.

2.2.2. Mammalian Cell Culture

2.2.2.1. Media preparation

Roswell Park Memorial Institute (RPMI) 1640 media (purchased from Lonza) was supplemented with 10 % dialysed Foetal Bovine Serum (dFBS) and 1.0 % Penicillin-Streptomycin antibiotic.

2.2.2.2. Cell seeding

Supplemented media was incubated at 37 °C for 15 min. One vial of $N_{2(l)}$ frozen SUM52-PE breast cancer carcinoma cells (Academic)[305] containing approximately $1x10^6$ cells (purchased from BioIVT) was added to 5 mL of RPMI 1640 media that had been incubated using a water bath at 37 °C for 20 min, mixed and centrifuged at 190 *g* for 5 min. The supernatant was removed *in vacuo* (with a sterilised glass pipette attached to a

vacuum pump) and the cell pellet resuspended in 5 mL of fresh media. The cells were then plated into 100 x 20 mm tissue culture treated dishes (purchased from Corning) and supplemented with 5 mL of fresh media to give a total volume of 10 mL. The dish was then incubated at 37 ºC and supplemented with 4 % CO_2. Cell confluence was regularly monitored with microscopy (microscope purchased from Leica DM IL).

2.2.2.3. Cell passage

When cells reached ~90 % confluence (as measured using microscopy), passage was performed as follows: Supplemented RPMI 1640 media and 0.05 % trypsin-ethylenediaminetetraacetic acid (EDTA) were incubated at 37 °C for 15 minutes. Cells were washed with 5 mL of Phosphate Buffered Saline (PBS, purchased from Gibco) and PBS was removed *in vacuo*. 1 mL of trypsin-EDTA was added to the dish before incubation at 37 °C for 5 min. Then, 5 mL of fresh supplemented RPMI 1640 media was added to the dish to resuspend cells and the cells were centrifuged at 190 *g* for 5 minutes. Supernatant was removed *in vacuo,* then cells were resuspended in 5 mL of fresh supplemented media. Based on the original confluence of the dish, the appropriate fraction of the resuspended cells was re-plated and supplemented with fresh RPMI 1640 media to a final volume of 10 mL (generally one fifth of the cell pellet, after 4 days of incubation). The dish was then incubated at 37 °C with 4 % CO_2 supplementation until the next passage.

2.2.2.4. Cell lysis

Throughout the lysis procedure, cells were kept at 4 °C. PBS and lysis buffer were incubated for 4 °C for at least 1 hour. The media was removed *in vacuo* and the cells were washed with PBS, then PBS was removed *in vacuo* (this washing was then repeated once). Lysis buffer (Tris-HCl (20 mM, pH 7.5), Triton X-100 (0.05 % *v/v*), sodium deoxycholate (0.5 % *w/v*), NaCl (150 mM), EDTA (10 mM), protease inhibitor (1 tablet/ 10mL), and phosphatase inhibitor (1 tablet/ 50 mL), 500 µL) was added to the plate and the plate was subjected to orbital shaking (Stuart, see-saw rocker SSL4) at 4 °C for 30 min. Cells were scraped off the plate using a cell-scraper (catalogue number: 10069170, purchased from Corning) and centrifuged at 16,000 *g* for 30 minutes. Supernatant was extracted and protein quantification was performed using a Bradford assay[306] (see Section 2.2.3.1). The cell lysate was stored at -80 ºC until further use.

2.2.2.5. Cell storage

After trypsin digestion, cells were centrifuged at 190 *g* for 5 min and the supernatant was removed *in vacuo*. Dimethyl sulfoxide (DMSO) was added to supplemented RPMI 1640 media (10 % *v/v*) and 1 mL of this solution was used to resuspend cells before pipetting into a cryogenic tube and storage in a liquid nitrogen tank (BOC, EC 231-783-9).

2.2.3. Protein Blotting

SDS-PAGE followed by Western blotting was used to validate amplification and hyperphosphorylation of the SUM52-PE FGFR2 tyrosine kinase receptor relative to other cancer cell lines (MCF-7 and MDA-MB-231).

2.2.3.1.Bradford assay

Cell lysate protein concentration was measured using the Bradford assay[306]. Bovine Serum Albumin (BSA, purchased from Sigma-Aldrich) was dissolved in dH_2O and serially diluted to create the following concentrations: 0.00µg/mL (blank), 25.0 µg/mL, 125 µg/mL, 250 µg/mL, 500 µg/mL, 750 µg/mL, 1000 µg/mL, 1500 µg/mL, and 2000 µg/mL. Three replicates of each BSA concentration were pipetted (10 µL) into a 96 well plate (Perkin Elmer CulturPlate-96 flat bottom white-opaque) as well as three replicates of each SILAC cell lysate sample (10 µL) before gentle resuspension with 300 µL of Pierce™ detergent compatible Bradford reagent. The plate was incubated for 10 minutes before absorbance measurements at 595nm were taken using the Infinite 200 PRO (Tecan Life Sciences). The mean absorbance value for each BSA concentration was used to plot a standard curve and the mean absorbance value for each sample was used to acquired sample concentration using the standard curve.

2.2.3.2.Sodium dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

Tris(hydroxymethyl)aminomethane (Tris) and 30 % w/v acrylamide/methylene bisacrylamide solution (37.5:1 ratio), also known as Protogel, were used for gel synthesis as shown in Table 2.1.

Table 2.1 | Resolving and stacking gel solution constituents and amounts. Resolving volumes correspond to a 12 % gel.

Reagent	**Volume (µL)**	
	Resolving	**Stacking**
H2O	1580	1190
Tris (1.5 M, pH 8.8)	1200	0.0
Tris (0.5 M, pH 6.8)	0.0	500
SDS (10 %)	50	20
Protogel (30 %)	1920	270
Ammonium persulfate (APS**,** 10 %)	45	20
Tetramethylethylenediamine (TEMED)	5.0	5.0

12 % resolving gel solution was made up to 4.8 mL with APS and TEMED added last (Table 2.1), before solution was pipetted between a glass plate (purchased from Bio-Rad, 1.0 mm spacer) and spacer plate (purchased from Bio-Rad, 1.0 mm spacer). Pipetted solution was topped with 70 % ethanol to level the top of the resolving gel. After gel was polymerised (~20 minutes), stacking gel was made up as shown in Table 2.1, with APS and TEMED added last before pipetting above polymerised resolving gel. Then, a 10-well comb was added to the gel before stacking gel polymerisation for around 20 minutes. The well comb was removed prior to loading of protein sample onto the gel.

20 µL of 20 µg protein lysate was denatured by addition of 4 µL of x6 concentration Laemmli buffer and incubated at 96 °C for 2 minutes before pipetting of the samples into the wells. 1.50 µL of molecular weight ladder (PageRuler^{TM} Prestained Protein Ladder) was added to the first well of the gel. After the gel was loaded, the frame was placed into an electrophoresis cell (Mini-Protean Tetra System purchased from Bio-rad) with the cell connected to a power pack at 140 V (purchased from Bio-rad).

2.2.3.3.Western blotting

After SDS-PAGE was run to completion, protein bands were transferred onto an Immobilon®-FL polyvinylidene difluoride (PVDF) membrane (purchased from Merck) using the Mini-Protean Tetra System (purchased from Bio-rad) and connected to a powerpack at 100 V. Transfer was performed for 1 h at 4 ℃ using Transfer Buffer (1.92 mM Glycine, 24.9 mM Tris-Base, 10 % methanol). Membrane was then incubated with 5 % BSA in TBS-T (50 mM Tris-HCl, 150 mM NaCl, 0.1 % Tween-20, pH 7.5) for 1 hour before washing with TBS-T (three times, 5 min). Membrane was then incubated overnight in primary antibody (1:1000 dilution) in TBS-T with 3 % BSA. TBS-T washing was again performed three times for five minutes before incubation with secondary antibody (1:10,000 dilution) in TBS-T with 3 % BSA. Following high salt TBS-T washing (TBS-T with 5.0 M NaCl) five times for 5 minutes, membrane was washed for five minutes in TBS to remove Tween before imaging with Odyssey Imager, LI-COR. Primary antibodies: anti-FGFR 2 Receptor (D4L2V) (purchased from Cell Signalling Technology), anti-phospho-FGFR Receptor (Tyr653/654) #3471 (purchased from Cell Signalling Technology) and anti-α-tubulin (EP1332Y) (purchased from abcam) and secondary antibody: Goat anti-rabbit H&L (IRDye® 800CW, ab216773) (purchased from abcam) were used for Western blotting experiments.

2.2.4. Cell Viability Assays

2.2.4.1. Stock and active drug concentration preparation

Lyophilised SU4502 (purchased from Cambridge BioScience) was dissolved in 100 % DMSO and stored at -20 °C. Lyophilised AZD4547 (purchased from Stratech Scientific Ltd) was dissolved in 100 % DMSO and stored at -80 °C. The stock solutions of both

drugs were diluted into RPMI 1640 and Dulbecco's Modified Eagle Medium (DMEM) media for administration to cell lines: SUM52-PE and MDA-MB-231, respectively. SU5402 was diluted in the range: 0.1 µM – 32 µM at concentrations: 0.1 µM, 0.2 µM, 0.3 µM, 0.6 µM, 0.8 µM, 2.5 µM, 10 µM, 20 µM, and 32 µM. AZD4547 was diluted in the range 1.0 nM - 11 nM at concentrations: 1.0 nM, 2.0 nM, 3.0 nM, 4.0 nM, 5.0 nM, 6.0 nM, 7.0 nM, 8.0 nM, 9.0 nM, 10 nM, and 11 nM. These active concentrations were prepared by serial dilution. For both drugs, titration concentrations were based on previous work by Baker, K., *et al.*[307].

2.2.4.2. Drug titration assays

Titration assays for SU5402 and AZD4547 were performed against the SUM52-PE cell line to optimise drug concentration in SILAC experiments for 50 % growth inhibition (GI_{50}). The same titrations of the drug were performed against the negative control cell line: MDA-MB-231 to ensure cell death from both drugs were stimulated by receptor tyrosine kinase inhibition. Each drug-cell line experiment was performed in a separate 96-well plate (Perkin Elmer, CulturPlate-96, white-opaque, sterile and tissue culture treated), with each drug concentration administered to cells in triplicate together with a vehicle control (0.32 % DMSO) and positive control (fresh media with no drug or DMSO). The vehicle control of DMSO at 0.32 % was chosen as this was the highest DMSO concentration for any of the drug titrations (32 µM SU5402). Initially, 100 µL containing ~$7x10^3$ cells of SUM52-PE or MDA-MB-231 in RPMI 1640 or DMEM, respectively, were seeded in each well, and incubated at 37 °C with 4 % CO_2 supplementation for 24 hours before media was removed *in vacuo* and replaced with 100 µL of drug containing media as described in Section 2.2.4.1. Cells were then incubated

at 37 °C with 4 % CO_2 supplementation for 72 hours before cell viability analysis was performed using the CellTiter-Glo® Luminescent Cell Viability Assay kit (purchased from Promega). Each 96-well plate and its contents were incubated at room temperature for 30 minutes before 100 μL CellTiter-Glo® Reagent was added to each well and mixed for two minutes using an orbital shaker. Each plate was then incubated at room temperature for 10 minutes to stabilise luminescence signal before luminescence was measured (integration time: 0.25 s per well) using a plate reader (Infinite 200 PRO, Tecan Life Sciences).

2.2.5. SILAC Quantitative Proteomics Investigation of PTM Crosstalk in SUM52 Breast Cancer Cells

2.2.5.1. SILAC cell growth

Stable isotope labelled arginine and lysine were used for preparation of SILAC media (Table 2.2).

Table 2.2 | Materials used for SILAC incorporation into cells.

Material	Supplier	Abbreviation
L-Lysine-2HCl, 4, 4, 5,5-D4 for SILAC	Cambridge Isotope Laboratories	K4
L-Arginine-HCl U-13C6 for SILAC	Cambridge Isotope Laboratories	R6
L-Lysine-2HCl U13C6, 15N2 for SILAC	Cambridge Isotope Laboratories	K8
L-Arginine-HCl U13C6, 15N4 for SILAC	Cambridge Isotope Laboratories	R10
L-Lysine-2HCl	Thermo Fisher Scientific	K0
L-Arginine-HCl	Thermo Fisher Scientific	R0
L-Proline	Thermo Fisher Scientific	P0
RPMI 1640 Media for SILAC	Thermo Fisher Scientific	-

2.2.5.2. Preparation of SILAC media

RPMI 1640 for SILAC media (deficient in L-lysine and L-arginine) was supplemented with dFBS, and antibiotics as in Section 2.2.2.1. For each amino acid (L-Lysine/L-Arginine, see Table 2.2) to be added to the media, the amino acid (50 mg) was dissolved in 1 mL of RPMI 1640 for SILAC and then added to the media bottle to generate final lysine and arginine concentrations of 0.1 mg/mL. As well, 250 mg of unlabelled proline was dissolved in 1 mL RPMI 1640 media for SILAC and added to the media bottle to give a final additional proline concentration of 0.5 mg/mL for all SILAC labelled media. Light, medium, and heavy media were prepared as reported in Table 2.3.

Table 2.3 | Incorporation of L-lysine and L-arginine to generate SILAC media.

Label	Lysine	Arginine
Light	K0	R0
Medium	K4	R6
Heavy	K8	R10

2.2.5.3.Experiment design

2.2.5.3.1. Cell passage into SILAC media

Cells were grown from $N_{2(l)}$ frozen SUM52 samples in standard RPMI 1640 media as described in Section 2.2.2 for four passages (with 20 x 100 mm tissue culture treated dishes) before being moved into light, medium and heavy media, to generate three quantitative channels for SILAC proteomics investigation. After four cell doublings in SILAC media, samples were moved into 20 mm x 150 mm cell culture treated dishes (purchased from Corning), for at least two more cell doublings for increased protein lysate extraction upon cell lysis.

2.2.5.3.2. AZD4547 and SU5402 preparation

Lyophilised AZD4547 (10 mg, purchased from Stratech Scientific Ltd) was dissolved in 100 % DMSO before serial dilution in medium labelled RPMI media for SILAC for administration into medium labelled SUM2 cells. Lyophilised SU4502 (purchased from Cambridge BioScience) was dissolved in 100 % DMSO before serial dilution into heavy labelled RPMI media for SILAC for administration into heavy labelled SUM52 cells.

2.2.5.3.3. Cell treatment and lysis

All biological conditions were performed in triplicate. After 4 h serum starvation at 37 ºC, light cells were treated with 0.2 % DMSO (vehicle control), medium cells were administered with 8 nM AZD4547, and heavy cells were administered with 20 μM SU5402 before incubation for 30 minutes at 37 ºC. All three labelled cell lines were then treated with 10 μg/mL heparin and 20 ng/mL FGF1 before incubation for 30 minutes at 37 ºC. Following this, cell lysis was performed on all SILAC samples as described in Section 2.2.2.4 to generate cell lysates corresponding to the three conditions: vehicle control SUM52, AZD4547 administered SUM52 and SU5402 administered SUM52. Bradford assay was performed on each biological condition to acquire total protein concentration (as described in Section 2.2.3.1) before labelled lysates were pooled together with equal concentrations. Thus, three SILAC labelled biological replicates were acquired for LC-FAIMS-MS/MS analysis.

2.2.5.4.Trypsin digestion

For each SILAC SUM52 protein lysate biological replicate, 0.50 M ammonium bicarbonate was added to a final concentration of 50 mM before sample was reduced with 8mM dithiothreitol (DTT) and incubated at 37 °C for 45 minutes. Reduced sample was then alkylated using 20 mM iodoacetamide and incubated in the absence of light for 45 minutes before a 1:5 dilution with 50 mM ammonium bicarbonate. 2 mg/mL of trypsin (purchased from Promega) was added at a trypsin: protein ratio of 1:50 and incubated overnight at 37 °C (protein lysate content was determined by Bradford assay). The digestion was then arrested by addition of trifluoroacetic acid (TFA) to a final concentration of 0.5 %.

2.2.5.5. Sample desalting

2.2.5.5.1. Sep-Pak® C_{18} desalting

Initial desalting of digested lysate (after trypsin digestion) was performed using a Sep-Pak® C_{18} Plus Light cartridge (purchased from Waters). Initially, the cartridge was washed with 4 mL of 100 % acetonitrile before 1.5 mL of conditioning buffer (50 % acetonitrile, 0.5 % acetic acid) was added. Cartridge was then equilibrated with 4 mL of 0.1 % TFA before digested sample was loaded onto the Sep-Pak® column. Sample washing was performed with 4 mL of 0.1 % TFA before a further wash with 1 mL of 0.5 % acetic acid. Desalted peptide was eluted with 2 mL of elution buffer (50 % acetonitrile, 0.5 % acetic acid).

2.2.5.5.2. ZipTip C_{18} desalting

Final desalting pre-LC-FAIMS-MS/MS analysis was performed using a 5 µg capacity ZipTip C_{18} column (purchased from Millipore®). The ZipTip was attached to a P20 pipette and conditioned with 10 µL of 100 % acetonitrile by aspiration and dispensing twice. Next, equilibration was performed by aspiration and dispensing of 10 µL of 0.1 % TFA, twice. Sample peptides were dried *in vacuo* and resuspended in 10 µL of 0.1 % TFA before loading onto the ZipTip column by aspiration and dispensing (back into sample tube) ten times. Peptides were then washed with 10 µL 0.1 % TFA by aspiration and dispensing twice before peptide elution by aspiration of 10 µL 0.1 % formic acid/50 % acetonitrile and dispensing into a fresh sample tube. This was repeated ten times to ensure sufficient elution of sample peptides from ZipTip column. Samples were then dried *in vacuo* before reconstitution in 0.1 % formic acid for proteomics analysis. For 80 µg capacity ZipTip (Pierce C-18 Tips, 100 µL bed) desalting; the same protocol was followed with 100 µL volume instead of 10 µL.

2.2.5.6. TiO_2 phosphoenrichment

Sample was phospho-enriched using ~100 µg capacity TiO_2 spin tips (Titanosphere Phos-TiO_2 purchased from Hichrom). Initially, 20 µL of Buffer A (80 % acetonitrile, 0.4 % TFA) was used to condition spin tip with centrifugation at 3800 *g* for 2 minutes before 20 µL of Buffer B (25 % lactic acid, 60 % acetonitrile, 0.3 % TFA) was added to equilibrate tip with centrifugation at 3800 *g* for 2 minutes. Conditioning step was repeated before 1 mg of sample (reconstituted in 50 µL buffer B) was added to TiO_2 tip for adsorption followed by centrifugation at 1350 *g* for 10 minutes with flow-through

re-added to tip and centrifugation repeated. Then, 20 μL of Buffer B was added with centrifugation at 3800 *g* for 2 minutes, followed by the same centrifugation with 20 μL of Buffer A twice. Elution of phosphopeptide was performed by addition of 50 μL of 5 % ammonia solution with centrifugation at 1350 *g* for 5 minutes, followed by addition of 50 μL of 5 % pyrrolidine solution with centrifugation at 1350 *g* for 5 minutes. Solvent was removed *in vacuo* and peptides were ZipTip desalted before reconstituted in 0.1 % formic acid for bottom-up proteomics analysis.

2.2.5.7. LC-FAIMS-MS/MS instrumentation

Two non-phosphoenriched biological replicates were analysed by external stepping LC-FAIMS-MS/MS and three TiO_2 phosphoenriched biological replicates were analysed by internal stepping LC-FAIMS-MS/MS.

2.2.5.7.1. Liquid chromatography

The UltiMate™ 3000 RSLCnano liquid chromatography system (Thermo Fisher Scientific), with a 75 μm x 500 mm analytical column (Acclaim PepMap C_{18} 100 Å, 3 μm) and a 0.075 mm x 20 mm trap cartridge (Acclaim PepMap C_{18} 100 Å, 3 μm), was connected to a SilicaTip emitter using a Teflon® union. Column temperature was set at 40 ºC and column flowrate was set to 350 μL/min. Mobile phase A (0.1 % formic acid) and mobile phase B (100 % acetonitrile/0.1 % formic acid) were applied with an elution gradient from 3.2 to 44 % mobile phase B over 24 min. Total run time was 56 minutes.

2.2.5.7.2. FAIMS

The FAIMS Pro™ was set up in an online LC-FAIMS-MS/MS workflow between the LC and nanoESI system with inner and outer electrode temperatures set to 100 ºC, entrance plate voltage set to 250 V, carrier gas flow rate (N_2) set to 3.50 L/min, dispersion voltage set to -5000 V and the FAIMS electrode ion separation gap at 1.50 mm. For static CV conditions (for external stepping) applied to non-phosphoenriched SUM52 SILAC sample, the selected CV: -40.0 V, -45.0 V, -50.0 V, -55.0 V and -60.0 V was applied for each LC-FAIMS-MS/MS run. Each static CV was considered as a fraction within the external stepping condition. For phosphoenriched samples, internal stepping was performed, with the FAIMS device set to cycle between -45.0 V, -52.5V, -60.0 V and -67.5 V, each for 0.75 s with a total cycle time of 3 s.

2.2.5.7.3. Mass spectrometry

The Thermo Fisher Scientific Orbitrap Eclipse Tribrid mass spectrometer was used for this work. The mass spectrometer was externally calibrated using Pierce™ FlexMix™ calibration solution. nanoESI was performed by application of a voltage to a SilicaTip (FS360-75-15-N) emitter, via a HPLC liquid junction cross (P/N ES269). Spray stability and intensity were optimised by varying the SilicaTip electrospray voltage (1.4 kV – 2.0 kV) using Instrument Control Software version 3.3.2782.34 and varying SilicaTip positioning (in the *x*, *y* and *z* dimensions). Transfer capillary temperature was set to 275 ºC, RF lens was set to 30 %, precursor ion mass spectrum was acquired at a resolution of 120,000 with a mass range of 380 - 2000 *m/z*, Automatic Gain Control (AGC) was set to a target of 400,000 (100 %), maximum injection time was set to 50 ms, Monoisotopic Precursor Selection (MIPS) was enabled and precursor ions were isolated with a 1.2 *m/z*

window using the quadrupole mass filter. Dynamic exclusion was set to 15 s for a single charge state per precursor and charge states from 2^+ to 6^+ were selected for MS/MS. Data Dependant Analysis was set to Top10. MS^2 acquisition was performed in the Orbitrap with resolution set to 15,000, scan range set to 200-2000 *m/z* and AGC target set to 50,000 (100%). Higher energy collision induced dissociation (HCD) was applied with a normalised energy of 28 %.

2.2.5.8.Data processing and analysis

2.2.5.8.1. Data processing

Proteome Discoverer search engine was used with the hybrid search strategy of PMi Byonic node. RAW files were searched against the *Homo sapiens* (SwissProt TaxID 9606, 20444 proteins) FASTA file, generated September 2019 with default Proteome Discoverer and Byonic settings used unless otherwise stated. For all searches, trypsin cleavage was set to full and maximum number of common modifications per peptide was set to 4, precursor and fragment mass tolerance were set to 10 ppm and 0.05 Da, respectively. PTMs included in the search were: phosphorylation (+79.996 Da; S, T, Y), acetylation (+42.011; K, R), methylation (+14.016; K, R), nitrosylation (+28.990; C, Y), deamidation (+0.984; N), GG corresponding to ubiquitination (+114.042; K), and GlcNAcylation (+203.079; N, S, T). Dynamic N-terminal methionine loss (−131.040 Da), dynamic N-terminal methionine loss plus acetylation (−89.030 Da), dynamic methionine oxidation (+15.995 Da) and static cysteine carbamidomethylation (+57.021 Da) were included as standard modification in all searches. For all acetyl containing searches, N-terminal methionine loss plus acetylation was included as an acetylation PTM within the context of multi-PTM containing peptides. For PTM site localisation,

ptmRS[220] was used. An FDR of 0.01 for PSMs, peptide groups, and proteins was applied to the final Proteome Discoverer results file. SILAC 3plex (Arg6, Lys4 | Arg10, Lys8) quantification method was selected in Proteome Discoverer, Within the processing workflow, Minora Feature Detector node was selected for peak and feature detection with minimum trace length set to 5, signal to noise threshold set to 1, maximum ΔRT of isotope pattern multiplets set to 0.2 min and feature to ID linking PSM confidence set to 'at least high'. Within the consensus workflow, feature mapper and precursor ion quantifier nodes were selected with maximum chromatogram RT alignment set to 10 min with coarse parameter tuning. All other SILAC quantitation parameters were set as default. Processing and consensus search parameters in Appendix Section 8.3.

2.2.5.8.2. Data analysis

Quantification ratios for AZD4547 (medium quantitative channel) and SU5402 (heavy quantitative channel) conditions were compared against control condition (light quantitative channel), with all abundance ratios greater than 1.5 or less than 0.67 considered as significant abundance changes (vs control). All statistically significant multi-PTM peptides groups were present in at least two biological replicates (out of three biological replicates and out of two biological replicates for phospho-enriched internal stepping experiment and non-enriched external stepping experiment, respectively). All PSM MS/MS spectra from which actin cytoskeletal related proteins were discussed in Section 4.2.6 were manually validated. Proteins selected for gene ontology analysis were grouped and their accession numbers input into g:profiler[304] g:GOSt functional profiling (Ensembl 106, Ensembl Genomes 53 (database built on 2022-05-18)) with organism set to *Homo sapiens* and default advanced options selected. For network map generation,

grouped protein accession numbers were input into STRING[308] v 11.5 with ‘multiple proteins’ selected and organism set to *Homo sapiens*. Data was then exported for visualisation and figure generation in Cytoscape[309] version 3.9.1.

2.3. Methods for Chapter 5: Mass Spectrometry Based Investigation of a Novel PTM Regulator of Cyanobacteria Light Harvesting Protein Complex

2.3.1. Chemicals and Reagents

All buffers were made using type II ultrapure water unless otherwise stated. Buffers were pH adjusted using HCl or appropriate hydroxide solution and monitored with a Mettler Toledo benchtop pH Meter. All chemicals and reagents were purchased from Thermo Fischer Scientific unless otherwise stated. All procedures were performed at room temperature unless otherwise stated. All percentage solutions refer to volume of solute in volume of solvent (*v/v)* or mass (g) of solute in volume of solvent (*w/v*) depending on appropriate solution.

2.3.2. Cyanobacteria Cell Culture

Arthrospira maxima (reference 1475/9) and *Arthrospira platensis* were provided by the Culture Collection of Algae and Protozoa (CCAP)[310] biological resource centre, before being cultured in a 50:50 ratio of artificial seawater (ASW) and blue-green liquid media (BG) with a 12h: 12h light/dark cycle. The culture was grown in a 250 mL conical flask, enclosed in a cardboard box with a cavity through which an LED light was applied to the culture during the light half of the day *as described in J. Bellamy-Carter et al., 2022*[311]. Media preparation (described below) was based on CCAP recommendations, as described in CCAP knowledgebase[310].

2.3.2.1. Artificial seawater

2.3.2.1.1. Stock solution preparation for artificial seawater

Constitution for preparation of stock solutions for ASW (Table 2.4)

Table 2.4 | ASW stock solution constituents. Per 1 Litre, dissolved in medium as described in **Table 2.5**.

Stock solution 1 (salts). Per 1 L	$NaNO_3$	30.00 g
	Na_2HPO_4	1.20 g
	K_2HPO_4	1.00 g
Stock solution 2 (vitamins). Per 1 L	Biotin	0.0002 g
	Calcium pantothenate	0.02 g
	Cyanocobalamin	0.004 g
	Folic acid	0.0004 g
	Inositol	1.00 g
	Nicotinic acid	0.02 g
	Thiamine HCl	0.1 g
	Thymine	0.6 g

After preparation, stock solutions were autoclaved at 15 psi.

2.3.2.1.2. Preparation of soil extract

Soil was extracted from grid reference: SP 31827 94287 (Hartshill Haze Country Park), before large debris such as stone and invertebrates were removed using a 1 cm mesh sieve. Sieved soil was then spread for air drying with handpicked removal of smaller invertebrates and roots before periodical re-spreading. This cycle of spreading and hand-picking was repeated twice before soil was sieved through a 2 - 4 mm mesh and stored in an airtight container in a cool and light-free location. Twice the soils volume of distilled water was then added, and the contents autoclaved at 15 psi for two hours before being left to cool at room temperature. Supernatant was then aliquoted before another

round of autoclaving (15 min for 1 L aliquots). The resulting soil extract solution was refrigerated until required. Soil extract, stock solutions and tricine were then combined (Table 2.5).

2.3.2.1.3. Final volumes for preparation of artificial seawater

Constitution of soil extract, stock solutions and tricine for ASW preparation.

Table 2.5 | Medium constituents for artificial seawater preparation. Per 1 L.

Tricine	0.50 g
Stock solution 1	3.75 mL
Stock solution 2	2.50 mL
Soil Extract	25.00 mL

2.3.2.2. Blue Green medium (BG11)

Constitution of stock salts and trace metals for BG11 preparation (Table 2.6)

Table 2.6 | BG11 constitution for salt and trace metal stocks. Per 1 L.

Stock Salts (per 1L)	$NaNO_3$	150.0 g
	K_2PO_4	4.0 g
	$MgSO_4.7H_2O$	7.50 g
	$CaCl_2.H_2O$	3.60 g
	Citric acid	0.60 g
	Ammonium ferric citrate green	0.60 g
	$EDTANa_2$	0.10 g
	Na_2CO_3	2.00 g
Stock Trace Metals (per 1L)	H_3BO_3	2.86 g
	$MnCl_2.4H_2O$	1.81 g
	$ZnSO_4.7H_2O$	0.22 g
	$Na_2MoO_4.2H_2O$	0.39 g
	$CuSO_4.5H_2O$	0.08 g
	$Co(NO_3)_2.6H_2O$	0.05 g

Constitution of stock salts and stock trace metals for preparation of BG11 (.

Table **2.7**).

Table 2.7 | Constitution of BG11 media from salt and trace metal stocks as described in Table 2.6. Per 1 L

Stock Salts	10.0 mL
Stock trace metals	1.0 mL

BG11 media was made up to 1 L with deionised water, before pH adjustment to 7.1 (using 1 M NaOH/1 M HCl). Media was then autoclaved for 15 minutes at 15 psi.

2.3.3. Cyanobacteria Culture Incubation

2.3.3.1. High light intensity experiments

Arthrospira platensis was cultured as described in Section 2.3.2 before 10 mL was harvested in to a 25 mL conical flask. The flask was then incubated in a Micro-Pharos PBR^{TM} tank (purchased from Xanthella) with tank parameters set as follows:

Table 2.8 | Configuration of Micro-Pharos PBR^{TM} tanks for high light intensity experiments.

Condition	Configuration parameter
Light Intensity (mA)	150 (control: 30)
Light Cycle (day: night)	12h: 12h
Temperature (ºC)	20

All other tank parameters were set to default values. Cultures were incubated for 3 h, before lysis and protein extraction.

2.3.3.2. Hydrogen peroxide experiments

Arthrospira platensis was cultured as described in Section 2.3.2 before 10 mL was harvested in to a 25 mL conical flask. For the hydrogen peroxide incubation conditions, 100 μM H_2O_2 was added to 10 mL culture. The flask was then incubated in a Micro-Pharos PBR™ tank with tank parameters reported in Table 2.9:

Table 2.9 | Configuration of Micro-Pharos PBR™ tanks for hydrogen peroxide experiments.

Condition	Configuration parameter
Light Intensity (mA)	30
Light Cycle (day: night)	12h: 12h
Temperature (ºC)	20

All other tank parameters were set to default values. Cultures were incubated for 3 h, before lysis and protein extraction. Control condition followed the same protocol without addition of hydrogen peroxide.

2.3.4. Cyanobacteria Protein Extraction

2.3.4.1. Cyanobacteria sample lysis

Following incubation of cyanobacteria, 10 mL of sample was centrifuged at 16,000 *g* for 30 min before supernatant was removed. An equivalent volume of distilled water was added to cell pellet, followed by resuspension to mix pellet. Sample was then freeze-thaw cycled (-80 ºC for 20 mins to 20 ºC until completely thawed) five times to lyse the cells. Following this, sample was centrifuged at 16,000 *g* for 10 minutes, and supernatant collected for further analysis. The degree of lysis could be inferred from the colour of the supernatant, with successful lysis indicated by blue colour of supernatant. If

supernatant was clear, further rounds of freeze-thaw were performed until blue coloured supernatant appeared.

2.3.4.2. Cyanobacteria buffer exchange into 100 mM ammonium acetate for native mass spectrometry

Buffer exchange was performed at 4 ºC and refrigerated ammonium acetate was used. Initially, 30 kDa MWCO (Molecular Weight Cut-Off filter, purchased from Merck Millipore) was added to a 2 mL Eppendorf tube and washed with 500 µL dH_2O by centrifugation at 16,000 *g* for 5 minutes. MWCO was then equilibrated by centrifugation with 500 µL of 100 mM ammonium acetate at 16,000 *g* for 5 minutes. Next, supernatant from lysed sample was added to MWCO, and topped up to 500 µL with 100 mM ammonium acetate and resuspended before centrifugation at 16,000 *g* for 5 minutes. Then, flowthrough was discarded and 500 µL of 100 mM ammonium acetate was added to MWCO followed by centrifugation at 16,000 *g* for 5 minutes, with this step repeated a further 8 times. Sample was then pipetted out of MWCO and centrifuged at 16,000 *g* for 2 minutes, before supernatant was transferred to a fresh Eppendorf tube to remove any remaining cell debris. Sample was then stored at 4 ºC in dark (refrigerated) until further analysis was performed.

2.3.4.3. Absorbance based measurement of C-phycocyanin and allophycocyanin concentrations

Upon extraction of buffer exchanged C-phycocyanin and allophycocyanin (in 100 mM ammonium acetate), absorbance readings were taken using a DS-11 spectrophotometer

(purchased from DeNovix, Wilmington, NC, USA). For allophycocyanin, an absorbance peak maximum of 652 nm, with a molar extinction coefficient of 7×10^5 $M^{-1}cm^{-1}$ was used to extrapolate concentration. For C-phycocyanin, absorbance maxima at 280 nm, 620 nm and 652 nm were used to extrapolate concentration. Concentrations were calculated using the equation defined by Bennet & Bogorad[312].

2.3.5. Mass Spectrometry Methodology and Data Processing

All experiments were performed in triplicate unless otherwise stated.

2.3.5.1. Native mass spectrometry and native top-down mass spectrometry

2.3.5.1.1. Instrument set-up

All native mass spectrometry was performed using an Orbitrap Q-Exactive HF mass spectrometer. Instrumentation parameters such as: trapping gas, *m/z* range, mass resolution and quadrupole isolation were modified using customised instrument control software with permission from Thermo Fisher Scientific. nanoESI was performed using borosilicate glass capillaries that were pulled in-house using a P-1000 pipette puller (Sutter Instrument) and gold-coated using a sputter coater (purchased from Agar Scientific Ltd). nanoESI was performed in positive mode with voltage ranging from 1.4 kV – 1.8 kV and gas pressure set to 0.3 psi. For all native mass spectrometry experiments, source temperature was set to 250 ºC, in-source dissociation (SID) was set to 20, S-lens RF was set to 100, with trapping gas pressure at 5. All native MS were acquired at a resolution of 15,000, and a mass range of 1000 – 6000 *m/z* was used. Automatic Gain Control (AGC) and maximum ion injection time were set to 1×10^6 and

50 ms respectively. Samples were sprayed at a C-phycocyanin concentration of 5 μM (based on Bennet and Bogorad calculation, see Section 2.3.4.3) for native mass spectrometry analysis. For native top-down analysis of C-phycocyanin, the 11^+ 3434 *m/z* peak was isolated using the quadrupole before HCD was applied with a normalised HCD energy of 15 % for MS^2 acquisition of dissociated α- and β-phycocyanin. Methods based off *Sound, J. K. et al., 2021*[313].

2.3.5.1.2. Data processing and analysis

All mass spectra were visualised using Xcalibur 4.1 (Thermo Fisher Scientific), and manually deconvoluted. UniDec[314] version 4.2.2 was used to guide peak annotation for more complex spectra. For the predicted masses of allophycocyanin and C-phycocyanin monomer, dimer and hexamer; amino acid sequences for α-, αB-, and β-chains of allophycocyanin and α- and β-chains of phycocyanin were determined by metagenomics studies undertaken by CCAP and published findings of phycobilisome subunit protein masses by *Sound et al*.[313]. Calculated masses were adjusted for post-translation modifications that had been previously reported in the UniProt[315] database, including the phycobilin chromophore (+585.7 Da), removal of initiator methionine (-131.2 Da) and N4-methylAsp-72 substitution for Asp-72 (+14 Da).

2.3.5.2. LC-MS/MS for the characterisation of C-phycocyanin S-glutathionylation

C-phycocyanin concentration was determined by Bennet & Bogorad based absorbance analysis (2.3.4.3). Trypsin digestion was performed as described in Section 2.2.5.4, with the exception that iodoacetamide and DTT were not used (to prevent S-glutathione PTM

removal). Sep-Pak® and ZipTip desalting were performed as described in Section 2.2.5.5 before removal of solvent *in vacuo* and peptide reconstitution with 0.1 % formic acid. Liquid chromatography was performed as described in Section 2.2.5.7.1, with the exception that an Acclaim PepMap 100 C_{18} analytical column (75 µm x 150 mm) was used and nanoESI was performed using a Triversa Nanosource Spray (purchased from Advion) as appose to a SilicaTip. Mass spectrometry was performed using an Orbitrap Q-Exactive HF mass spectrometer. Transfer capillary temperature was set at 275 ºC, RF lens was set at 60 %, Orbitrap precursor ion mass spectra were acquired at 120,000, mass range was set to 360 – 1600, maximum injection time was set to 40 ms, charge states ranging from 2^+ to 8^+ were selected for MS^2 analysis, precursor ions were isolated with a 1.2 *m/z* window using the quadrupole mass filter, MS^2 acquisition was performed in the Orbitrap with a resolution of 15,000 and HCD was applied with a normalised energy of 28 %.

2.3.5.3. C-phycocyanin sequence alignment analysis

NIH BLAST blastp suite[316] was used to compare *Arthrospira platensis* C-phycocyanin, with amino acid sequence query range: 100 – 117, against subject orthologous C-phycocyanin of: *Synechocystis sp*., *Nostoc sp*, *Agmenellum quadruplicatum*, *Synechococcus sp*., *Mastigocladus laminosus*, *Cyanidum caldarium*, *Aglaothamnion neglectum*, *Porphyra purpurea*, *Neoporphyra haitanensis*, *Neopyropia yezoensis*, *Cyanophora paradoxa*, *Thermosynechococcus vestitus*, *Microchaete displosiphon*, *Galdieria sulphuraria*, *Rhodella violacea* and *Aphanizomenon flos-aquae*.

2.3.5.4. LC-FAIMS-MS/MS analysis of cyanobacteria for investigation of differential S-glutathionylation PTM

Sample preparation for LC-FAIMS-MS/MS was described in Section 2.3.5.2, with the exception that protein lysate concentration was acquired by Bradford assay (as described in Section 2.2.3.1).

2.3.5.4.1. Online liquid chromatography

The UltiMate™ 3000 RSLCnano liquid chromatography system (Thermo Scientific), with a 75 µm x 500 mm analytical column (Acclaim PepMap C_{18} 100 Å, 3 µm) and a 0.075 mm x 20 mm trap cartridge (Acclaim PepMap C_{18} 100 Å, 3 µm), was connected to a SilicaTip emitter using a Teflon® union. Column temperature was set at 40 ºC and column flowrate was set to 350 µL/min. Mobile phase A (0.1 % formic acid) and mobile phase B (100 % acetonitrile/0.1 % formic acid) were applied with an elution gradient from 3.2 to 44 % mobile phase B over 24 min. Total run time was 56 minutes.

2.3.5.4.2. FAIMS

The FAIMS Pro™ was set up in an online LC-FAIMS-MS/MS workflow between the LC and nanoESI system with inner and outer electrode temperatures set to 100 ºC, entrance plate voltage set to 250 V, carrier gas flow rate (N_2) set to 3.50 L/min, dispersion voltage set to -5000 V and the FAIMS electrode ion separation gap at 1.50 mm. LC-FAIMS-MS/MS was operated in external stepping mode, with individual 56 minute LC-FAIMS-MS/MS runs collected as fractions corresponding to FAIMS CV's:

-40.0 V, -45.0 V, -50.0 V, -55.0 V and -60.0 V for high light intensity experiments and -45.0 V, -52.5 V and -60.0 V for hydrogen peroxide experiments.

2.3.5.4.3. Mass spectrometry

The Thermo Fisher Scientific Orbitrap Eclipse Tribrid mass spectrometer was used for this work. The mass spectrometer was externally calibrated using Pierce™ FlexMix™ calibration solution. nanoESI was performed by application of a voltage to a SilicaTip (FS360-75-15-N) emitter, via a HPLC liquid junction cross (P/N ES269). Spray stability and intensity were optimised by varying the SilicaTip electrospray voltage (1.4 kV – 2.0 kV) using Instrument Control Software version 3.3.2782.34 and varying SilicaTip positioning (in the *x*, *y* and *z* dimensions). Transfer capillary temperature was set to 275 ºC, RF lens was set to 30 %, Orbitrap precursor ion mass spectra were acquired at 60,000 with a mass range of 350 - 2000 *m/z*, Automatic Gain Control (AGC) was set to a target of 400,000 (100 %), maximum injection time was set to 50 ms, Monoisotopic Precursor Selection (MIPS) was enabled and precursor ions were isolated with a 1.6 *m/z* window using the quadrupole mass filter. MS^2 acquisition was performed in the Orbitrap with resolution set to 30,000 and AGC target set to 50,000 (100%). Higher energy collision induced dissociation (HCD) was applied with a normalised energy of 30 %, dynamic exclusion was set to 60 s for a single charge state per precursor and charge states from 2^+ to 8^+ were selected for MS/MS. Data Dependant Analysis was set to Top10.

2.3.5.4.4. Data processing & analysis

2.3.5.4.4.1. Data processing

A 64 GB Intel(R) Core (TM) i9-10900 CPU @ 2.80 GHz with 12 virtual cores was used with Windows 10 Enterprise N for all data processing. All RAW files were processed and analysed using Thermo Proteome Discoverer version 2.5. All database searches were performed using Sequest HT[317]. The default settings were predominantly used in Proteome Discoverer unless otherwise stated. *Arthrospira platensis* (strain NIES-39 / UTEX 3086 / IAM M-135, Proteome ID: UP000006803), taken from UniProt[315] was used for the protein database. For all searches, trypsin cleavage was set to full, with maximum number of missed cleavages set to 3. Precursor mass tolerance was set to 10 ppm, with fragment mass tolerance set to 0.05 Da. Target FDR for proteins, peptide groups and PSMs was set to 0.01. Data was searched against PTMs: phosphorylation (79.996 Da; S, T, Y, H), acetylation (+42.011; K, R, N-terminal), methylation (+14.016; K, R), glutathionylation (+305.068 Da, C), phycocyanobilin (+586.279 Da, C), oxidation (+15.995 Da, M), N-terminal methionine loss (-131.040 Da, M) and N-terminal methionine loss plus acetylation (-89.030 Da, M). Total common modifications parameter was set to 4.

2.3.5.4.4.2. Data analysis

For LC-FAIMS-MS/MS high light intensity experiments, samples were prepared in biological triplicate. For an S-glutathionylated peptide isoform to be considered; at least one PSM corresponding to this peptide was present in 2/3 biological replicates. For any protein that was derived from these validated S-glutathionylated peptide isoforms; at least two separate peptide groups had to be present. Proteins derived from S-

glutathionylated peptides, upon exposure of cyanobacteria to high light intensity, were then grouped and analysed for enrichments in gene ontology using STRING[308] v.11.5. Protein interaction network figures were generated using Cytoscape v3.9.1[309]. For LC-FAIMS-MS/MS hydrogen peroxide experiments, the same analysis as above were performed; with the exception that samples were prepared in biological duplicate. For an S-glutathionylated peptide isoform to be considered; at least one PSM corresponding to this peptide was present in 2/2 biological replicates.

2.3.5.4.5. Determining optimal FAIMS external stepping CVs for identification of S-glutathionylation PTM

After LC-FAIMS-MS/MS was performed on high light intensity experimental samples, all S-glutathionylated PSMs were assigned to the compensation voltage at which the PSM was identified to determine the most favourable FAIMS CVs for S-glutathionylation analysis.

3. FAIMS Enhances the Identification of PTM Crosstalk

The following work has been published in the Journal of Proteome Research and was written by K. R. Adoni (first author) and Dr A. C. Leney[115]*. Permission to use figures and data was granted by the Journal of Proteome Research. All LC-FAIMS-MS/MS data corresponding to this chapter is available on Proteome Xchange (Identifier: PXD026448).*

3.1. Overview

Proteoform characterisation is one of the great challenges of modern biochemistry, and development of novel methodologies for the detection and analysis of proteoform variability is required to properly understand protein biochemistry. Further, the mechanistic formation of multiple PTM containing proteoforms, via a technique known as PTM crosstalk, must be investigated. Whilst PTM-crosstalk has been probed with upregulation of one PTM followed by Western blotting of a potential complimentary PTM[318] or the analysis of synthetic peptide libraries for PTM-crosstalk sequence motif investigations[53–55]; LC-MS/MS has become the gold standard for PTM crosstalk characterisation[318], owing to its accurate, high-throughput detection of amino acid site localised PTMs. LC-FAIMS-MS/MS, as a bottom-up proteomics workflow that incorporates High Field Asymmetric Waveform Ion Mobility Spectrometry (FAIMS) after LC separation and before MS and tandem MS *m/z* analysis, has shown promise in the context of improved proteome converage[133], increased multi-phosphorylated peptide identification[319,320] and the separation of peptide isoforms that differ in the amino acid site of the PTM[321]. In this chapter, an unbiased, enrichment-free LC-FAIMS-MS/MS methodology was developed for the detection and characterisation of multiple PTM containing peptides from the tryptic digestion of HeLa S3 cervical cancer cell lysate, as candidates for positive PTM-crosstalk. Comparison of this methodology to a standard LC-MS/MS workflow revealed a 6-fold improvement in the detection of such peptides, with ~40 % of these PTM crosstalk sites not previously reported.

3.2. Results and Discussion

3.2.1. LC-FAIMS-MS/MS Enhanced the Identification of Peptides and Proteins

Tryptic peptides from the protein lysate of human cervix cancer cell line, HeLa S3, were used as a quality control standard for LC-FAIMS-MS/MS from which the performance of the methodology could be monitored, with its expression of over 15,000 proteins with relevant PTMs providing the ideal platform for our purposes[322]. Due to the varying properties of tryptic peptides, LC-FAIMS-MS/MS replicates were performed using a range of FAIMS static compensation voltages (0, −45.0, −52.5, −60.0, −67.5, −75.0, and −90.0 V) and the sum of the different peptides identified at each compensation voltage was compared to the number of peptides identified when the FAIMS source was removed from the instrument (standard LC-MS/MS). A workflow for this methodology is illustrated (Figure 3.1). FAIMS compensation voltage (CV) choices were informed by previous work from Coon and co-workers[133]. They had analysed tryptic peptides derived from the K562 human cell line at CVs ranging from -10.0 V to -120.0 V, identifying the total number of peptides and unique peptides (that are not identified at other CVs) for each compensation voltage in one-hour LC-FAIMS-MS/MS runs. From this, they concluded that the best four CVs for maximised peptide identification were: -45.0 V, -60.0 V, -75.0 V and -90.0 V. Peptides identified from -10.0 V to -45.0 V were almost entirely singly charged, whilst the more informationally relevant multiply charged peptides tended to preferential transmit below -45.0 V[133]. Ultimately, improved peptide identification would yield an increase in the number of multiple PTM containing peptides that were identified, from which positive PTM crosstalk could be investigated. Notably, Coon et al. reported a 3.7-fold increase in the number of unique peptides identified per protein when FAIMS was added to the bottom-up proteomics workflow[133].

As such, improved peptide identification, and thus protein identification was expected when comparing LC-FAIMS-MS/MS with standard LC-MS/MS on trypsin digested protein lysate of HeLa S3 cervical cancer cells.

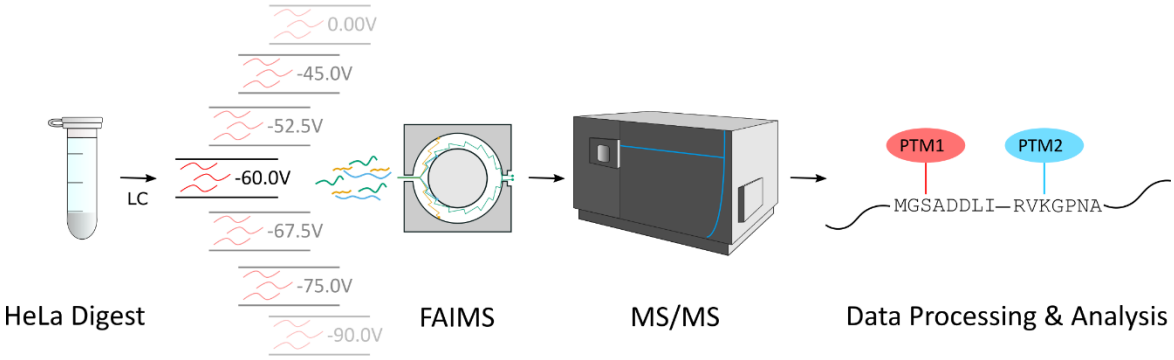


Figure 3.1 | Proteomics workflow for LC-FAIMS-MS/MS. *Figure taken with permission from Journal of Proteome Research (Adoni, K. R., et al)*[115].

The addition of FAIMS to the LC-MS/MS workflow improved peptide and protein identifications. Importantly, both internal stepping (whereby the FAIMS device cycles across CVs: -45.0 V, -60.0 V, -75.0 V and -90.0 V every 3 seconds in one LC-FAIMS-MS/MS run) and external stepping (whereby the FAIMS device is static at each individual CV for the entire LC-FAIMS-MS/MS run, with all data from each LC-FAIMS-MS/MS run pooled together) yielded an improvement in the total number of peptide identifications. These data validated the LC-FAIMS-MS/MS workflow, demonstrating that FAIMS enabled more precursor peptide ions to be identified before fragmentation and MS/MS analysis (Figure 3.2). Further, peptide identifications were transmitted in a normally distributed manner across CVs, centred around -60.0 V, except for -90.0 V, which detected minimal identifications. Despite the low number of peptide identifications at -90.0 V (1339 peptides), this translated to a disproportionately high

number of protein identifications (1763 proteins), indicating the importance of these -90.0 V peptide identifications in deriving proteins within the sample. Further, the same peptides often matched multiple isoforms of a protein, leading to a higher number of protein identifications than peptide identifications.

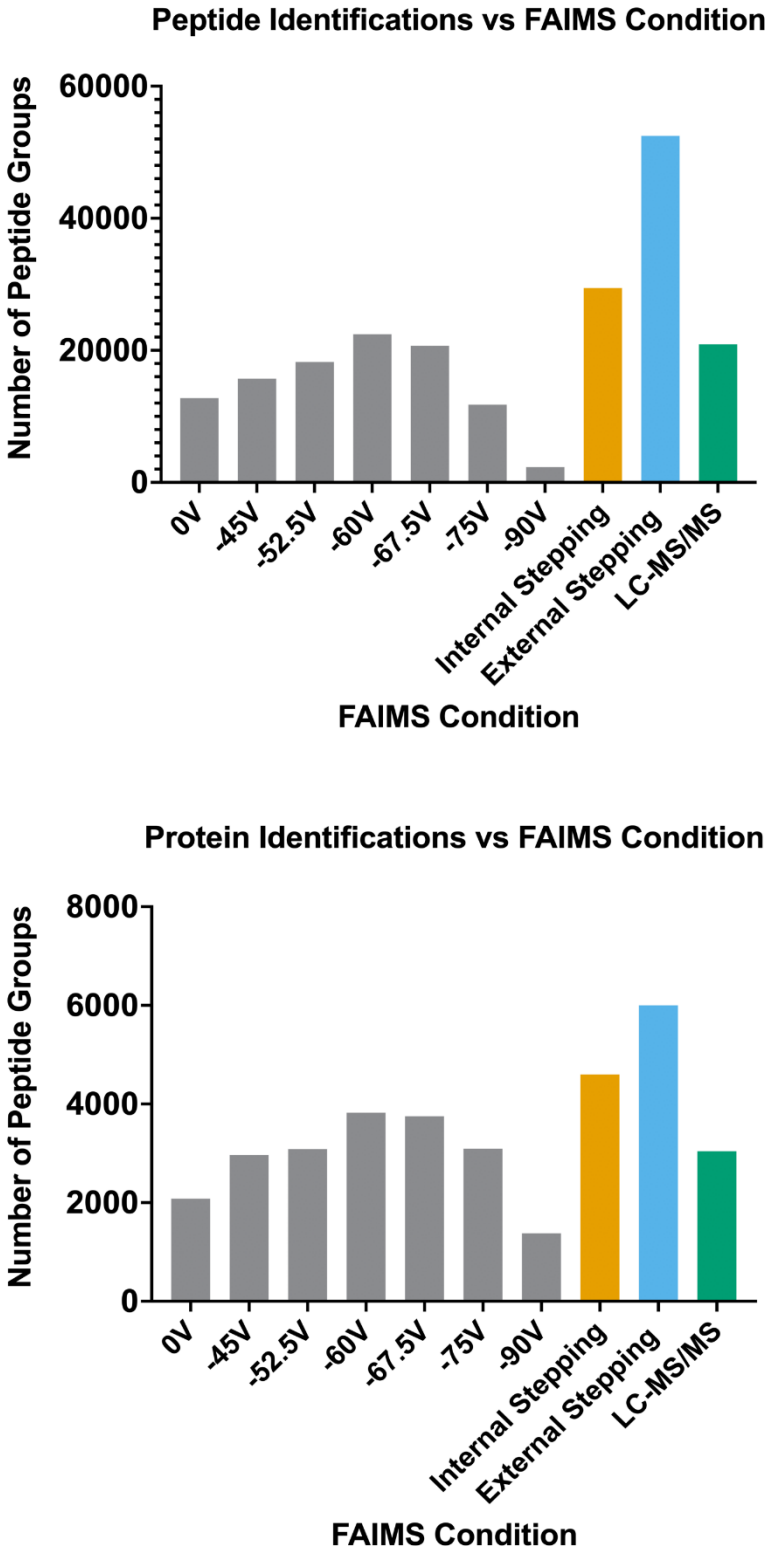


Figure 3.2 | LC-FAIMS-MS/MS improved peptide and protein identifications relative to standard LC-MS/MS. (A) Number of peptide groups identified at different static compensation voltages as well as for internal stepping, external stepping, and standard LC-MS/MS. (B) Number of proteins identified for same conditions as (A). *Figure taken with permission from Journal of Proteome Research (Adoni, K. R., et al)*[115].

3.2.2. LC-FAIMS-MS/MS Enhanced the Identification of Multi-PTM Peptides as Candidates for Positive PTM Crosstalk

3.2.2.1. LC-FAIMS-MS/MS enhanced the identification of multi-PTM peptides involving phosphorylation, acetylation, and methylation

With the LC-FAIMS-MS/MS workflow corroborating previous literature relating to improved proteome coverage upon the incorporation of FAIMS[133], the relationship between the different FAIMS CV's and the identification of multiple PTM containing peptides was next investigated. Multi-PTM containing peptides could be considered as candidates for positive PTM crosstalk, based on their co-occurrence on the peptide. To identify these candidate PTM-crosstalk peptides, the theoretical database was expanded to include any two or more of the following PTMs: S/T/Y phosphorylation, R/K/N-terminal acetylation, and R/K methylation. These three PTMs were chosen because of their biological relevance as well as the existence of databases with curated modification sites[141,142,210,323]. Previous work by Shvartsburg and co-workers had characterised a shift in collisional cross-section of peptides upon post-translational modification, leading to modified peptide ion mobilities[324]. As was predicted based on the differential charge, shape and collisional cross section leading to differential ion mobility, LC-FAIMS-MS/MS improved the identification of these multi-PTM containing peptides, as candidates for positive PTM-crosstalk. Static CV LC-FAIMS-MS/MS runs between -45.0 V and -67.5 V increased the number of multiple PTM containing peptide identifications, compared with standard LC-MS/MS (Figure 3.3). Overall, 179 multi-PTM containing peptides, corresponding to 158 proteins, were reported using this external stepping LC-FAIMS-MS/MS workflow. This represented a 4-fold improvement in identifications compared to standard LC-MS/MS (Figure 3.3). Also of

note, -90.0 V did not generate any validated multi-PTM containing peptides as candidates for positive crosstalk across all PTM combinations, in corroboration with the relatively poor performance of -90.0 V in overall peptide identifications.

PTM containing peptides often occur at low stoichiometry relative to their non-modified isoforms[66]. Thus, searching for multi-PTM containing peptides as candidates for positive crosstalk amplifies this problem. The challenge of detection of these peptides is related to the high background of un-modified and mono-modified peptides that co-elute into the mass spectrometer and are preferentially selected for analysis. LC-FAIMS-MS/MS was shown to enhance the detection of peptides and therefore the proteins that they derived from (Figure 3.2), and this translated to the improved detection of multi-PTM containing peptides as candidates for positive crosstalk (Figure 3.3). This data supported the theory that the ion mobility dimension of the workflow specifically enabled low abundance peptides to transmit through the FAIMS device whilst higher abundance peptide ions collided with the FAIMS electrode and were neutralised. These low abundance peptide ions were therefore more likely to be identified as precursor ions. Subsequently, higher energy collision induced dissociation (HCD) was performed for MS/MS mass analysis and information relating to the characterisation of the peptide, together with site localised modifications was acquired. Furthermore, multi-PTM containing peptides as candidates for positive crosstalk were transmitted through the FAIMS device from -0.00 V through to -75.0 V. Data acquisition above 0.00 V was not performed, however even without an upper limit, the broad range of compensation voltages through which these ion were transmitted provided an opportunity to filter across the peptidome of a digested sample without the intrusion of higher abundance background peptide ions that compete for precursor identification in the Orbitrap after

elusion from the liquid chromatography analytical column. As demonstrated in Figure 3.2, despite the poor peptide identification of -90.0 V (Figure 3.2a), this corresponded to a relatively high protein identification for -90.0 V external stepping (Figure 3.2b), suggested that these peptides were uniquely identified at -90.0 V, leading to the identification of proteins that would otherwise be missed. Hence, the -90.0V FAIMS CV may be useful for the identification of fruitful peptides for protein detection, that are otherwise difficult to characterise. For the purposes of positive PTM-crosstalk, however, -90.0 V should be omitted.

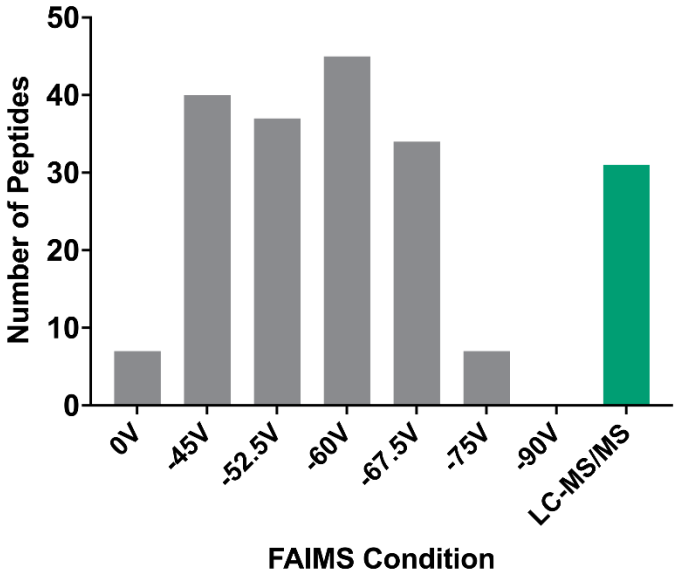


Figure 3.3 | FAIMS improved the identification of multi-PTM peptides as crosstalk candidates. Multi-PTM containing peptides as candidates for positive crosstalk at each FAIMS CV was compared to standard LC-MS/MS. -45.0 V, -52.5 V, -60.0 V and -67.5 V all identified more multi-PTM peptides than standard LC-MS/MS. *Figure taken with permission from Journal of Proteome Research (Adoni, K. R., et al)*[115].

3.2.2.2. LC-FAIMS-MS/MS can be tailored for specific PTMs when searching for multi-PTM containing peptides

Multi-PTM containing peptides, as candidates for positive crosstalk, were preferentially transmitted across static CVs ranging from 0.00 V to -75 V (Figure 3.3). This posed the question of whether bespoke FAIMS CVs could be applied to specifically transmit multi-PTM containing peptides to enrich for the PTM-combination that was being investigated. The relationship between FAIMS CV and multi-PTM combinations for multi-PTM containing peptides as candidates for positive crosstalk was therefore investigated. The results demonstrated a clear relationship between PTM combination and FAIMS CV (Figure 3.4), with Table 3.1 reporting the best four CVs when performing external stepping FAIMS for different PTM combinations.

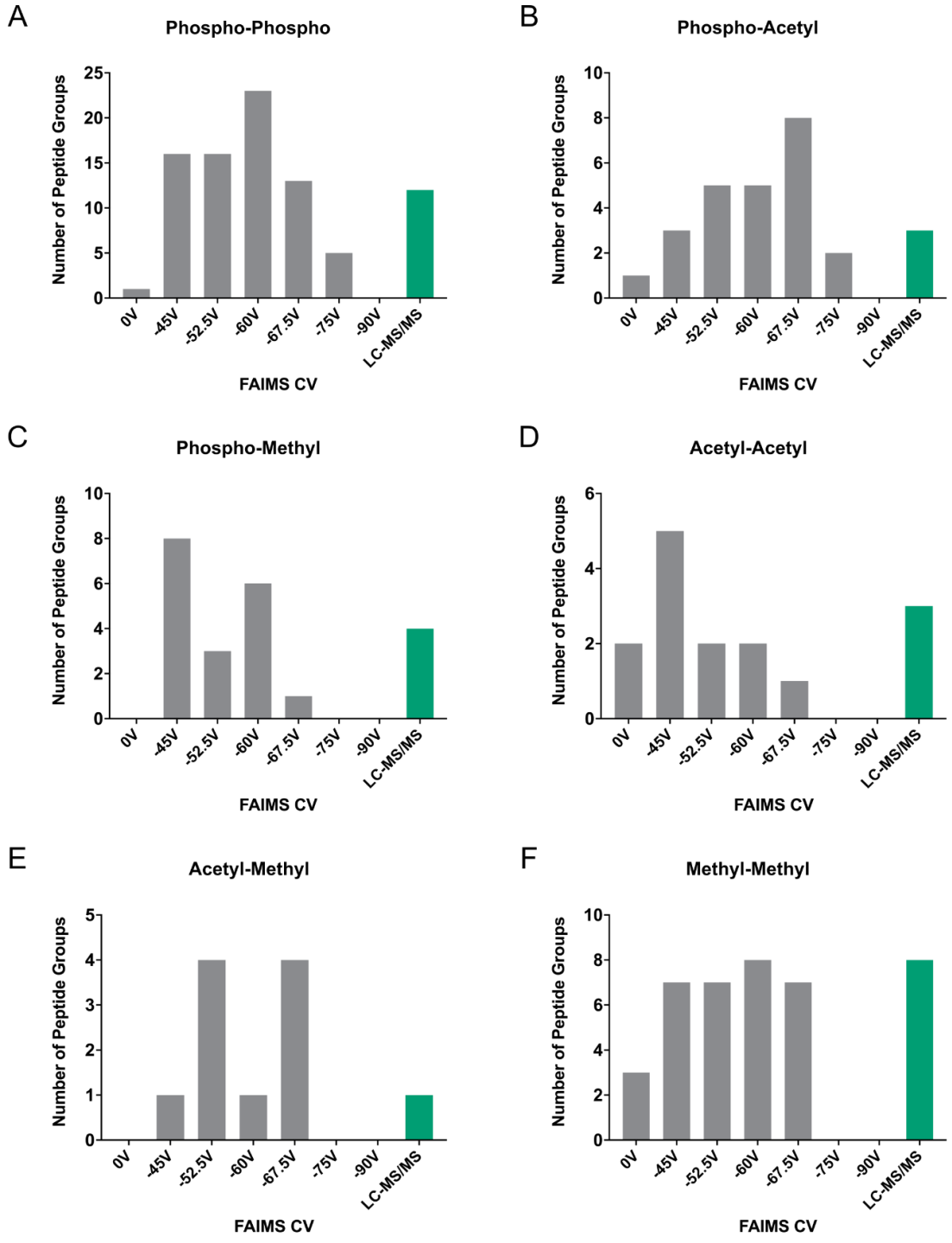


Figure 3.4 | FAIMS improved the identification of candidate PTM crosstalk sites across the different PTM combinations. The number of peptide groups in which multi-PTM peptides were observed as a function of FAIMS compensation voltage (CV) for the PTM combinations: multi-phosphorylation (phospho−phospho) (A), phosphorylation−acetylation (phospho−acetyl) (B), phosphorylation−methylation (phospho−methyl) (C), multi-acetylation (acetyl−acetyl) (D), acetylation− methylation (acetyl−methyl) (E), and multi-methylation (methyl−methyl) (F). The number of peptide groups detected with the FAIMS source removed from the instrument (LC-MS/MS) is included for each PTM combination (green). *Figure taken with permission from Journal of Proteome Research (Adoni, K. R., et al)*[115].

Table 3.1 | Best CVs for FAIMS External Stepping of different PTM combinations.

	Top 4 Ranked CVs (V)			
PTM Combination	**1**	**2**	**3**	**4**
Multi-Phospho	-60	-52.5, -45.0		-67.5
Phospho-Acetyl	-67.5	-52.5, -60.0		-45.0
Phospho-Methyl	-45.0	-60.0	-52.5	-67.5
Multi-Acetyl	-45.0	0.00, -52.5, -60.0		
Acetyl-Methyl	-52.5, -67.5	-45.0, -60.0		
Multi-Methyl	-67.5	-45.0, -52.5, -67.5		

The ion mobility of a peptide is dependent on several factors including mass, charge, and collisional cross-sectional area. When considering all multiple PTM containing peptides, a normal distribution of identifications ranging from 0.00 V to -75.0 V was identified, centred around 60.0 V. However, further analysis of the individual dual-PTM combinations demonstrated differential patterns of peptide transmission through the FAIMS device, corresponding to different PTM combinations. For example, multi-phospho-peptides preferentially transmitted at -60.0 V, whereas phospho-methyl peptides were mostly transmitted at -45.0 V. These findings supported the theory that each chemical modification of a peptide impacts the ion mobility of that peptide in a specific way. The combination of multiple PTMs on the peptide could further modify the peptide such that the CV at which it was best transmitted through the FAIMS device was altered. Myriad physical factors impact how the peptide ion behaves when between the FAIMS electrodes at a specific CV, from the charge distribution of electrons of the peptide, to the electrostatic interactions between different regions of the peptide leading to folding and the relative permittivity of the solvent clusters to the alternating high and low fields of the FAIMS device. Based on this data, using FAIMS increased the

identification of peptide ions with multiple PTMs, relative to standard LC-MS/MS. Further, specific CVs could be selected for internal or external FAIMS CV stepping based on the PTM combinations that was being investigated (Table 3.1).

3.2.3. FAIMS Internal Stepping Accelerated the Identification of Multi-PTM Containing Peptides as Candidates for Positive Crosstalk

Although external stepping demonstrated the improved identification of multi-PTM containing peptides as candidates for positive crosstalk; each static CV LC-FAIMS-MS/MS run took 3 hours, resulting in long experimental run times to acquire all the data for an external stepping LC-FAIMS-MS/MS experiment. Considering the use of 7 CVs, including -90.0 V which yielded no multi-PTM identifications, this represented 21 hours of run time for one technical replicate. The improved identification of peptide and proteins when using the internal stepping FAIMS condition had already been demonstrated (Figure 3.2). Consequently, the utility of cycling through FAIMS CVs using the internal stepping function of the FAIMS Pro^{TM} for the identification of multi-PTM containing peptides was investigated.

3.2.3.1. FAIMS internal stepping boosted the identification rate for multi-PTM peptides when searching for S/T/Y phosphorylation, K/R/Nt acetylation, and K/R methylation

CVs: -45.0 V, -60.0 V, -75.0 V and -90.0 V were chosen as internal stepping CVs, based on previous literature in the field of LC-FAIMS-MS/MS bottom-up proteomics[111,133,188,325] (although -90.0 V proved to be a poor CV for the identification

of multi-PTM containing peptides, with 0 identifications). Internal stepping identified over double as many multi-PTM containing peptides as that of the standard LC-MS/MS methodology, demonstrating its relevance for the detection of candidate peptides for positive PTM crosstalk in the same time frame as standard LC-MS/MS (Figure 3.5). Further, upon studying the overlap of peptide identifications between external stepping, internal stepping, and standard LC-MS/MS; just ten of the 179 multi-PTM peptides were identified across all three proteomics methods, demonstrating the poor overlap of these techniques for multi-PTM peptide identification. Further, just 5 peptides were identified in both external stepping and standard LC-MS/MS, and no peptide overlap was observed between internal stepping and standard LC-MS/MS (Figure 3.5). Together, these findings supported the hypothesis that the incorporation of FAIMS within a bottom-up proteomics workflow allowed for the identification of unique multi-PTM peptide as candidates for positive PTM crosstalk, that were not accessible with standard LC-MS/MS workflows.

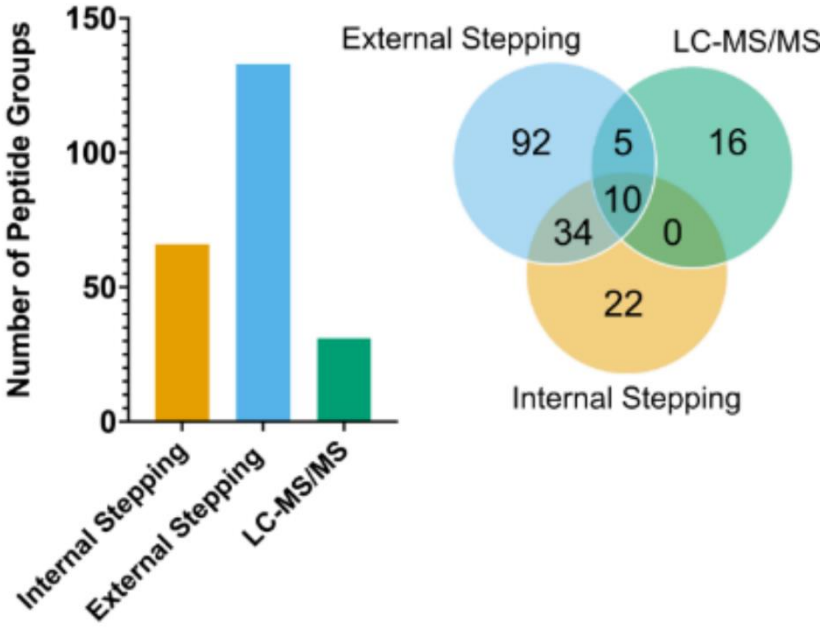


Figure 3.5 | Comparison of internal stepping, external stepping, and standard LC-MS/MS for multi-PTM peptide identifications. Multi-PTM containing peptides as candidates for positive crosstalk for PTMs: S/T/Y phosphorylation, R/K/N-terminal acetylation, and R/K methylation. Venn diagram demonstrates the relatively poor overlap of FAIMS with standard LC-MS/MS for the identification of multi-PTM containing peptides. *Figure adapted with permission from Journal of Proteome Research (Adoni, K. R., et al)*[115].

3.2.3.2. Internal stepping had a differential impact on multi-PTM peptide identification for different PTM-combinations

As presented in Section 3.2.2.2, multi-PTM containing peptides as candidates for positive crosstalk were transmitted across multiple CVs, with different multi-PTM combinations preferentially enriching with specific CVs. Therefore, the relationship between internal stepping using four CVs (-45.0 V, -60.0 V, -75.0 V and -90.0 V), compared to external stepping and standard LC-MS/MS, for specific PTM combinations, was investigated (Figure 3.6). Of note, LC-FAIMS-MS/MS internal stepping identified fewer phosphorylation/methylation containing peptides than did standard LC-MS/MS (Figure 3.6e). Not all the peptides identified in the standard LC-MS/MS were identified in the internal stepping LC-FAIMS-MS/MS condition, perhaps

relating to the differential ion mobility characteristics of FAIMS transmitted peptides against standard LC-MS/MS peptides that do not traverse the cyclic ion mobility device. The overlap between internal and external stepping was lower than expected across all the PTM-combinations, as illustrated by the Venn diagrams corresponding to each PTM-combination (Figure 3.6). For example, of the 68 multiple phosphorylation containing peptides that were detected with LC-FAIMS-MS/MS methodologies (internal and/or external stepping), just 28 % were identified by both stepping conditions. Whilst these findings can be partially reconciled by considering that just 4 internal stepping CVs were used compared to 7 external stepping CVs for external stepping, further analysis of the differential physiochemical properties of the transmitted peptides for both FAIMS conditions should be performed for better characterisation of these differences.

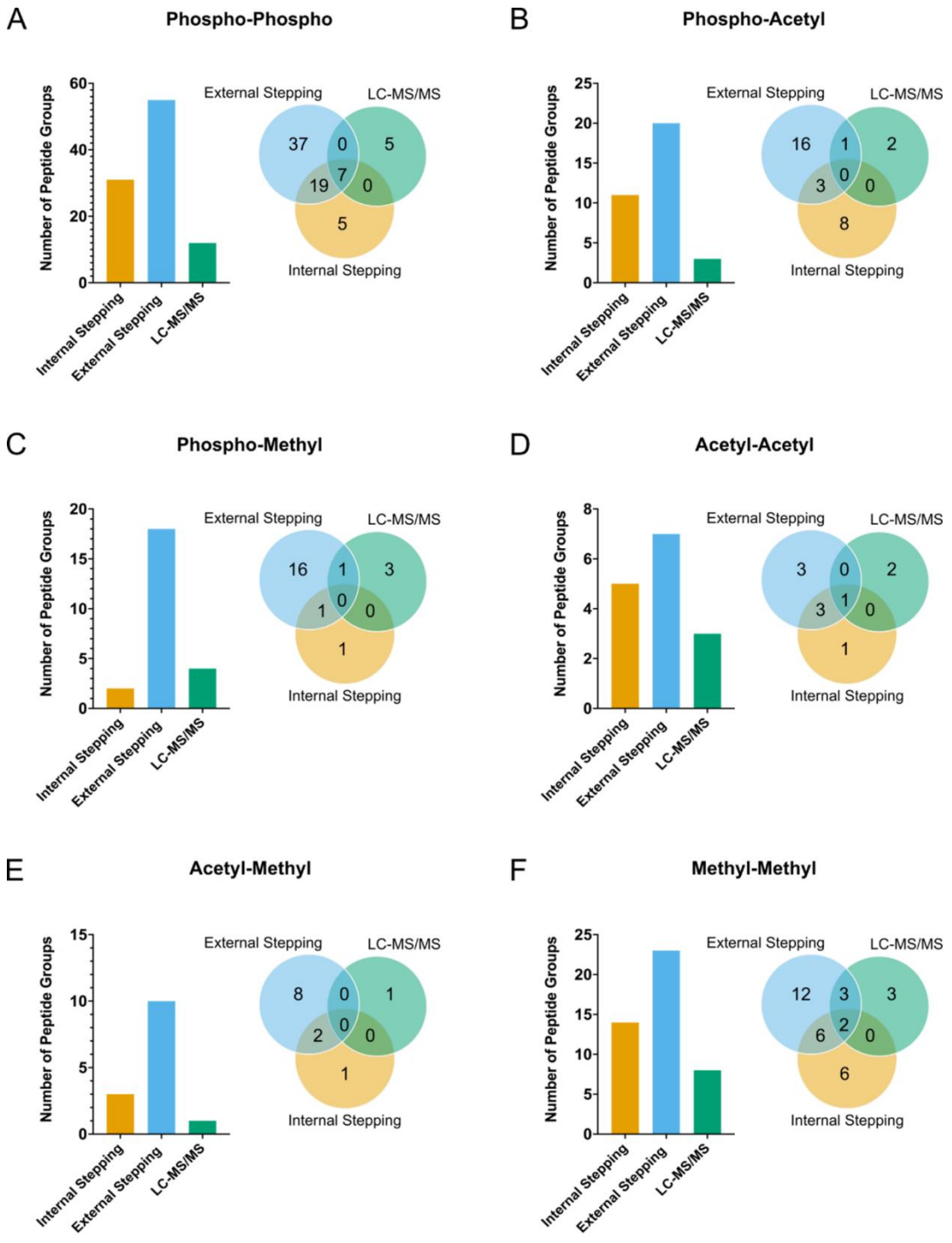


Figure 3.6 | Internal stepping accelerated crosstalk identification rates compared to external stepping. The number of identified multi-PTM peptides versus FAIMS internal stepping, FAIMS external stepping, and standard LC-MS/MS for the PTM combinations: phospho−phospho (A), phospho−acetyl (B), phospho−methyl (C), acetyl−acetyl (D), acetyl−methyl (E), and methyl−methyl (F). Venn diagrams show overlapping peptides detected across the three proteomics conditions. *Figure taken with permission from Journal of Proteome Research (Adoni, K. R., et al)*[115].

Considering the run time and costs of externally stepping across seven CVs with 3-hour LC-FAIMS-MS/MS, it could be argued that these issues outweigh the benefits of this experimental set-up. Hence, the use of internal stepping condition with the FAIMS Pro^{TM} device cycling across -45.0 V, -60.0 V, -75.0 V and -90.0V over a 3 second time-period accelerated the rate of identification of multi-PTM containing peptides as candidates for positive crosstalk. This workflow was investigated as a compromise between improved total number of identifications and rate of identifications of multi-PTM containing peptides as candidates for positive crosstalk. The 2-fold improvement in these identifications on standard LC-MS/MS, compared to a 4-fold improvement when comparing external stepping to standard LC-MS/MS demonstrates the compromise of time and identifications. Based on these results, if time and running costs were irrelevant, externally stepping across CVs: 0.00 V, -45.0 V, -52.5 V, -60.0 V, -67.5 V and -75.0 V would generate the highest number of multi-PTM containing peptides as candidates for positive crosstalk. Considering the distribution of identifications across the FAIMS CVs for specific PTM-combination containing peptides with external stepping (Figure 3.4), the choice of which four CVs that the FAIMS device cycles across in an internal stepping LC-FAIMS-MS/MS run could be specifically tailored for the PTM combination being investigated. To this end, the most appropriate four internal stepping CVs for the different PTM-combinations including phosphorylation (S, T, Y), acetylation (K, R) and methylation (K, R) were reported (Table 3.1).

3.2.4. LC-FAIMS-MS/MS Improved Candidate Positive PTM Crosstalk Identifications with an Increased Repertoire of PTMs within Database Searching

3.2.4.1. LC-FAIMS-MS/MS for the identification of multi-PTM containing peptides as candidates for positive PTM crosstalk using open-PTM searching

Having validated that LC-FAIMS-MS/MS improved the identification of multi-PTM containing peptides, and internal stepping across four CVs increased the rate of identification of these peptides; the viability of increasing the repertoire of PTMs that could be searched within the data processing pipeline was considered. As discussed in Chapter 1, over 400 PTMs have now been reported, with each one imparting a specific biochemical alteration that modifies protein behaviour. For complete characterisation of a protein, the combinatorial decoration of these PTMs on the protein must be investigated. Thus, searching additional PTMs would impart more biological relevance onto the data and generate more multi-PTM containing peptides as candidates for positive crosstalk, from which proteoform variability could be derived. Initially, the feasibility of using open-PTM searching was explored. Open-PTM searching describes a data processing strategy whereby, rather than limiting the search to a pre-defined set of specific PTMs, any peptide with a mass difference to the unmodified precursor ion ranging from -150 to 500 Da (or user defined equivalent) would be identified (discussed in Introduction Section 1.6.5.3). The mass difference of the peptide would therefore correspond to one or multiple PTMs[326,327]. To directly compare open-searching to more traditional closed search strategies, a comprehensive multi-protease phosphoproteomics dataset[328], downloaded from the Proteomics Identification Database (PRIDE PXD001428)[329] database, was processed using MSFragger[213] (open searching) and

Sequest HT[135,317] (closed searching). Open search parameters included peptide mass defects ranging from -150 – 500 Da, and closed search parameters included dynamic S/T/Y phosphorylation, dynamic methionine cleavage with N-terminal acetylation, dynamic methionine oxidation, and static cysteine carbamidomethylation. This comprehensive phosphoproteomics dataset was chosen due to the existence of multiple phosphorylation containing peptides as candidates for positive crosstalk. Sequest identified 1489 multi-phospho PSMs compared to 0 from MSFragger. Further to this, Sequest identified 12803 phosphorylated PSMs compared to 5197 from MSFragger and a total of 13079 PSMs compared to 6756 identified using MSFragger (Figure 3.7).

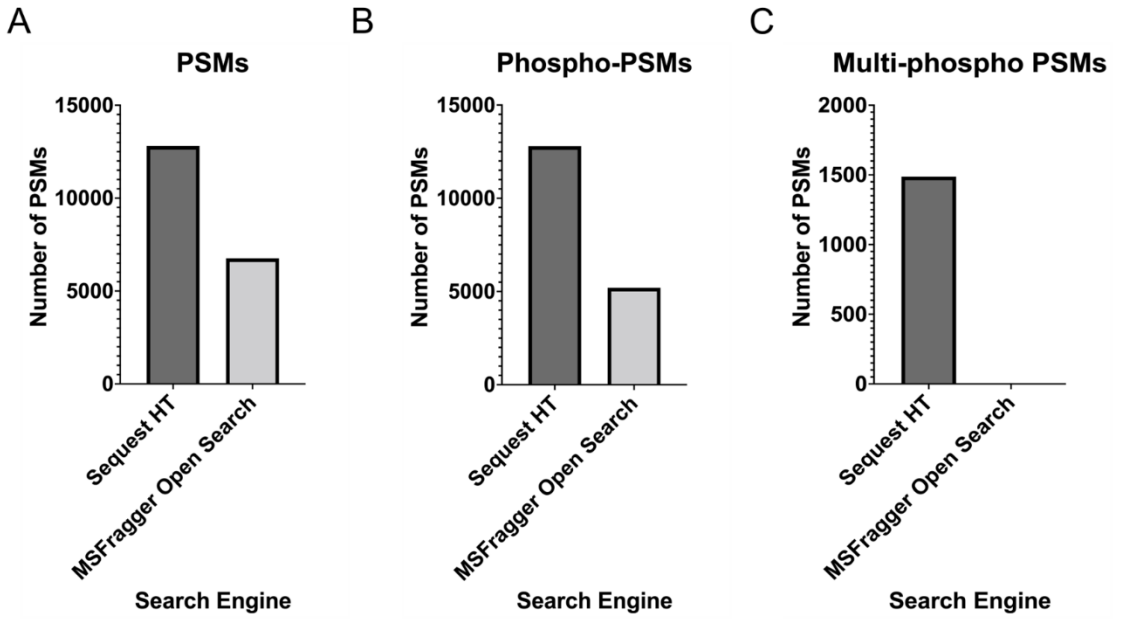


Figure 3.7 | Comparison between Sequest HT and MSFragger open searching. Sequest HT outperformed MSFragger open searching in number of PSMs (A), number of phosphorylated PSMs (B) and number of multi-phosphorylated PSMs (C).

The fundamental objective of our LC-FAIMS-MS/MS workflow was to improve the identification of multi-PTM containing peptides as candidates for positive crosstalk. Thus, the choice of data processing software was a key factor in determining how many

MS/MS spectra would be accurately converted into putative multiple PTM containing peptides. Open searching did not perform well compared the more traditional closed PTM search using Sequest HT (Figure 3.7). 5197 phospho-PSMs were identified in the MSFragger open search, representing 77 % of the dataset. Interestingly, the second most abundant PTM identified with this open-searching strategy (431 PSMs) was pyrophosphorylation (S, T), whereby an initial enzymatically catalysed phosphorylation event facilitates a non-enzymatic β-phosphorylation[330]. Inositol pyrophosphates play the roles of signalling and metabolic messengers and are ubiquitously expressed in eukaryotic cells[330]. However, these pyrophosphate identifications could have represented misassigned multiply phosphorylated peptides as both delta masses correspond to +159.937 Da and the analysis was performed on a comprehensive multi-protease phosphoproteomics dataset that likely contains many multi-phosphorylation containing peptides. Of note, MS/MS analysis for this work was performed in the ion trap[328], with its relatively low resolution mass analysis perhaps facilitating these misassignments. Further to this, analysis of this dataset with MSFragger did not identify any other PTMs beyond unannotated mass shifts, sulfur dioxide and gylcerylphosphorylethanolamine. MSFragger based open searching also generated no examples of multi-PTM containing peptides, suggesting it to be sub-optimal for our purposes (with the search parameters that were selected). Comparatively, the data generated from Sequest HT vindicated the use of a closed PTM-searching system to be more suited to the identification of multi-PTM containing peptides as candidates for positive crosstalk. Of the closed-PTM searching databases; PMi Byonic™ by Protein Metrics was chosen due to its reported ability to handle multiple pre-defined PTMs in the same search[210] (as discussed in Introduction Section 1.6.5.3). Thus, Byonic[207] was

selected as a compromise between the broad PTM identification opportunities of open-searching and the reduced PTM identification search space of closed searching.

3.2.4.2. LC-FAIMS-MS/MS for the identification of multiple-PTM containing peptides as candidates for positive PTM crosstalk with an increased repertoire of PTMs in the closed PTM search strategy

Since MSFragger based open-searching proved less applicable for the identification of multi-PTM containing peptides as candidates for positive crosstalk, the practicality of increasing the number of PTMs in a closed-search based strategy was next probed. Table 3.2 describes the PTMs used, with site localisation and mass defects.

Table 3.2 | PTMs included in broad closed PTM database searching of LC-FAIMS-MS/MS data. Each PTM is reported with its corresponding target amino acid for modification and imparted mass shift on the precursor tryptic peptide.

PTM	**Amino Acid**	**Δ Mass (Da)**
Phosphorylation	S, T, Y	+79.996
Acetylation	K, R	+42.011
Methionine loss + Acetylation	Protein N-terminus	-89.030
Methylation	K, R	+14.016
O-GlcNAc	S, T	+203.079
Ubiquitination	K	+114.042
Deamidation	N	+0.984
Nitrosylation	C, Y	+28.990

O-GlcNAc has been reported to compete with phosphorylation in a reciprocal crosstalking manner[53]. Further to this, complex interplay between O-GlcNAc and phosphorylation has been identified across all compartments of the cell[40]. Ubiquitin and

phosphorylation positive crosstalk has been termed ‘phosphodegrons’, relating to the phosphorylation stimulated ubiquitination of proteins that signal their degradation via the 26S proteosome[331]. Nitrosylation facilitated ubiquitination leading to protein degradation has also been identified in plants[332]. Deamidation, a spontaneous and irreversible modification, has been associated with ageing[333], neurodegenerative-[334] and autoimmune-diseases[335] via degradation signalling[336]. Thus, the PTMs included in the closed search (Table 3.2) were selected due to the biologically relevant nature of these PTMs, their curated identification in databases[141,142,337,338] and their reported involvement in PTM-crosstalk. In total 398 multi-PTM containing peptides as candidates for positive crosstalk were identified, compared to 179 when the closed-PTM search was limited to phosphorylation (S, T, Y), acetylation (K, R) and methylation (K, R) (Figure 3.8). Of these multi-PTM containing peptides, external stepping yielded a 5.3-fold improvement compared to standard LC-MS/MS. Internal stepping across -45.0 V, -60.0 V, -75.0 V and -90.0 V generated a 2.1-fold improved identification in multi-PTM containing peptides compared to standard LC-MS/MS, although, again, no multi-PTM peptides were identified from scans at -90.0 V. In total, LC-FAIMS-MS/MS identified 6-fold more multi-PTM containing peptides as candidates for positive crosstalk than did just standard LC-MS/MS. Further, of the multi-PTM containing peptides that were identified in this dataset, 159 have not previously been reported. Thus, this work suggested that by reducing background peptide ions during the ion mobility dimension of this bottom-up proteomics workflow, low stoichiometry multi-PTM containing peptides are more likely to be detected as precursor ions and characterised with MS/MS.

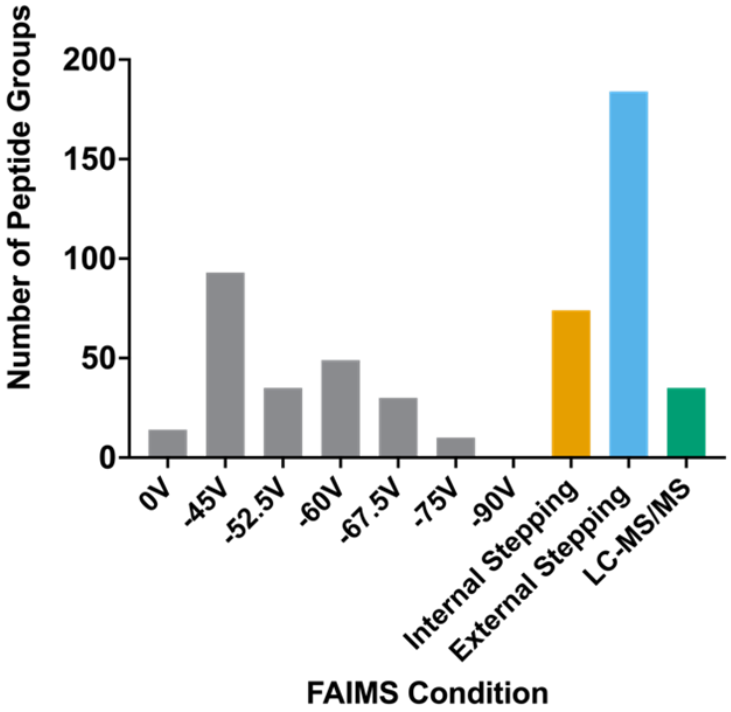


Figure 3.8 | Expanding the repertoire of PTMs demonstrated that more candidate PTM crosstalk sites were identified. The number of identified multi-PTM peptides for different FAIMS conditions including internal stepping (orange) and external stepping (blue), standard LC-MS/ MS (green), as well as each static LC-FAIMS-MS/MS CV (grey). The PTMs included in the database search were phosphorylation (S, T, Y), deamidation (N), nitrosylation (C, Y), acetylation (K, R, N-terminal), methylation (K, R), ubiquitination (K), and GlcNAcylation (S, T). *Figure taken with permission from Journal of Proteome Research (Adoni, K. R., et al)*[115].

3.2.5. Investigating the Order of PTM Occurrence in Positive PTM Crosstalk Using LC-FAIMS-MS/MS

The above data validated the potential of FAIMS for the increased identification of multi-PTM containing peptides as candidates for positive crosstalk in a bottom-up proteomics workflow. However, for a peptide that is modified by multiple PTMs, information relating to the order of occurrence of these modifications is necessary to characterise the relationship of positive PTM crosstalk. Thus, whether the proteomics workflow developed in this chapter could be used to probe the order of modification in multi-PTM containing peptides was next investigated. It was hypothesised that the

presence of multiple isoforms of a peptide, with varying levels of modification, could be used to propose a mechanism for the specific priming of positive crosstalk, leading to the generation of a specific protein proteoform. To demonstrate this; MS/MS spectra corresponding to both the monophosphorylated (containing a serine-444 phosphosite, Figure 3.9a) and twice-phosphorylated (containing serine-437 and serine-444 phosphosites, Figure 3.9b) isoforms of the peptide amino acid residues: 432-455 from protein FACT Complex Subunit SSRP were identified and annotated (Figure 3.9). Importantly, the isoform with a single serine-437 phosphorylation was not identified. These data indicated a possible phospho-priming mechanism of positive PTM crosstalk, with serine-444 phosphorylation required to stimulate serine-437 phosphorylation. Further work, such as Western blotting against each PTM upon stimulation of serine-444 phosphorylation, or synthetic peptide *in-vitro* analysis of serine-437 phosphorylation after serine-444 phosphorylation (and vice versa) would be required to corroborate this model for phospho-priming positive PTM crosstalk.

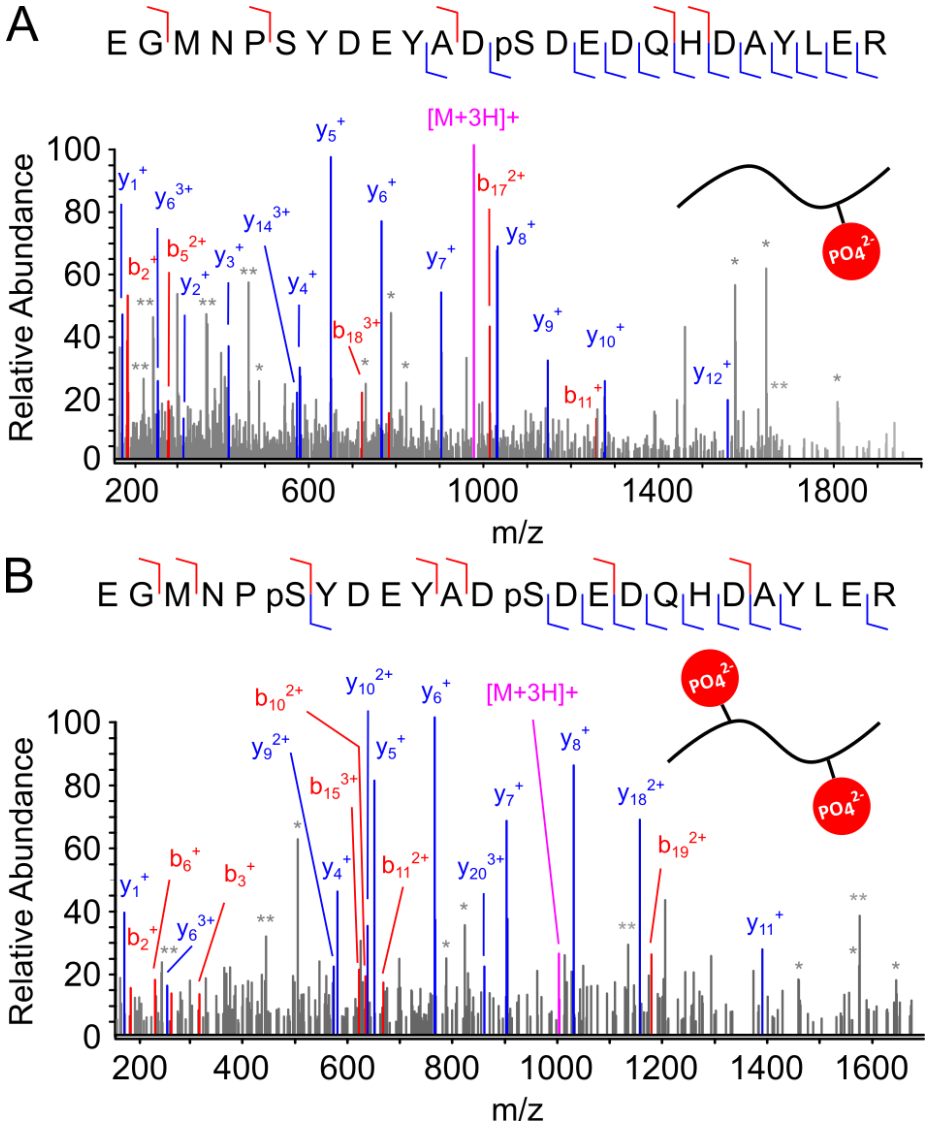


Figure 3.9 | HCD MS/MS spectra for peptide sequence 432-455 from FACT complex subunit SSRP1 in its monophosphorylated (A) and dual-phosphorylated peptidoform (B). Red peaks represent *b* ions, blue peaks represent *y* ions. Pink peaks represent the precursor ion. Peaks marked * and ** represent *b* and *y* ions with phosphate loss and internal fragment ions, respectively. Schematic representation of each peptidoform is included, with PO_4^{2-} representing the phosphorylation PTM.

In the case of Polyadenylate Binding Protein 2 (PABP2); whereby different isoforms of peptide amino acid residues: 1-22 were identified, positive crosstalk between acetylation and methylation could be inferred (Figure 3.10). These three isoforms, together with the unmodified isoform, were the only isoforms of this peptide that were identified,

suggesting that the initial N-terminal acetylation could facilitate arginine-16 methylation followed by arginine-19 methylation. Again, further work would be required to validate this hypothesis.

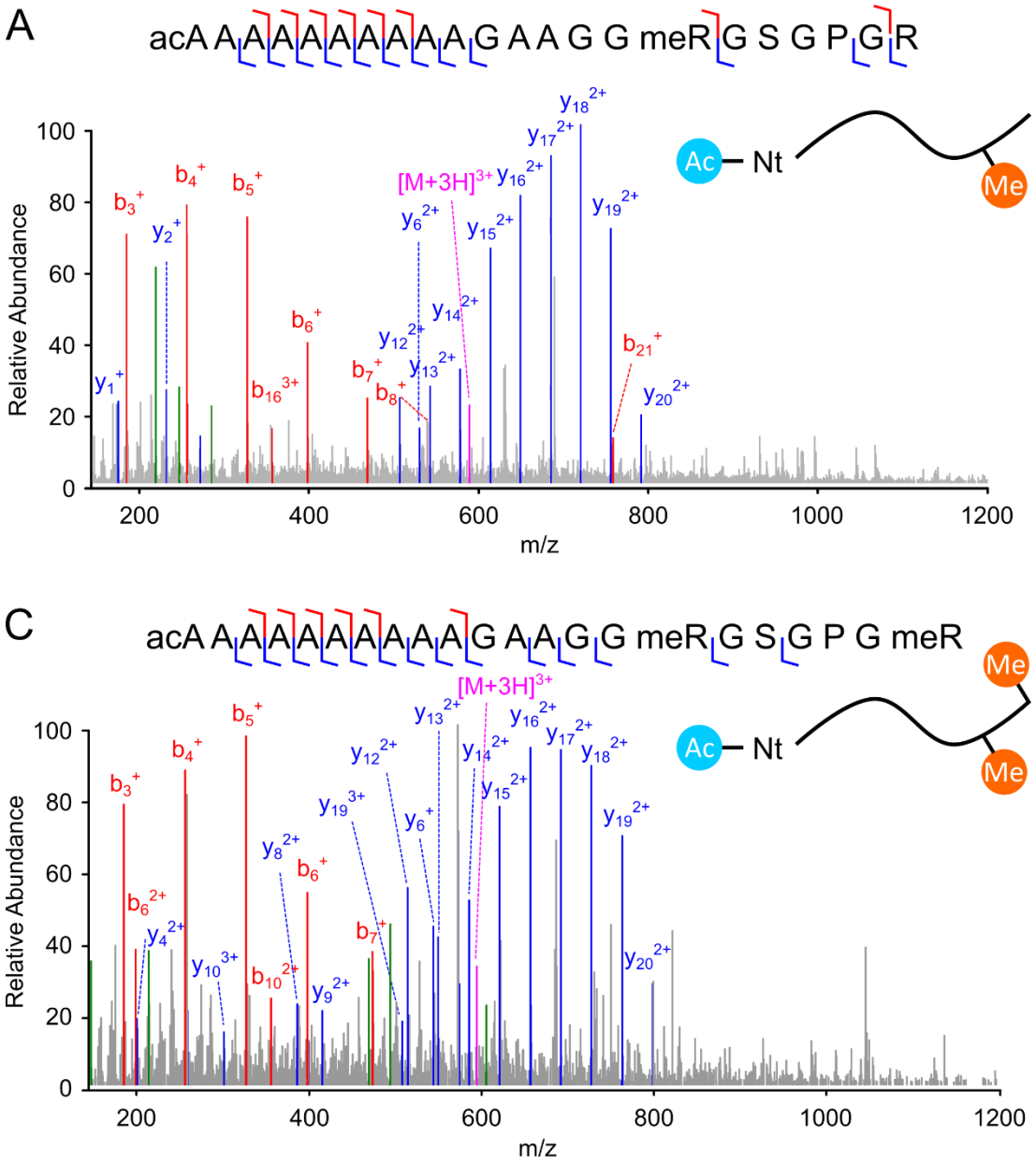


Figure 3.10 | HCD MS/MS spectra for the peptide sequence 1-22 of PABP2. The peptide was identified in its N-terminally acetylated and arginine-16-methylated peptidoform (A), its N-terminal acetylated, arginine-16-methylated and arginine-22 methylated peptidoform (B) as well as being identified as an N-terminally acetylated peptidoform. Red peaks represent *b* ions, blue peaks represent *y* ions. Pink peak represents precursor ion and green peaks represent internal fragments. Schematic representation of each peptidoform is included, with Nt representing the N-terminal site, Ac representing the acetylation PTM and Me representing the methylation PTM.

3.2.6. Proteins Derived from Different Multi-PTM Combination Peptides Differentially Enrich for Specific Gene Ontologies

Specific PTMs have their own well characterised roles in the biological regulation of protein systems; for example histone acetylation is associated with increased gene expression[339], glycosylation is involved in protein folding, transport and localisation as well as host-pathogen interaction[340], and histone methylation is associated with compaction of chromatin[341]. As such, the role of multiple-PTM combinations in specific biological processes was next investigated. As discussed above, the data from these experiments generated novel examples of multi-PTM containing peptides as candidates for positive crosstalk, across all PTM-combinations that were searched. Thus, whether proteins derived from the individual PTM combinations enriched for certain biological processes or cell compartments would be of interest to the PTM-crosstalk field. g:profiler[304] gene ontology analysis software was used to explore these possibilities (Table 3.3). PTM combinations involving phosphorylation (S, T, Y), acetylation (K, R) and methylation (K, R) were specifically selected for this purpose.

Table 3.3 | Gene Ontology Enrichment Analysis of proteins derived from different PTM-combinations. PTMs: phosphorylation, acetylation, and methylation**.** Analysis performed with g:profiler[304]

PTM Combination	Biological Process	Cell Compartment	Number of Proteins
Multi-Phospho	RNA binding and stability	Nucleus	71
Phospho-Acetyl	Protein import	Membranes	30
Phospho-Methyl	RNA binding and activity	Nucleus	23
Multi-Acetyl	Cytoskeletal dynamics	Cytoskeleton	10

Acetyl-Methyl	Low Density Lipoprotein (LDL) receptor binding	Lumen (ER)/Extracellular matrix	12
Multi-Methyl	RNA binding	Cytoplasm	32

The enrichment of multi-acetylated candidates for PTM crosstalk in cytoskeletal dynamics was supported by previous data that associates lysine acetylation with cytoskeletal assembly and dynamics via its modification of tubulin, cortactin and mDia2[342] (with the importance of these regulatory mechanisms investigated in Chapter 4). Further, arginine methylation of RNA binding proteins has been reported as a mechanism of cellular function and differentiation[343]. Ultimately, the differential enrichment of proteins derived from each PTM-combination for biological processes and cell compartments suggested the existence of some specialised roles for each multiple-PTM combination in cell biology. However, examples of each PTM-combination have been identified across several different biological processes, indicating a highly complex regulatory system for PTM crosstalk and its relationship with biological functions.

3.2.7. Limitations of the LC-FAIMS-MS/MS Methodology for the Identification of Positive PTM Crosstalk

3.2.7.1. Positive PTM crosstalk often occurs between amino acids sites that are far apart in primary sequence.

The interaction of PTMs with each other, described as PTM-crosstalk, is often a tertiary and quaternary structure phenomenon[344]. The proximity of the two modified amino acids

could be far apart in primary sequence, with multiple arginine/lysine residues between them. As such, trypsin digestion of the protein into peptides eradicates any information pertaining to the protein proteoforms in the context of the two or more PTMs that are from separate tryptic peptides. This could be partially nullified using alternative protein digestion techniques that facilitate longer peptide generation relative to standard bottom-up trypsin digestion. Although there was an order of magnitude fewer multi-PTM containing peptides compared to unmodified peptides in the dataset, multi-PTM peptides were longer on average than their non-modified counterparts (Figure 3.11). This suggested that longer peptides are more likely to contain multiple PTMs, and their longer sequence is generally accompanied by higher charge, a factor that was considered in the decision to set a charge state range of 3^+ - 8^+ for precursor ions in the mass spectrometry methodology. Whether this preselected precursor charge state range aided in the identification of multi-PTM containing peptides as candidates for positive crosstalk is debatable, as the loss of multiply modified 2^+ peptide ions possibly outweighed the biasing of the instrument to pick up longer peptides that were more likely to be modified. Examples of proteases that result in longer peptides include: ArgC, AspN, Chymotrypsin, GluC and LysC[345]. Another approach for the generation of longer peptides involves reducing trypsin incubation time or temperature. Arguably, both of these processes delve into middle-down proteomics territory, and the use of FAIMS with middle down proteomics to resolve PTM isomers has shown success[346].

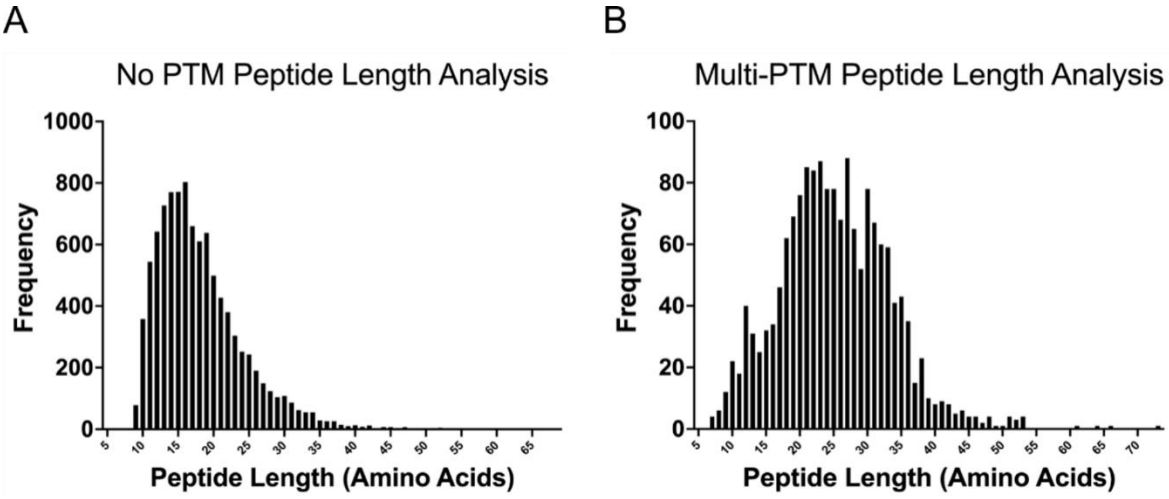


Figure 3.11 | Multiple PTM-containing peptides were longer in sequence than the unmodified peptides identified in this study. To generate this data, all peptide group identifications within the study were filtered for high confidence peptides using an FDR of 0.01 (with their identification across biological replicates not considered). The frequency of peptide groups that contained no modifications (A) or a combination of two or more modifications (B) were then plotted against their sequence length.

3.2.7.2. Negative and reciprocal PTM-crosstalk cannot be deduced by bottom-up LC-FAIMS-MS/MS

Whilst this workflow demonstrated an improvement in the identification of multi-PTM containing peptides as candidates for positive crosstalk; two important PTM crosstalk mechanisms were not considered. Negative PTM crosstalk, whereby one PTM inhibits the occurrence of another PTM[331], was not a focus of the workflow. Reciprocal PTM crosstalk, whereby PTMs compete to modify the same amino acid or amino acids in close proximity[39], was also ignored. The data processing focussed on validating multi-PTM containing peptides as candidates for positive crosstalk and mono-modified peptides with negative or reciprocal crosstalk relationships were ignored. LC-FAIMS-MS/MS' ability to improve PTM identification has been reported[319,320] and our dataset contained thousands of mono-modified peptides. Purpose built data processing software

could deconvolute putative negative and reciprocal crosstalking relationships between PTMs based on their peptide isoforms, as well as predicting important sequence motifs as pre-requisites for PTM-crosstalk. To this end, EpiProfile 2.0[347] is a computational platform that uses retention time prediction and peak detection from a data independent acquisition (DIA) workflow to identify isobaric peptides with differential PTM localisation.

3.2.7.3. Ion trap analysis of HCD fragmented peptides hindered accurate PTM site localisation

To ensure optimal throughput of the HeLa digested lysate, the linear ion trap was selected over the Orbitrap for fragment mass analysis. Thus, HCD fragments could be analysed by the ion trap whilst precursor ions were simultaneously scanned in the Orbitrap to improve the duty cycle. Consequently, the utilisation of all three mass analysers in the Tribrid instrument (including the quadrupole for *m/z* filtering) led to improved peptide identifications[109]. However, the lower mass resolution of the ion trap compared to the Orbitrap dictated that a fragment mass tolerance of 0.5 Da was chosen when processing the RAW data. This relatively low MS/MS mass tolerance hindered the confident assignment of PTM site localisation, as well as increasing the likelihood of PTM and amino acid misassignment[66,217]. This issue was considered in the data analysis, with all validated multi-PTM containing peptides as candidates for positive crosstalk passing a rigorous set of criteria (see Materials and Methods Section 2.1.3). These criteria possibly led to an underestimation of multi-PTM containing peptides as candidates for positive crosstalk. Despite this, the relative ambiguity of *b* an *y* ion identification in the ion trap cannot be fully rectified with statistical robustness and the

false localisation rate of PTMs was compromised by using the ion trap mass analyser for MS/MS analysis.

3.2.8. Future Directions of LC-FAIMS-MS/MS for PTM Crosstalk Characterisation

The above experiment demonstrated the benefits of using such a workflow in the detection of multi-PTM containing peptides as candidates for positive crosstalk. As the combinatorial impact of PTMs on protein regulation takes on a central role in the study of protein biochemistry, so too will the need for techniques that identify low stoichiometry multiply modified peptides in bottom-up proteomics. As such, LC-FAIMS-MS/MS, with its ability to supress background non-modified peptide ions in favour of multiply modified peptide identifications, is well positioned to meet this challenge. Deconvoluting the myriad PTM combinations at the N-terminus of histone proteins, dubbed 'the histone code', is an example of this new challenge in PTM analysis and FAIMS has been used in a middle-down context for this purpose[48]. However, the high-throughput nature of LC-FAIMS-MS/MS as a bottom-up technique is more attractive for global PTM-omics studies. Many PTM combinations have now been shown to regulate each other[331], for example; with 'phosphodegrons', as phosphosites that prime E3 ligases to ubiquitinate the protein and signal it for 26S proteasomal degradation, now recognised as key regulators of protein turnover in many signalling pathways[348]. Multi-PTM containing peptides as candidates for positive crosstalk such as this must first be detected, for biochemists to fully understand how PTMs regulate proteins at the proteoform level.

3.3. Conclusions

Incorporation of FAIMS into an automated LC-MS/MS workflow (LC-FAIMS-MS/MS), without prior peptide fractionation or PTM enrichment, reported a 6-fold improvement in the number of multi-PTM containing peptides as candidates for positive crosstalk when compared to standard LC-MS/MS. Of note, 40 % of the multi-PTM containing peptides as candidates for positive crosstalk identified in the dataset have not previously been reported, demonstrating the capacity of LC-FAIMS-MS/MS in the investigation of PTM crosstalk. Additionally, through the optimisation of the FAIMS parameters in an internal stepping workflow, the effectiveness of LC-FAIMS-MS/MS in the context of PTM crosstalk analysis was demonstrated on a timescale comparable to standard LC-MS/MS experiments, doubling the number of multi-PTM sites detected. PTM crosstalk is now a widely appreciated mechanism of protein regulation. As such, the incorporation of this workflow into routine bottom-up proteomics experiments could greatly improve our understanding of protein behaviour, including that of signalling pathways across the human proteome. Further, the technique could add valuable detail to our mechanistic understanding of how dysfunctional regulation of proteins from aberrant PTM crosstalking relates to disease phenotypes.

4. SILAC LC-FAIMS-MS/MS for the Investigation of Novel PTM Crosstalk in SUM52 Breast Cancer

Proteomics data corresponding to this chapter including protein, peptides and PSM hits, as well as all MS/MS spectra are available on the University of Birmingham Research Archive (UBIRA): https://doi.org/10.25500/edata.bham.00000933 (Pure ID: 188436277)

4.1. Overview

As discussed in Chapter 3, LC-FAIMS-MS/MS can be used to identify novel multi-PTM containing peptides as candidates for positive PTM crosstalk. Such peptides, within a bottom-up proteomics workflow, can provide valuable information regarding combinatorial PTM regulation across protein systems[349]. In this chapter, LC-FAIMS-MS/MS was applied to investigate positive PTM crosstalk levels and their association with the cancer phenotype. We focussed on a SUM52 triple negative breast carcinoma cancer cell line, which harbours a Fibroblast Growth Factor Receptor (FGFR) 2 gene amplification[350]. SILAC labelling was used to investigate the changes in combinatorial PTM status upon administration of two different FGFR inhibitor drugs; AZD4547 (FGFR kinase specific inhibitor) and SU5402 (multi-target receptor tyrosine kinase inhibitor). Such small molecule drugs bind the Adenosine Triphosphate (ATP) binding cleft of the receptor tyrosine kinase to competitively inhibit kinase activity, causing lethality by prevention of downstream signalling[264,351,352]. The identification of novel PTMs, that were differentially regulated by FGFR inhibition, revealed changes in the combinatorial PTM status of proteins in biological pathways including mitosis, protein synthesis and cytoskeletal regulation. These proteoform changes, by decoration with multiple PTMs, were considered in the context of the hallmarks of cancer that lead to the cancer phenotype, which describes six biological capabilities that are acquired during the development of human cancer[353]. One such hallmark, the activating invasion and metastasis of cancer cells, was further investigated in the context of actin cytoskeletal regulation[353]. Within the cytoskeletal protein regulating machinery, 69 manually validated PTM sites as candidates for positive PTM crosstalk were detected, whose abundance was significantly modified with inhibition of the FGFR signalling pathway.

Of these, 54 % have not previously been reported. The differential patterns of PTM regulation for SUM52 cells with and without FGFR inhibition shed light on the combinatorial-PTM coordination of proteins as cancer cells adapt their cytoskeletal dynamics towards a cell migration phenotype, priming the eventual metastasis of cancer tissue.

4.2. Results and Discussion

4.2.1. SUM52 Cells Overexpress FGFR2 Kinase

SUM52 cell lines overexpress the FGFR2 gene, resulting in overactive FGFR2 downstream signalling of the RAS/MAP/ERK/RSK and PI3K/PDK1/AKT pathways amongst others[252]. Terminal targets of these kinase-dependant pathways lead to increased gene expression for proliferation, cell motility, metabolism and survival related proteins[177,350]. Cultured SUM52 cells were tested for FGFR2 gene amplification leading to FGFR2 overexpression and hyperphosphorylation. Western blotting of SUM52 lysate against anti-FGFR2, anti-phosphoFGFR2 (phosphotyrosine-656, phosphotyrosine-657) and anti-tubulin loading control antibodies validated the overexpression of FGFR2 as well as hyperphosphorylation of FGFR2 at activation loop tyrosine-656 and tyrosine-657, which are critical for FGFR2 kinase activation via autophosphorylation (Figure 4.1).

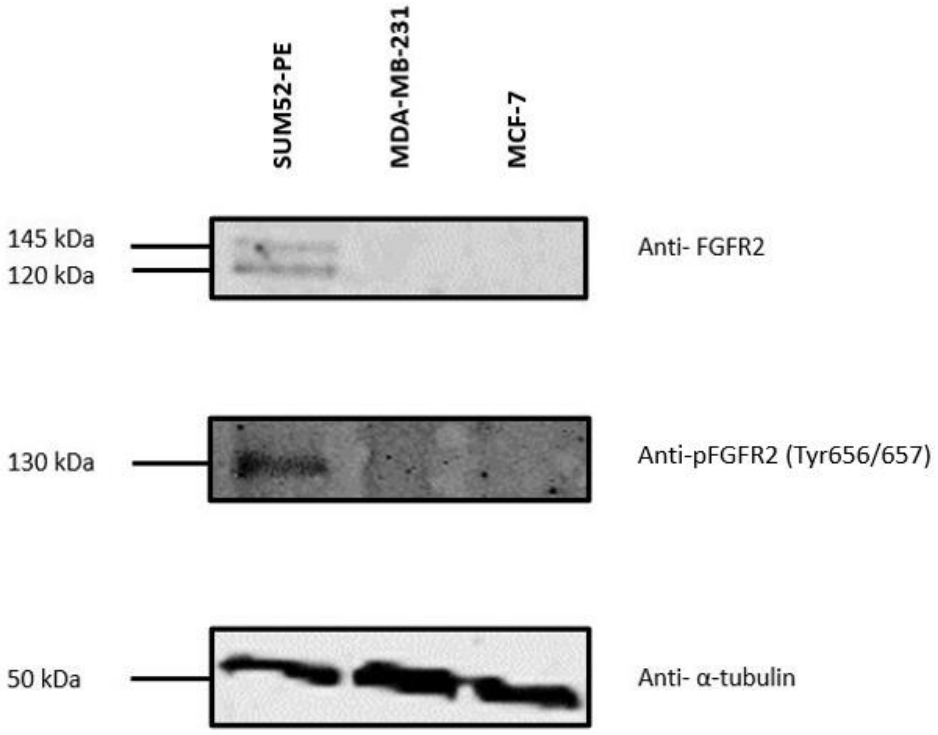
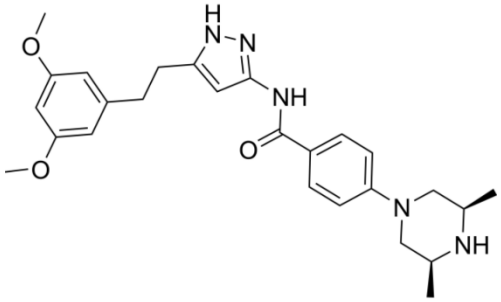


Figure 4.1 | Western blotting validated the overexpression of FGFR2 and FGFR2 activation loop phosphotyrosine-656 and phosphotyrosine-657 in SUM52. This was compared to non-FGFR gene amplified breast cancer cell lines MDA-MB-231 and MCF-7. The unannotated Western blot is provided in Appendix Section 8.4.

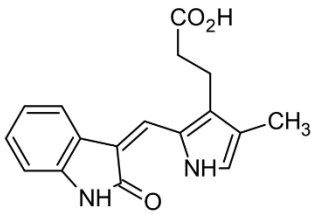
4.2.2. Cell Viability Assays to Optimise FGFR Inhibitor Concentrations for SUM52 cell line

As the SUM52 cell line is addicted to FGF activation for cell viability, these cells are highly sensitive to FGFR inhibitor drugs[354]. SU5402 is a potent inhibitor of FGFR, binding to the ATP pocket and inhibiting autophosphorylation of receptor tyrosine kinases: FGFR, Vascular Endothelial Growth Factor Receptor, and Platelet-Derived Growth Factor Receptor, with different half maximal inhibitory concentration (IC_{50}) values for each receptor[355]. AZD4547 represents a newer and more specific inhibitor of FGFR, with minimal activity on other receptor tyrosine kinases[356]. Similar to SU5402, AZD4547 localises to the FGFR ATP binding cleft and inhibits FGFR

autophosphorylation[356]. As both drugs supress downstream signalling of FGFR2, their differential effect on the PTM status of proteins across different biological protein systems compared to a non-drug administered SUM52 sample (drug-free control) was used to probe the combinatorial impact of PTMs downstream of FGFR kinase within the SUM52 breast cancer cell line. As such, optimal AZD4547 and SU5402 GI_{50} (half maximal growth inhibitory concentration) concentrations for SUM52 were acquired. Cell TiterGlo® based drug titration viability assays were performed on SUM52 and MDA-MB-231 cell lines. Whilst SUM52 cells were highly responsive to FGFR inhibition, MDA-MB-231 cells were insensitive to these FGFR inhibitor molecules (Figure 4.1). The drug titration assays elucidated a GI_{50} of 8 nM and 20 μM for AZD4547 and SU5402 respectively (Figure 4.3).



AZD4547



SU5402

Figure 4.2 | Structures for AZD4547 and SU4502.

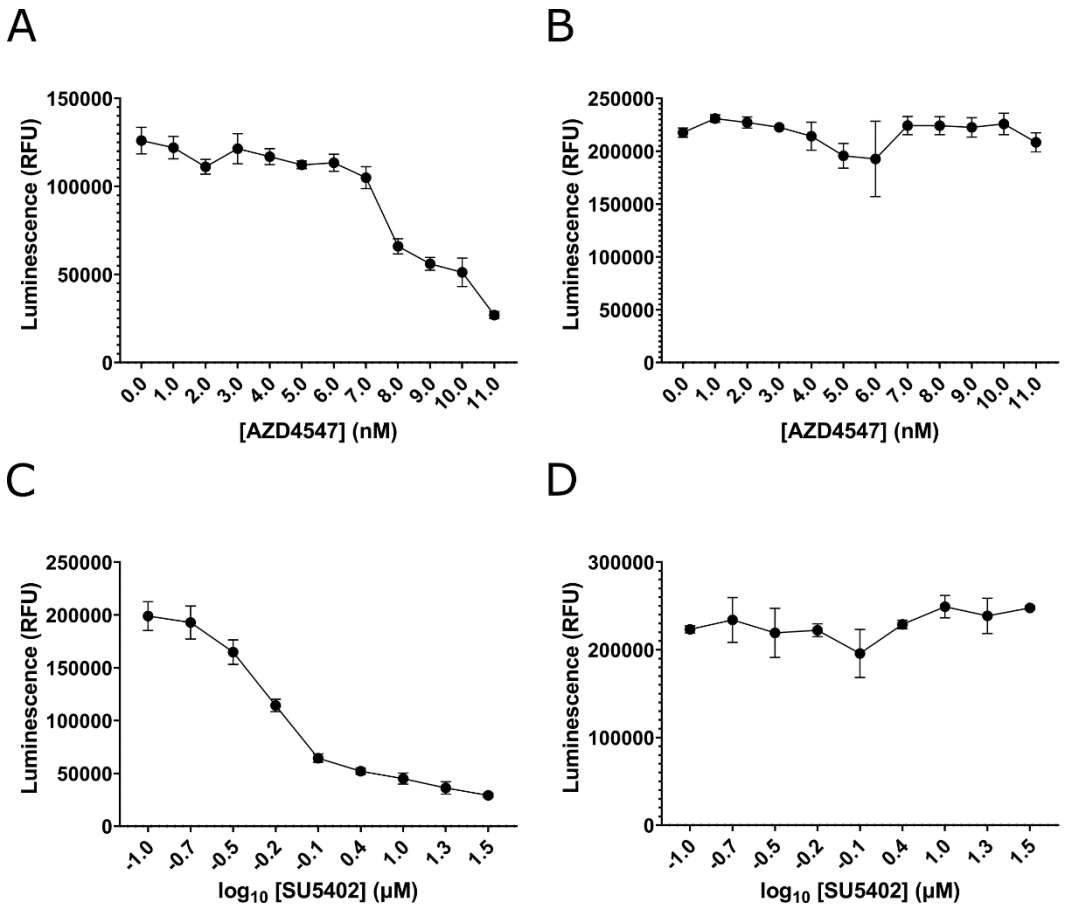


Figure 4.3 | Cell TiterGlo based drug titration assays. Luminescence correlates to viable cell population at variable drug concentrations. (A) Administration of AZD4547 to SUM52 with concentrations ranging from 0.0 nM to 11 nM. (B) Administration of AZD4547 to MDA-MB-231 with concentrations ranging from 0.0 nM to 11 nM. (C) Administration of SU5402 to SUM52 with concentrations ranging from 0.0 μM – 32 μM, displayed using a logarithmic scale. (D) Administration of SU5402 to MDA-MB-231 with concentrations ranging from 0.0 μM – 32 μM, displayed using a logarithmic scale. All assays were performed in triplicate with standard deviation per concentration represented with error bars. MDA-MB-231 cell viability was not affected by AZD4547 or SU5402. Graph generation and statistical analysis for standard deviation performed in GraphPad Prism 9.

4.2.3. SILAC LC-FAIMS-MS/MS Analysis of SUM52 Revealed Differential Inhibition of FGFR2 Activation Loop Autophosphorylation for AZD4547 vs SU5402

4.2.3.1. AZD4547 and SU5402 downregulated FGFR2 autophosphorylation with differential levels of inhibitory activity

FGFR2 tyrosine-656 and tyrosine-657 autophosphorylation at the activation loop are required for maximal FGFR2 kinase activity[357] (Figure 4.4a). Thus, to validate the FGFR inhibitory effect of AZD4547 and SU5402, the differential abundance of the phosphopeptide corresponding to residues 650-659 (containing activation loop phosphosites at tyrosine-656 and tyrosine-657) were probed using LC-FAIMS-MS/MS with SILAC quantitation for light (control), medium (AZD4547) and heavy (SU5402) metabolic labels (SILAC experiment workflow illustrated in Figure 4.5). Chromatogram traces corresponding to the different labels revealed the inhibition of this autophosphorylation activity upon administration of both FGFR inhibitors (Figure 4.4b). Overall abundance of the FGFR2 peptide corresponding to residues 197-203 (identified using external stepping LC-FAIMS-MS/MS), distinct from the activation loop, was not significantly modified by either FGFR inhibitor, indicating no significant change in overall FGFR2 protein abundance across the conditions (Figure 4.4b). TiO_2-phosphoenrichment pre-internal stepping LC-FAIMS-MS/MS was used to identify activation loop phosphopeptide residues 650-659 (containing phosphotyrosine-656 and phosphotyrosine-657). SILAC quantitation revealed this peptide to be significantly reduced by both drugs (Figure 4.4b). AZD4547 downregulated the abundance of this phosphopeptide by more than 2-fold, whilst SU5402 completely abrogated the identification of this dual phosphorylated peptide. This suggested more potent inhibition

of FGFR2 and its downstream signalling with SU5402 compared to AZD4547 at the concentrations administered. Consequently, changes in the PTM status of proteins would be expected to be different between the two FGFR inhibitors, with the PTM status of SU5402 data resulting from complete loss of FGFR2 downstream signalling, whereas AZD4547 changes would result from partial downregulation of FGFR2 signalling. Further to this, the mono-phosphorylated FGFR2 peptide 650-659 was identified as both phospho-peptide isoforms in the control (phosphotyrosine-656 and phosphotyrosine-657), indicating no significant priming effects of one phosphosite to the other. SILAC quantitation of the FGFR2 activation loop peptide and a non-activation loop peptide validated the specific downregulation of FGFR2 kinase activity (and consequently, its downstream signalling pathways) upon both FGFR inhibitors AZD4547 and SU5402, as previously reported[177,264].

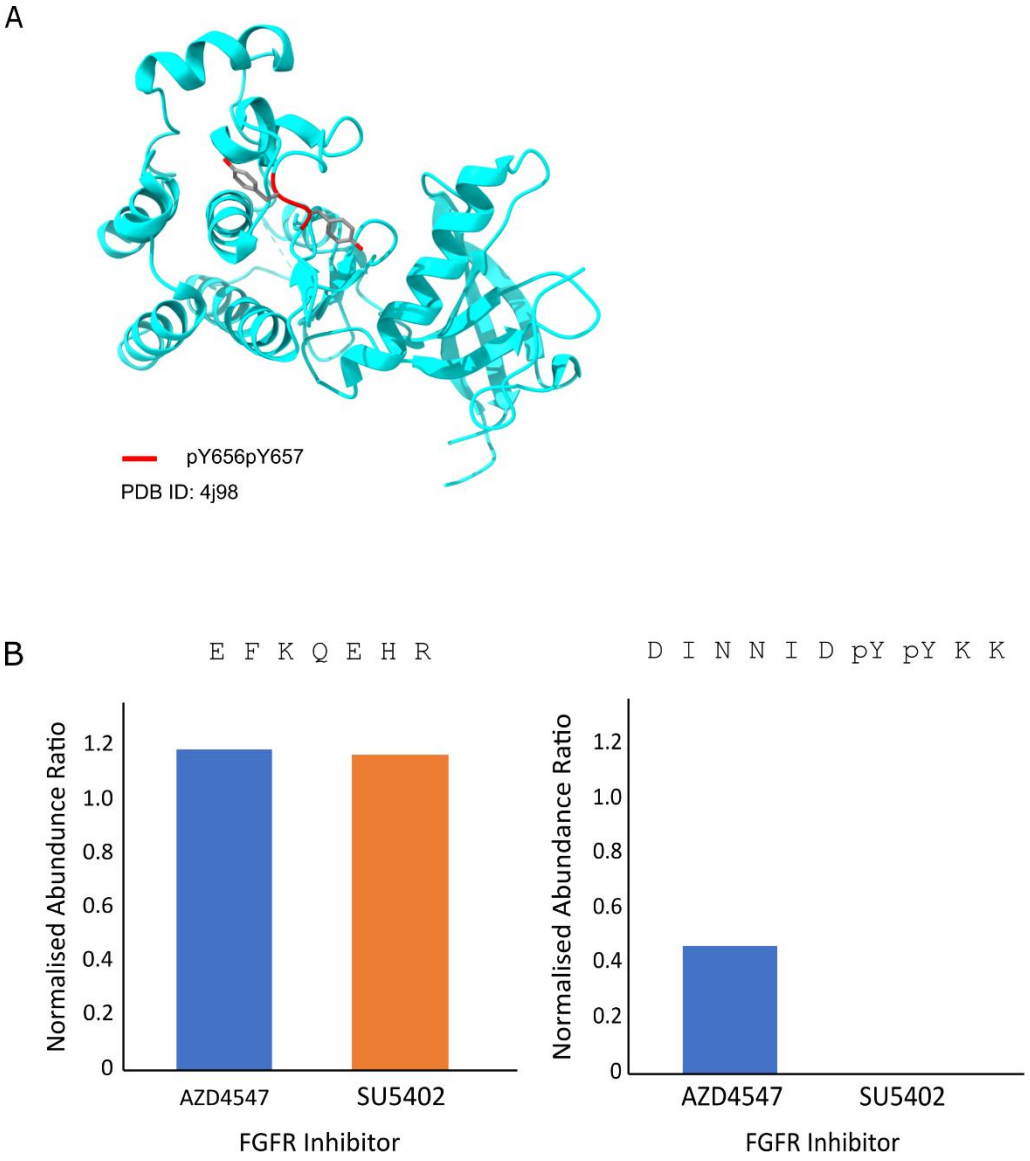


Figure 4.4 | Both FGFR2 inhibitors downregulated activation loop phosphorylation of tyrosine-656 and tyrosine-657 with downregulation of subsequent downstream signalling. (A) PDB Crystal structure of FGFR2[358] with tyrosine-656 and tyrosine-657 annotated (red). PDB ID: 4j98. Structure visualisation and annotation performed using Chimera v1.3[359]. (B) SILAC normalised abundance ratio (vs control) plot for FGFR2 peptide amino acid sequence residues 197-203 (identified in external stepping non-enriched LC-FAIMS-MS/MS experiment, left) and dual-phosphorylated activation loop FGFR2 peptide amino acid sequence residues 650-659 (identified in internal stepping TiO2-enriched LC-FAIMS-MS/MS experiment, right). Both drugs downregulated phosphotyrosine-656 and phosphotyrosine-657 autophosphorylation, with SU5402 abrogating these phosphosites.

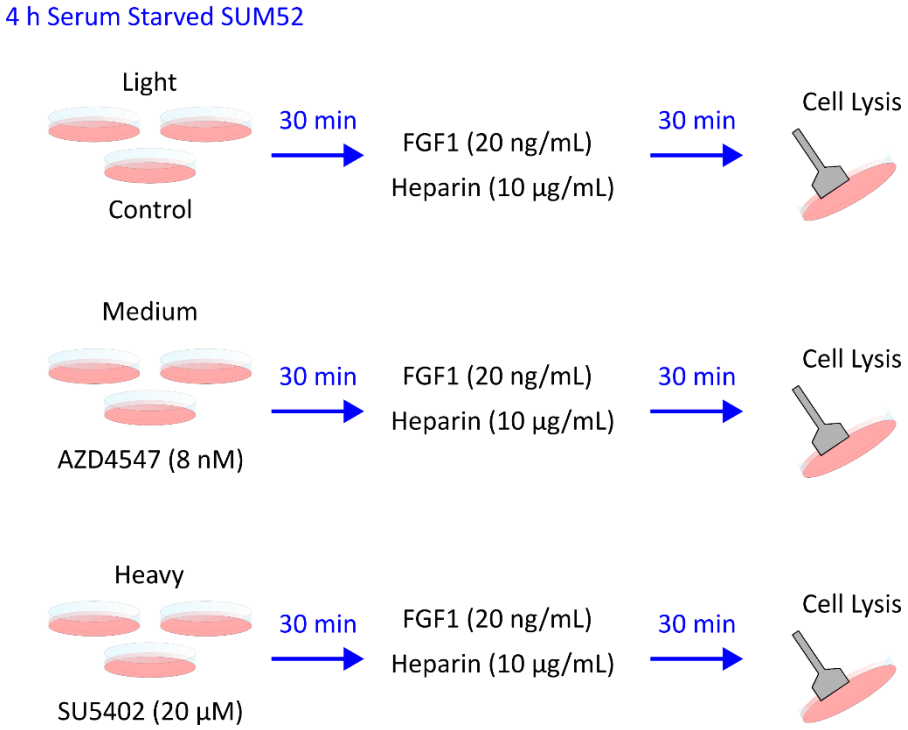


Figure 4.5 | Workflow for SUM52 SILAC cell treatment. After cell lysis, protein concentrations were acquired by Bradford assay before equal amounts of protein content from each quantitative channel were pooled for sample preparation and SILAC LC-FAIMS-MS/MS proteomics.

4.2.4. SILAC LC-FAIMS-MS/MS Revealed Differential Global Proteome Profiles for AZD4547 vs SU5402 FGFR Inhibitor Drugs in SUM52 Cells.

4.2.4.1. AZD4547 and SU5402 stimulated differential global proteome abundance profiles in SUM52 cells

Initially, the differential abundance of proteins from the external stepping LC-FAIMS-MS/MS dataset was investigated to determine global changes in proteome profiles. Previous work by Coon and co-workers[133], as well as the data presented in Chapter 3, have demonstrated the capacity of LC-FAIMS-MS/MS to increase the number of peptide and protein identifications. Thus, investigation of how the abundance of SUM52 proteins were modified by inhibition of FGFR signalling would provide context for the

differential PTM patterns that would then be studied. Overall, 1152 proteins were identified from external stepping LC-FAIMS-MS/MS experiments. For the AZD4547 condition corresponding to upregulated proteins, 29 proteins had a log2 fold change of 1.5 or greater, with 65 proteins being upregulated with a p-value of less than or equal to 0.05. For the AZD4547 condition corresponding to downregulated proteins, 105 proteins had a log2 fold change of 0.5 or lower, with 148 proteins being downregulated with a p-value of less than or equal to 0.05. For the SU5402 condition corresponding to upregulated proteins, 26 proteins had a log2 fold change of 1.5 or greater, with 63 proteins being upregulated with a p-value of less than or equal to 0.05. For the SU5402 condition corresponding to downregulated proteins, 86 proteins had a log2 fold change of 0.5 or lower, with 139 proteins being downregulated with a p-value of less than or equal to 0.05 (Figure 4.6). Principal Component Analysis, of data acquired from non-enriched external stepping LC-FAIMS-MS/MS, revealed differential changes in the global proteome abundance profiles of AZD4547 and SU5402 FGFR inhibitor administered SUM52 samples relative to the FGFR inhibitor-free control (Figure 4.5). The total change in the abundance of proteins across the proteome was evident when comparing the FGFR inhibitor samples to the drug free sample, as expected based on previous SUM52-FGFR inhibitor experiments[177,360] (Figure 4.6c). Differential proteome abundance profiles for each FGFR inhibitor drug suggested distinct changes in global protein abundances between AZD4547 and SU5402 administration. Taken together with the findings of higher FGFR2 inhibitory activity of SU5402 relative to AZD4547 at the concentrations used (Figure 4.4b), SUM52 proteins responded differently for the two drugs.

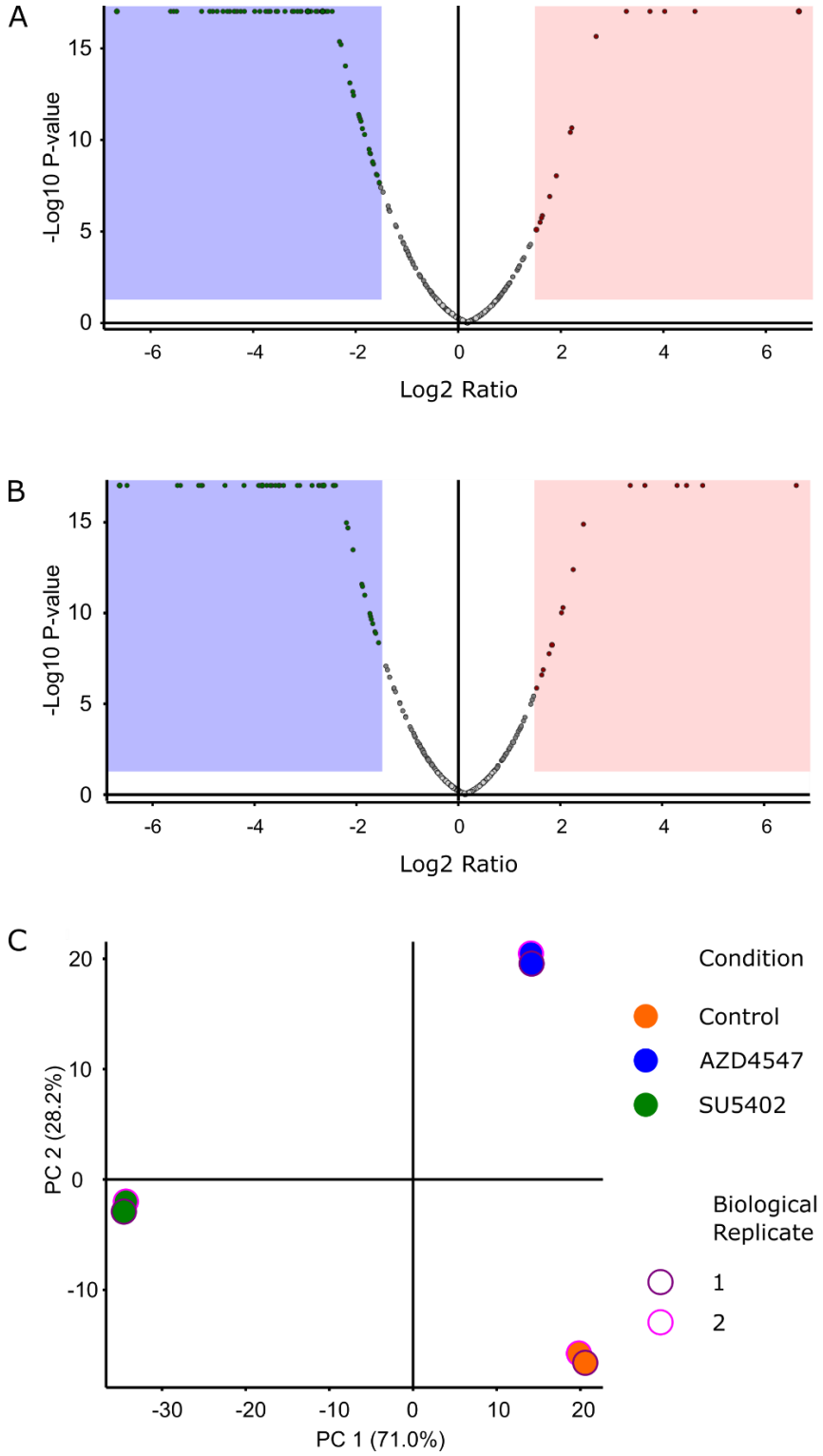


Figure 4.6 | Differential protein abundance profiles upon FGFR inhibition. (A) Volcano plot corresponding to upregulated and down regulated proteins upon FGFR inhibition with AZD4547. Log2 ratio of 1.5 or greater are within red (upregulated) and blue (downregulated) sections. (B) Same as (A) but corresponding to SU5402. (C) Principal Component Analysis of protein abundance profiles for control, AZD4547 and SU5402, with both biological replicates for each SILAC quantitation channel included.

4.2.4.2. FGFR inhibitors revealed downregulation of cytoskeletal proteins

Whilst the differential response of SUM52 to the two FGFR inhibitors is of clinical and biochemical interest, Section 4.2.4 and 4.2.5 focus specifically on the analysis of differentially regulated protein abundances with both drugs compared to the SUM52 control, as such proteins (whose abundance has been significantly modified by both FGFR inhibitors) were more likely to be affected by hyperactive FGFR kinase signalling in SUM52 breast cancer cells, promoting the biological capabilities that are acquired for the cancer phenotype (as characterised by the hallmarks of cancer[353]). Gene ontology enrichment analysis was performed on proteins identified from external stepping LC-FAIMS-MS/MS of non-enriched sample (sample preparation workflow illustrated in Figure 4.5). Proteins whose abundance was downregulated by both FGFR inhibitors enriched for cytoskeletal regulation and pro-inflammatory Receptors for Advanced Glycation Endproducts (RAGE) (although just 3 proteins were associated with the RAGE binding molecular function) (Table 4.1). These findings suggested a change in cytoskeletal architecture upon inhibition of the FGFR signalling pathway. For example, the abundances of titin, myosin, kinesin, dynamin, catenin and gelsolin were all downregulated upon the addition of both AZD4547 and SU5402. All gene ontology data in this thesis satisfied an adjusted p-value of 0.05 as described in Raudvere, U *et al*[304].

Table 4.1 | Gene ontology enrichment analysis of downregulated proteins upon AZD4547 and SU5402 administration. Analysis performed using g:profiler[304]. For specific protein identifications, see Appendix Section 8.5.

Category	Description	Number of proteins
GO Molecular Function	Cytoskeletal protein binding	15
GO Molecular Function	Actin filament binding	7
GO Molecular Function	RAGE receptor binding	3

4.2.4.3. FGFR inhibitors revealed upregulation of ribosomal and RNA related proteins

The inhibition of FGFR kinase activity stimulated an increase in the abundance of ribosomal proteins and RNA related proteins (Table 4.2). These findings were indicative of an increase in the activity of the cell's protein translation machinery, with the upregulated biosynthesis of 40S and 60S ribosomal subunits, translation initiation factors and tRNA ligases (including threonine and histidine). FGFR signalling is linked to upregulated ribosomal protein synthesis for sustained proliferation[361]. However, the upregulation of these proteins upon FGFR inhibition suggested a more nuanced relationship between FGFR and ribosomal protein abundance, as the inhibition of FGFR alone did not supress the synthesis of protein translation machinery, instead upregulated these components.

Table 4.2 | Gene ontology enrichment analysis of upregulated proteins upon AZD4547 and SU5402 administration. Analysis performed using g:profiler[304].

Category	Description	Number of proteins
GO Molecular Function	RNA Binding	17
GO Molecular Function	Structural Constituent of Ribosome	11

Ultimately, gene ontology enrichment analysis of significantly upregulated (normalised abundance ratio of ≥ 1.5) and downregulated (normalised protein abundance ratio of ≤ 0.67) protein abundances upon FGFR inhibition revealed a change in the cellular proteome away from the generation of a dynamic cytoskeletal architecture as has been reported in previous work[362,363]. Here, upregulation was defined as proteins with a foldchange upon addition of a drug (relative to control) of 1.5, satisfying the p-value of

≤ 0.05 (based on 3 biological replicates). FGFR inhibition also appeared to increase ribosomal protein synthesis, in contradiction of literature reports directly linking FGFR signalling to increased ribosome synthesis[364,365].

4.2.5. SILAC LC-FAIMS-MS/MS Revealed Changes in the Combinatorial PTM Status of Proteins Upon FGFR Inhibition

4.2.5.1. Multiple-PTM regulation of proteins were associated with the cancer phenotype in SUM52 cells

As demonstrated in Chapter 3, LC-FAIMS-MS/MS can be used to improve the identification of multi-PTM containing peptides that may be considered as candidates for positive PTM crosstalk[115]. Further to this, multiple peptide isoforms with differential PTM site localisations can be isolated and characterised[115]. External stepping LC-FAIMS-MS/MS and TiO_2-phosphoenriched internal stepping LC-FAIMS-MS/MS datasets were combined to generate a comprehensive representation of PTM dynamics to increase multi-PTM peptide identifications. In applying this technique to SUM52 breast cancer cells, with SILAC quantitation to distinguish changes in multi-PTM status upon the inhibition of FGFR signalling; the combinatorial PTM regulation of SUM52 cells towards a cancer phenotype (as described by the hallmarks of cancer[353]) was probed. FGFR inhibition directed the cellular phenotype away from migration and invasion as seen by the downregulation of proteins involved in cytoskeletal architecture (Table 4.1), corroborating the drugs' role in inhibiting cell migration within breast cancer[366], colon cancer[367] and pancreatic cancer[368] as well as the previously reported anti-cell migratory effect of AZD4547[369] and SU5402[370] specifically. Both drugs also inhibited such biological processes as proliferation, metabolism, and survival. Thus,

combinatorial PTMs that were abrogated by both drugs could be involved in PTM regulation of proteins towards the cell migration phenotype, including tumour metastasis. However, from a global stand-point; these PTMs may also relate to the regulation of proliferation, metabolism and survival signalling pathways[177,350]. Whilst 31 and 44 multi-PTM peptides as candidates for positive PTM crosstalk were upregulated for AZD4547 and SU5402, respectively, 221 and 202 multi-PTM peptides as candidates for positive PTM crosstalk were downregulated, respectively, with PTM combinations identified across the PTMs: phosphorylation (S, T, Y), acetylation (K, R), methylation (K, R), ubiquitination (K) and nitrosylation (C, Y) (Figure 4.7). Specifically, the downregulation of many PTMs suggested their significance for generating the cancer phenotype including cell motility, survival, proliferation, and metabolic rewiring, with stimulation of this behaviour initiated by FGFR signalling[177,350].

A AZD4547

	Phospho		Acetyl		Methyl		GG		Nitrosylation	
Phospho	0	0	5		5		3		1	
Acetyl	5		41	5	9		9		3	
Methyl	6		48		10	5	5		6	
GG	18		56		38		24	3	4	
Nitrosylation	10		27		30		19		5	0

B SU5402

	Phospho		Acetyl		Methyl		GG		Nitrosylation	
Phospho	3	1	5		1		2		1	
Acetyl	20		42	5	5		6		4	
Methyl	21		49		27	2	2		10	
GG	21		57		38		25	0	2	
Nitrosylation	11		28		32		21		5	0

Upregulated

Downregulated

Figure 4.7 | Number of multi-PTM peptides with a significant change in abundance upon FGFR inhibition with AZD4547 (A) and SU5402 (B). Upregulated multi-PTM containing peptides are coloured in orange/red, with a colour gradient representing the amount of multi-PTM peptide identifications for each specific PTM combination relative to the rest of the upregulated multi-PTM combinations (darker colour correlates to more identifications). Downregulated multi-PTM containing peptides are coloured in blue, with a colour gradient representing the amount of multi-PTM peptide identifications for each specific PTM combination relative to the rest of the downregulated multi-PTM peptide combinations (darker blue correlates to more identifications).

4.2.5.2. Combinatorial PTM dynamics were associated with changes in cytoskeleton, mitosis, and protein synthesis machinery

Proteins that derived from multi-PTM containing peptides, whereby the multi-PTM peptide status was significantly modified by both FGFR inhibitors, were grouped and gene ontology analysis revealed enrichment for cytoskeletal regulation, mitosis, and the protein synthesis machinery (Table 4.3). Regulation of cytoskeletal architecture is required for the cancer cell's ability to migrate in a directional manner across tissue via

dynamic changes in actin-based lamellipodia, invadopodia and filopodia[371]. These structures are dynamically synthesised and broken down through interactions with accessory proteins[372], changes in transcriptional mRNA splicing[373] and post-translation modification of both the F-actin (filamentous-actin) polymers, G-actin (globular actin) monomers, and their regulatory accessory proteins[374]. The abundance of 69 PTM sites on proteins involved in cytoskeletal regulation were significantly modified with FGFR inhibition. Of these sites, 37 have not previously been reported, with 28 occurring simultaneously with another PTM in a multiple PTM containing peptide. Such dynamic modification upon FGFR inhibition suggested the orchestrated and combinatorial regulation of cellular cytoskeletal dynamics to facilitate cell migration, stimulated from active FGFR signalling.

Table 4.3 | Gene ontology enrichment analysis of proteins derived from upregulated and downregulated multi-PTM containing peptides with AZD4547 and SU5402. Analysis performed using g:profiler[304]. All GO groups represent statistically significant ($p \leq 0.05$) protein networks within the overall group of upregulated and downregulated proteins.

Upregulated		
Category	**Description**	**Number of proteins**
GO Biological Process	cell cycle process	20
GO Biological Process	cellular macromolecule biosynthetic process	18
GO Biological Process	cellular amide metabolic process	18
GO Biological Process	peptide metabolic process	17
GO Biological Process	amide biosynthetic process	16
GO Biological Process	mitotic cell cycle	16
GO Biological Process	translation	15
GO Biological Process	peptide biosynthetic process	15
GO Biological Process	mitotic cell cycle process	15
GO Molecular Function	cell adhesion molecule binding	14
GO Molecular Function	cadherin binding	13
GO Biological Process	regulation of RNA splicing	9
GO Biological Process	regulation of mRNA splicing, via spliceosome	7
GO Biological Process	regulation of mRNA processing	7

GO Biological Process	chromosome condensation	5
Downregulated		
Category	**Description**	**Number of proteins**
GO Biological Process	cytoskeleton organization	33
GO Biological Process	cellular response to stress	33
GO Biological Process	regulation of localisation	33
GO Biological Process	cell cycle	32
GO Biological Process	establishment of protein localisation	31
GO Biological Process	cellular amide metabolic process	30
GO Biological Process	cellular macromolecule biosynthetic process	29
GO Molecular Function	cytoskeletal protein binding	27
GO Biological Process	peptide metabolic process	27
GO Biological Process	amide biosynthetic process	26
GO Biological Process	cell cycle process	25
GO Biological Process	translation	24
GO Biological Process	peptide biosynthetic process	24
GO Biological Process	regulation of organelle organization	24
GO Molecular Function	cell adhesion molecule binding	23
GO Molecular Function	structural molecule activity	23
GO Biological Process	positive regulation of cellular component organization	23
GO Biological Process	actin filament-based process	22
GO Biological Process	mitotic cell cycle	21
GO Molecular Function	cadherin binding	20
GO Molecular Function	actin binding	20
GO Biological Process	actin cytoskeleton organization	20
GO Biological Process	chromosome organization	19
GO Biological Process	mitotic cell cycle process	19
GO Molecular Function	actin filament binding	17
GO Molecular Function	ATP-dependent activity	17
GO Biological Process	positive regulation of organelle organization	16
GO Biological Process	muscle system process	15
GO Biological Process	actin filament organization	15
GO Biological Process	muscle contraction	13
GO Biological Process	regulation of cellular component size	12
GO Biological Process	protein folding	11

GO Biological Process	RNA localisation	9
GO Biological Process	telomere maintenance	8
GO Biological Process	cytoplasmic translation	8
GO Biological Process	multicellular organismal movement	6
GO Biological Process	musculoskeletal movement	6
GO Biological Process	RNA-templated DNA biosynthetic process	6

4.2.5.2.1. Cellular mitosis machinery related proteins underwent combinatorial-PTM changes with FGFR inhibitors

Proteins relating to mitosis were not enriched in the datasets for upregulated and downregulated protein abundances upon FGFR inhibition. However, many proteins relating to the mitosis biological process exhibited a change in their combinatorial PTM status upon the inhibition of the FGFR signalling pathway by either SU5402 or AZD4547 (or both drugs) (Figure 4.8). SUM52 cells display a proliferation phenotype, and FGFR inhibitors have been shown to cause G1 growth arrest, preventing SUM52 cells from reaching the mitosis stage of the cell cycle[360]. The structural maintenance of chromosomes family proteins: SMC1A and SMC4, condensing complex subunit 3, DNA mismatch repair protein Msh6 as well as multiple tubulin microtubule related proteins were found to be differentially PTM regulated in a combinatorial manner upon addition of both FGFR inhibitors). Of all mitosis related proteins derived from multi-PTM peptides, 71% involved downregulation of the multi-PTM containing peptide upon inhibition of the FGFR signalling pathway. Further, 85 % of these proteins derived from at least two distinct multi-PTM containing peptides mapped across their amino acid sequence (Table 4.4). These data indicate the orchestrated regulation of multiple PTMs on cell cycle related proteins that may be associated with the SUM52 mitosis driven

proliferation phenotype that is stimulated by FGFR kinase. As such, the abrogation of many PTMs that occurred concomitantly within a protein was indicative of reduced cell division activity upon FGFR inhibition, despite no significant change in the abundance of cell cycle related proteins with FGFR inhibition.

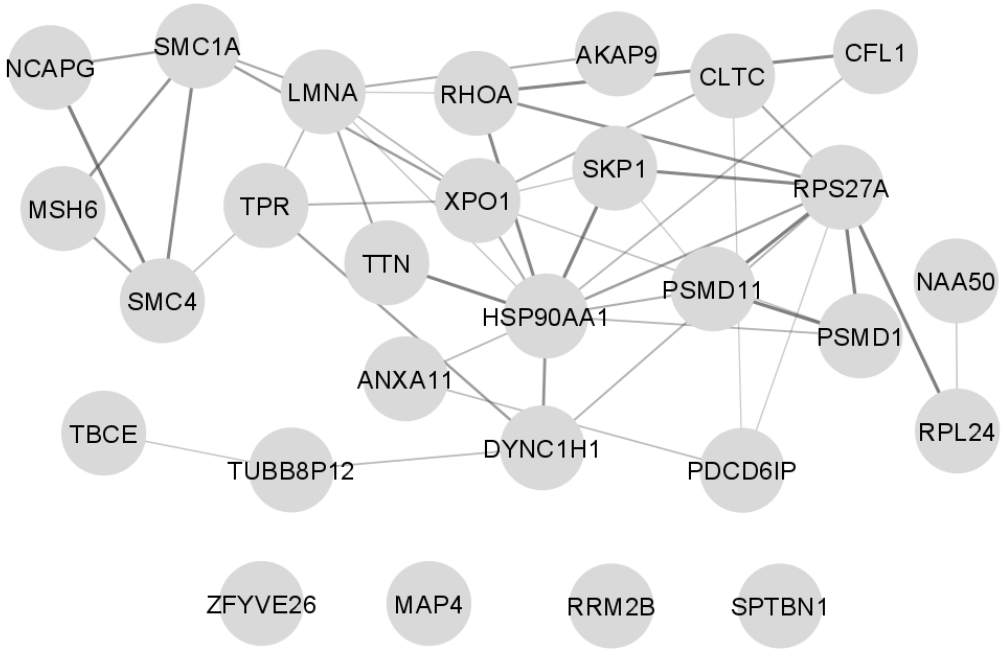


Figure 4.8 | Network map of mitosis-related proteins with changes in their multi-PTM status upon FGFR inhibition. Node codes represent proteins as shown in Table 4.4 (below). Network edges represent interactions as determined by STRING, with darker edge colour representing more interactions. Analysis performed using STRING v11.5, illustration designed in Cytoscape v3.9.1.

Table 4.4 | Mitosis-related proteins, as identified from gene ontology enrichment of mitosis related biological process, that were derived from multi-PTM containing peptides whose abundance was significantly modified with FGFR inhibition. Number of distinct multi-PTM containing peptides corresponding to each protein is also reported. Protein codes relate to Figure 4.8.

Protein	Protein Code	Number of FGFR inhibitor regulated multi-PTM peptides
Titin	TTN	35
Nucleoprotein TPR	TPR	8
Spectrin beta chain, non-erythrocytic 1	SPTBN1	6
Structural maintenance of chromosomes protein 4	SMC4	6
Zinc finger FYVE domain-containing protein 26	ZFYVE26	6
Structural maintenance of chromosomes protein 1A	SMC1A	4
Tubulin-specific chaperone E	TBCE	4
26S proteasome non-ATPase regulatory subunit 1	PSMD1	3
Cofilin-1	CFN1	3
26S proteasome non-ATPase regulatory subunit 1	PSMD1	2
A-kinase anchor protein 9	AKAP9	2
Annexin A11	ANXA11	2
Clathrin heavy chain 1	CLTC	2
Condensin complex subunit 3	NCAPG	2
DNA mismatch repair protein Msh6	MSH6	2
Heat shock protein HSP 90-alpha	HSP90AA1	2
Microtubule-associated protein 4	MAP4	2
N-alpha-acetyltransferase 50	NAA50	2
Programmed cell death 6-interacting protein	PDCD6IP	2
S-phase kinase-associated protein 1	SKP1	2
Transforming protein RhoA	RHOA	2
Tubulin beta 8B	TUBB8P12	2
Ubiquitin-40S ribosomal protein S27a	RPS27A	2

26S proteasome non-ATPase regulatory subunit 11	PSMD11	1
60S ribosomal protein L24	RPL24	1
Exportin-1	XPO1	1
Prelamin-A/C	LMNA	1

4.2.5.2.2. Protein synthesis machinery related proteins underwent combinatorial-PTM changes with FGFR inhibition

As discussed above, SILAC LC-FAIMS-MS/MS analysis of the SUM52 cancer cell line suggested a decrease in mitotic cell activity upon inhibition of the FGFR kinase signalling pathway, corroborating previous work[360]. For a cell to undergo rapid proliferation such as in cancer, the G1 and S phases of the cell cycle must generate the required protein mass for complete duplication into two daughter cells. Therefore, the cancer phenotype is often associated with an increased rate of protein synthesis[375]. Ribosomal proteins are known targets of the RAF/MEK/ERK signalling pathway downstream of FGFR kinase, via ribosomal protein S6 kinase[376]. Thus, the modified combinatorial PTM status of six 60S ribosomal subunits (60S Ribosomal protein-L27a, -L10, -L34, -L5, -L6 and -L7a) upon FGFR inhibition within this dataset was consistent with previous findings[177,375]. Further to this, tRNA ligases play a key role in charging tRNA molecules before peptide bond formation inside the ribosomal complex[377]. tRNA ligases: Leucine-, Valine-, Phenylalanine-, Bifunctional glutamate/proline-, Lysine- and tRNA (guanine(10)-N2)-methyltransferase homolog all underwent significant changes in their multi-PTM status upon inhibition of FGFR signalling, with the majority of these multi-PTM peptides downregulated (Figure 4.9, Table 4.5). FGFR activity is linked to ribosome biogenesis[364], implicating FGFR inhibition in downregulating this process and yielding reduced ribosomal protein abundance. Added to this, AZD4547 and SU5402

have both been associated with reduced protein synthesis via the mammalian target of rapamycin (mTOR) pathway[177,265,378]. However, protein abundance related changes towards reduced protein synthesis were not identified in the investigation of protein abundances upon addition of FGFR inhibitors. On the contrary, ribosomal proteins and RNA binding proteins were found to be upregulated with FGFR inhibition (Section 4.2.4.3). This may suggest an orchestrated role of multiple-PTM regulation for the downregulation of protein synthesis activity, without relative decreases in protein abundance. Alternatively, the changes in combinatorial PTM status on protein translation related proteins that were identified after FGFR inhibition could stimulate downregulation in their abundance, with these effects only detectable beyond the incubation time of the drugs in these experiments (1 hour). This demonstrates the significance of combinatorial PTMs for the regulation of ribosomal protein synthesis in the metabolic context of cancer.

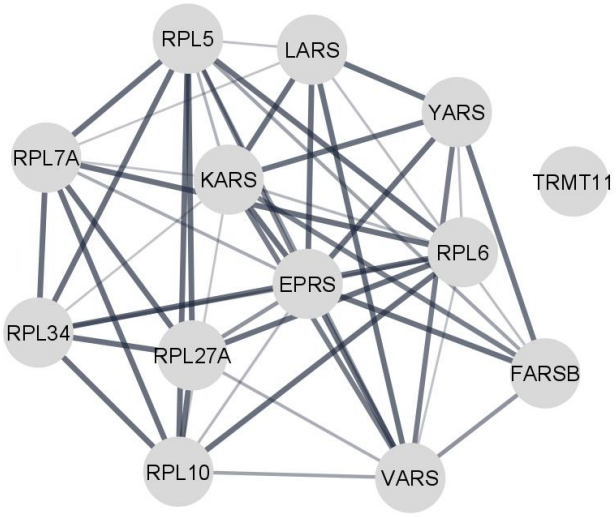


Figure 4.9 | Network map of protein synthesis-related proteins with changes in their multi-PTM status upon FGFR inhibition. Node codes represent proteins as shown in Table 5 (below). Network edges represent interactions as determined by STRING, with darker edge colour representing more interactions. Analysis performed using STRING v11.5, illustration designed in Cytoscape v3.9.1.

Table 4.5 | Protein synthesis related proteins, as identified from gene ontology enrichment of all protein synthesis related biological process, derived from multi-PTM containing peptides whose abundance was significantly modified upon FGFR signalling. Number of distinct multi-PTM containing peptides corresponding to each protein is also reported. Analysis performed using g:profiler[304]. Protein codes related to Figure 4.9.

Protein	Protein Code	Regulated multi-PTM peptides
Leucine--tRNA ligase cytoplasmic	LARS	7
Valine--tRNA ligase	VARS	5
60S ribosomal protein L27a	RPL27a	3
Phenylalanine--tRNA ligase beta subunit	FARSB	3
60S ribosomal protein L10	RPL10	2
60S ribosomal protein L34	RPL34	2
60S ribosomal protein L5	RPL5	2
60S ribosomal protein L6	RPL6	2
60S ribosomal protein L7a	RPL7a	2
Bifunctional glutamate/proline--tRNA ligase	EPRS	2
Lysine--tRNA ligase	KARS	2
tRNA (guanine(10)-N2)-methyltransferase homolog	TRMT11	2
Tyrosine--tRNA ligase, cytoplasmic	YARS	2

4.2.6. Actin Cytoskeletal Regulation may be Orchestrated by Dynamic PTMs

The SUM52 cell migratory phenotype is dictated by highly orchestrated regulatory systems[379]. For example. complex signalling pathways from transmembrane receptors such as G-Protein Coupled Receptors (GPCRs)[380], integrins[381], and receptor tyrosine kinases[381] (including FGFR) stimulate actin polymerisation, microtubule assembly and cell-cell junction development[382]. Further, mRNA splice variations translate to protein isoform variants which, intern, can form heteromeric complexes, modulating protein behaviour[373]. Added to this, multiple accessory proteins play a critical role in mediating cytoskeletal development through scaffold formation and recruitment or inhibition of other proteins, with accessory protein isoform variability imparting further regulation[383]. Multiple layers of highly sophisticated regulation fine-tune the cell's ability to

dynamically modify its morphology as it establishes a motile, migratory phenotype. However, conflicting data suggest more, as yet unidentified, regulatory mechanisms are involved in facilitating the dynamic behaviour of the cytoskeleton[384]. For example, profilin, an actin accessory protein, has demonstrated actin polymerisation and depolymerisation activity in different cellular contexts, with little known about the contributory factors towards this differential behaviour[385]. Further, another actin depolymerising accessory protein; gelsolin, has been identified as a key facilitator of cell motility, which contradicts data that demonstrates its redundancy in normal mouse developmental processes[386,387]. As such, more accurate characterisation of the proteoforms that these proteins can access in different physiological environments is necessary to reconcile these ambiguities.

4.2.6.1. Combinatorial PTM regulation of myosins and tropomyosins were modified by FGFR inhibition

Actin dependant molecular motors, myosins, are important for the regulation of cancer cell migration and invasion as they crosslink F-actin to provide force, powering cell motility (amongst other functions)[388]. Tropomyosins represent a class of regulatory proteins that bind actin and mediate its association with myosin as well as other accessory proteins[389]. The inhibition of the FGFR kinase signalling pathway resulted in the downregulation in abundance, accompanied by changes in the combinatorial PTM status, of myosin-2, -9, -IIIb, -13, -14 and -15. Further to this, FGFR inhibition resulted in changes in the combinatorial PTM status of tropomyosins α-1 chain, α-3 chain, and α-4 chain (without significant changes in protein abundance of α-1 or α-3). These data suggested the suppression of the cells ability to undergo changes in morphology as a

prerequisite to cell migration (as a known phenotype of FGFR hyperactive cells[256]) with PTMs playing a key role in their downregulated activity upon FGFR inhibition. Thus, the combinatorial impact of these PTMs on the activity of myosins and tropomyosins may be relevant to the cancer cell migratory phenotype. For example, myosin-2 promotes directional migration by aiding cytoskeletal lamellipodia formation towards the cells leading edge and promoting tail retraction[390]. Within the SUM52 control SILAC condition, novel methylations at myosin-2 lysine-1079 and lysine-1422 as well as novel nitrosylation at C541 were identified (Table 4.6). The methylations exclusively occurred concurrently, indicating possible crosstalk between these sites. Further, phosphorylation at threonine-1423 has previously been reported[142], and this may occur in an alternative regulatory proteoform of the protein. Importantly, these identified PTMs were abrogated with FGFR inhibition by both AZD4547 and SU5402. Myosin-9 abundance was not significantly modified by either FGFR inhibitor, however both drugs stimulated serine-1943 phosphorylation (a previously reported PTM[218,391,392]). Whilst only tropomyosin α-4 chain protein abundance was significantly modified by FGFR inhibitors, with its downregulation with AZD4547 and upregulation with SU5402, 15 unique responses of specific PTMs were observed across the three tropomyosins with FGFR inhibition, indicating heavy multiple-PTM regulation of this actin regulatory protein family (Table 4.6). Whilst 14 of the 15 identified PTMs have previously been reported[142] (tyrosine-215 nitrosylation being the exception, with differential regulation between AZD4547 and SU5402), the significant changes in abundance of these PTMs upon FGFR inhibition could provide valuable information to the regulatory framework of tropomyosin modulation in the context of FGFR stimulated actin polymerisation for lamellipodia

formation. Further, tropomyosin α-1 chain and α-4 chain are known biomarkers for gastric and colon cancer, respectively[393,394].

Table 4.6 | Changes in protein abundance and PTM status of myosins and tropomyosins upon FGFR inhibition with AZD4547 and SU5402. MS/MS spectra of all peptides contributing to this table were manually validated.

Protein Abundance
No Change
Downregulated (both drugs)
Downregulated (AZD4547) Up regulated (SU5402)
Downregulated (AZD4547) No change (SU5402)

Peptide Abundance
No Change
Downregulated
Upregulated

	Phospho		Acetyl		Methyl		Ubiquitin		Nitro	
	AZD4547	SU5402	AZD4547	SU5402	AZD4547	SU5402	AZD4547	SU5402	AZD4547	SU5402
Tropomyosin alpha-1 chain			K30		K251		K37			
Tropomyosin alpha-3 chain					K218	K162, R168, K218				
Tropomyosin alpha-4 chain			K212, K223	K212, K223			K214	K214	Y215	Y215
Myosin-2					K1079, K1422		K228	K228		
Myosin-9	S1943						K1724	K1724	C541	
Myosin-IIIb					R551					
Myosin-13			K916		R18		K922, K924			
Myosin-14	S1969, S1989				R826					
Myosin-15			K1782, K1783				K1769			

4.2.6.2. Combinatorial PTM regulation of actin accessory proteins were modified by FGFR inhibitors

4.2.6.2.1. Arp2/3 differential acetylation may regulate lamellipodia formation

Cytoskeletal accessory proteins play a significant role in the regulation of cytoskeletal dynamics. For example, the Arp2/3 protein dimer forms the Arp2/3 complex with nucleation promoting factors to generate branched actin polymers during lamellipodia formation for cell migration[384,395,396]. These structures, in coordination with myosin-2, generate cell cortex contractility to drive leading edge extension in cancer cell migration[397]. SILAC LC-FAIMS-MS/MS identified 22 FGFR inhibitor sensitive PTM sites across 9 actin related proteins (Table 4.7). Protein abundance corresponding to subunit 2 of the Arp2/3 heterodimer was downregulated by both FGFR inhibitors whereas subunit 3 was upregulated by both FGFR inhibitors, demonstrating stoichiometric changes in the abundance of complex subunit ratios for the Arp2/3 dimer in response to FGFR inhibition. Although no combinatorial PTM regulation was identified in the Arp2/3 heterodimer, Arp2/3 protein complex subunit 2 lysine-265 underwent significantly increased acetylation upon stimulation with both FGFR inhibitors, with this modification not previously reported in curated PTM databases[21,141,142]. Whilst growth factor receptor signalling is known to stimulate the formation of branched actin polymers for lamellipodia, inhibition of EGFRs has been shown to block such development[398]. As such, Arp2/3 may represent an important modulator of the receptor tyrosine kinase stimulated lamellipodia formation, with inhibition of growth factor-based signalling facilitating changes within the acetylation status of Arp2.

4.2.6.2.2. PTM crosstalk may regulate actin-lamellipodia formation via cofilin

Another important actin regulatory protein, cofilin, is necessary for receptor tyrosine kinase stimulated formation of lamellipodia for cell migration[399]. Further, Arp2/3 and cofilin work in synergy to generate lamellipod protrusion[400]. Cofilin is known to bind actin polymers and stimulate depolymerisation, reducing F-actin stability[401]. Whilst cofilin-1 (present in non-muscle cells[401]) and cofilin-2 (dominates mature skeletal and cardiac muscles[401]) protein abundances did not change upon FGFR inhibition, cofilin-1 and cofilin-2 N-terminal methionine loss followed by acetylation (not previously reported) and serine-3 phosphorylation (previously reported[402]) occurred concurrently in the SUM52 control condition, and were downregulated upon FGFR inhibition with both drugs (Table 4.7). In cofilin-1, the peptide sequence: MASGVAVSDGVIK with methionine cleavage and N-terminal acetylation was detected as both phosphorylated (Figure 4.10b) and non-phosphorylated (Figure 4.10a) peptidoforms (serine-3 phosphorylation), and serine-3 phosphorylation was not detected without N-terminal acetylation of alanine following methionine cleavage. Both peptidoforms were abrogated by FGFR inhibition, suggesting the possibility of co-translational N-terminal methionine removal and acetylation to facilitate serine-3 phosphorylation of cofilin, stimulated from upstream FGFR signalling. LIM kinase phosphorylation of cofilin serine-3 inhibits cofilin's ability to bind and regulate actin, thus sequestering cofilin activity and stabilising F-actin polymers[402]. Conversely, slingshot phosphatase (SSH) can reactivate cofilin via dephosphorylation at serine-3[403,404]. This phosphorylation activity is believed to act as a switch, with non-phosphorylated cofilin binding actin and catalysing depolymerisation. Preference for the serine-3 phosphorylated proteoform in the SILAC control condition suggested that SUM52 cells may stimulate stable actin

polymerisation to generate leading edge lamellipodia via sequestration of cofilin. The downregulation of serine-3 phosphorylation (and N-terminal methionine-loss followed by acetylation) of cofilin-1 and cofilin-2 upon FGFR inhibition suggested a reduction in stable F-actin formation for lamellipodia formation and cell migration, possibly stimulated by the inhibition of upstream FGFR signalling. Further to the above findings relating to the combinatorial PTM-regulation of actin accessory proteins for the regulation of lamellipodia formation, other proteins including spectrin, α-actinin, stathmin, filamin-A and shroom-3 all underwent differential PTM regulation upon FGFR inhibition (Table 4.7).

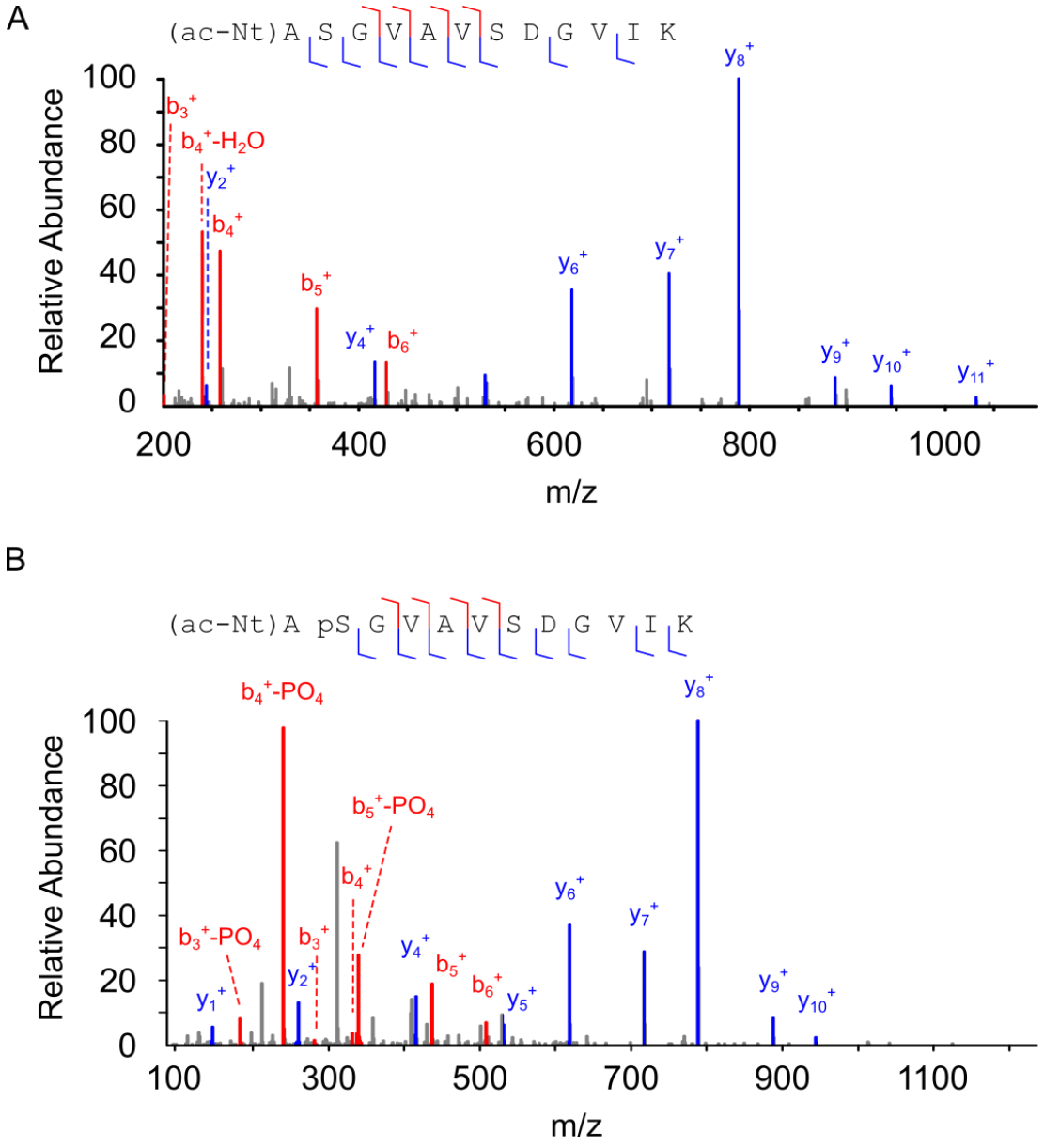


Figure 4.10 | Fragmentation mass spectra for peptides corresponding to cofilin-1. (A) MS/MS spectrum of N-terminally acetylated cofilin-1 (peptide: 2-13). (B) MS/MS spectrum of N-terminally acetylated and serine-3 phosphorylated cofilin-1 (peptide: 2-13). The detection of N-terminally acetylated peptidoform with and without phosphorylation together with no N-terminal acetylation independent phosphorylation suggests positive crosstalk between these modifications, with the acetylation possibly priming serine-3 phosphorylation.

Table 4.7 | Changes in protein abundance and PTM status of actin accessory proteins upon FGFR inhibition with AZD4547 and SU5402. MS/MS spectra of all peptides contributing to this table were manually validated.

Protein Abundance
No Change
Downregulated (both drugs)
Upregulated (both drugs)
Upregulated (AZD4547) Down regulated (SU5402)

Peptide Abundance
No Change
Downregulated
Upregulated

	Phospho		Acetyl		Methyl		Ubiquitin	
	AZD4547	SU5402	AZD4547	SU5402	AZD4547	SU5402	AZD4547	SU5402
Cofilin-1	S3		Nt					
Cofilin-2	S3		Nt					
Arp 2/3 complex subunit 2				K265				
Spectrin beta chain, non erythrocytic	T2337, T2242				R1675			
Alpha-actinin-4					R311, R319		K323	
Stathmin	S46, S16	S46, S38						
	S38	S16						
Filamin-A			K906					
LIM domain and actin binding protein 1	T487				R1937		K1801	
Protein shroom-3			K1809. K1928, R1936					

4.2.6.3. Combinatorial PTM regulation of signalling proteins upstream of actin cytoskeleton

4.2.6.3.1. 14-3-3 combinatorial PTM regulation may modulate actin polymerisation for lamellipodia formation

The post-translation modification of cofilin is regulated by SSH and LIM kinase as discussed above. FGFR signalling proteins upstream of these effectors also regulate the PTM status of cofilin, via SSH and/or LIM kinase. For example, the 14-3-3 phospho-binding protein, a global regulator of every major cellular function[405], can occur as seven isoforms. 14-3-3 ζ/δ isoform variant can bind serine-3 phosphorylated cofilin, preventing SSH phosphatase activity and consequently sequestering its actin depolymerisation activity[406]. Thus, the combinatorial regulatory PTM status of 14-3-3 isoforms could provide useful information in distinguishing signalling activity that leads to actin polymerisation, with phospho-acetyl crosstalk known to occur across human paralogs of 14-3-3[407]. 14-3-3 ζ/δ acetylation at lysine-3, a previously reported modification[408], was observed in the control SILAC condition and abrogated by both FGFR inhibitors. Thus, this acetylation may promote binding of 14-3-3 ζ/δ to phospho-cofilin, driving increased F-actin stability for lamellipodia formation. Added to this, 14-3-3 γ phosphorylation at threonine-139 (previously reported[409]), acetylation at arginine-4 and N-terminal methionine loss followed by acetylation were identified in the SUM52 control condition, with these PTMs also abrogated upon FGFR inhibition for both drugs (except arginine-4 acetylation, whose downregulation was observed specifically with SU5402). Further acetylation on 14-3-3 ε (arginine-4) and 14-3-3 β/α (N-terminally post-methionine cleavage) was observed, with N-terminal acetylation of 14-3-3 β/α downregulated with both FGFR inhibitors (both acetylation's not previously reported).

In summary, these 14-3-3 PTMs, whose occurrence was abrogated upon inhibition of FGFR kinase, could suggest the importance of these PTMs to facilitate the cancer phenotype of SUM52 cells, stimulated by hyperactive FGFR2. Further, considering the known regulation of cofilin by 14-3-3, the modified combinatorial PTM status of the multiple 14-3-3 isoforms upon FGFR inhibition may contribute to the complex regulation of lamellipodia formation and cell migration via modified sequestration of cofilin, especially considering that the majority of these combinatorial PTMs were not present upon inhibition of the FGFR signalling pathway.

4.2.6.3.2. Multiple signalling proteins upstream of actin polymerisation undergo changes in combinatorial PTM status.

Signalling proteins that converge at Rac GTPase (Rac) were found to be regulated by multiple PTMs. Rac GTPase is heavily involved in stimulating actin polymerisation[410] as phosphoinositide-3-kinase (PI3K) dependant Rac exchanger 1 functions as a Guanine Nucleotide Exchange Factor (GEF) for the activation of Rac GTPase, leading to downstream activation of serine/threonine protein kinase (PAK1), cortactin and filamin and thus promoting actin polymerisation and lamellipodia formation (Figure 4.11). PI3K dependant Rac exchanger 1 was acetylated at lysine-1214 (not previously reported), with this modification downregulated by both FGFR inhibitors (Table 4.8). PI3K dependant Rac exchanger 1 did not show any significant change in expression, perhaps revealing the importance of its PTM status in finetuning its behaviour. Further, another GEF that activates Rac, VAV2, was found to be ubiquitinated at lysine-764, with acetylation at either lysine-768 or lysine-770 occurring alongside it (all sites not previously reported) (Table 4.8). These modifications were all abrogated by both FGFR inhibitors. VAV2

abundance was also downregulated upon FGFR inhibition by both drugs. Further downstream, Rac GTPase localises serine-418 phosphorylated Src cortactin to the cell periphery where cortactin can then bind F-actin and activate Arp2/3 leading to lamellipodia formation and subsequent cell migration[411] (Figure 4.11). Src substrate cortactin phosphorylation at serine-418 was identified in the SILAC control cells, with this PTM representing a known positive regulatory modification by ERK1/2 down stream of FGFR kinase[412]. However, this phosphosite was not abrogated by FGFR inhibition via AZD4547 and was upregulated with SU5402. These data suggest a more complex regulatory role for serine-418 than promoting actin polymerisation in an isolated manner, perhaps involving other PTMs via a crosstalk mechanism. Another signalling protein upstream of cortactin, the multi-domain integrator of signalling pathways involved in cytoskeletal organisation, GIT1[413], undergoes highly dynamic modification with 68 PTMs previously reported[142]. This complex combinatorial PTM regulation may facilitate the promiscuous binding of GIT1 to many signalling proteins as well as its context dependant localisation[413]. LC-FAIMS-MS/MS analysis identified two multi-PTM peptides, each containing two phosphosites at: serine-385, serine-388 and serine-592, serine-596 (all four phosphosites previously reported[142,392,414,415]). Whilst AZD4547 FGFR inhibition upregulated serine-592 and serine-596 phosphorylation, SU5402 down regulated these phosphosites. Further, SU5402 downregulated serine-385 and serine-388 phosphorylation, with no significant change in these phosphosites upon AZD4547 FGFR inhibition. These data demonstrate the highly complex combinatorial PTM regulation of GIT1 in relation to cytoskeletal development, with further identification of modifications likely required to characterise proteoform variability for

different interactions with other proteins, to gain a deeper understanding of GIT1 behaviour in the context of actin polymerisation for cell migration.

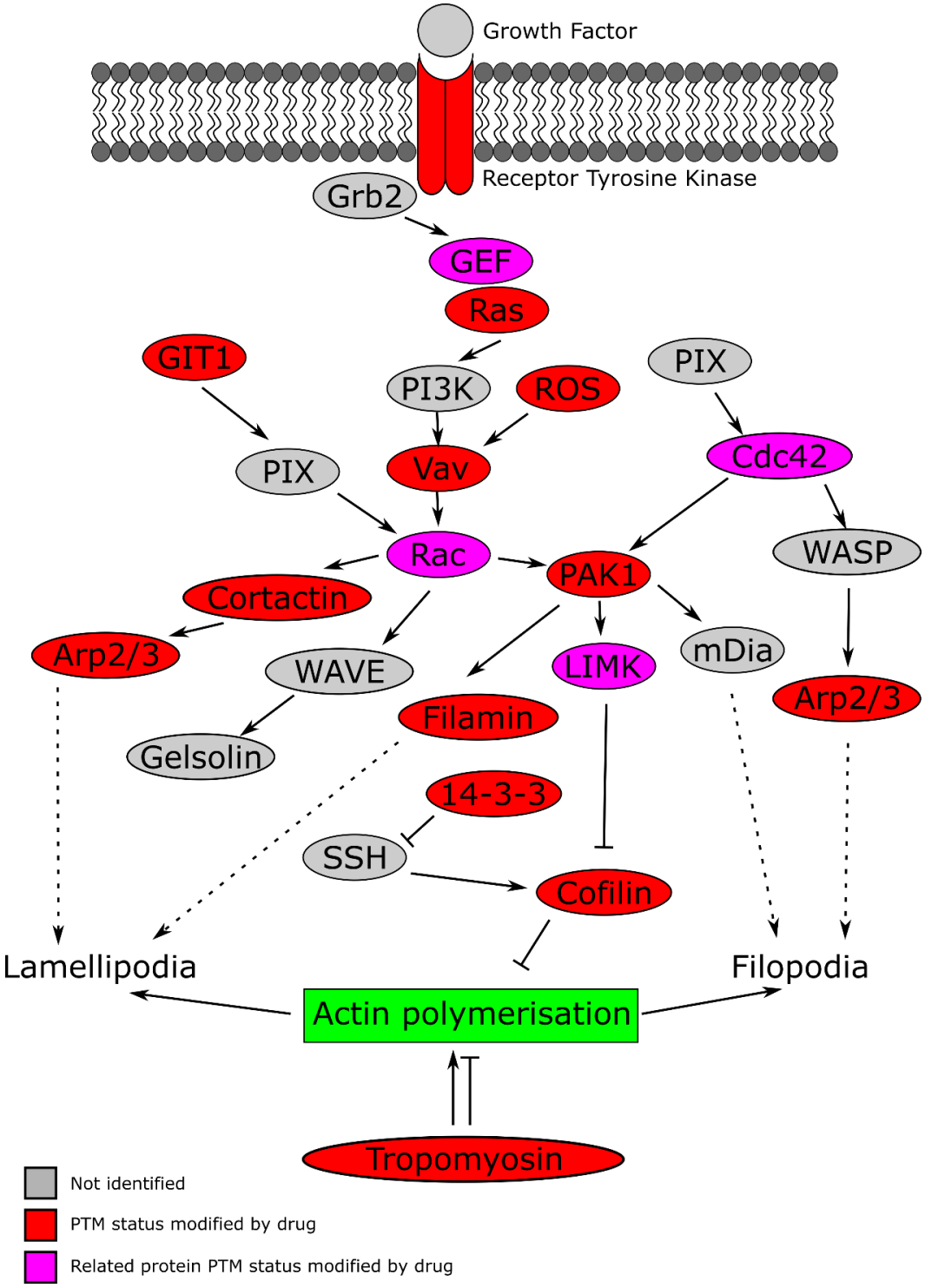


Figure 4.11 | Receptor tyrosine kinase stimulated signalling pathways for the formation of lamellipodia via actin polymerisation, leading to cell migration. Red represents proteins derived from peptides with a PTM status that was FGFR inhibitor sensitive. Pink represents the identification of peptides that correspond to a protein of related function to the displayed protein, with these related proteins demonstrating an FGFR inhibitor sensitive PTM status. Grey represents proteins that were not identified.

Protein Abundance
No Change
Downregulated (both drugs)

Peptide Abundance
No Change
Downregulated
Upregulated

	Phospho		Acetyl		Methyl		Ubiquitin		Nitro	
	AZD4547	SU5402	AZD4547	SU5402	AZD4547	SU5402	AZD4547	SU5402	AZD4547	SU5402
PAK1	S174									
Cdc42 effector protein-1	S350, S353									
	S292	S292								
PI-3,4,5-triphosphate-dependant Rac exhanger 1										
Vav2 GTP exchange factor (GEF)			K1214				K764			
Src substrate cortactin	S418	S418	K768, R770, K73							
LIM domain and actin binding protein 1	T487									
14-3-3 protein zeta/delta			K3							
14-3-3 protein gamma	T139		Nt				K142			
			R4	R4						
14-3-3 protein epsilon				R4						
14-3-3 protein beta/alpha			Nt							
proto-oncogene tyrosine-protein kinase ROS				R360						
Arf GTPase activating protein GIT1	S592, S596, S385, S388	S592, S596, S385, S388							C395	C395
Jupiter microtubule associated homolog 1	S87									

Table 4.8 | Changes in protein abundance and PTM status of cytoskeletal signalling proteins upon FGFR inhibition with AZD4547 and SU5402. MS/MS spectra of all peptides from which proteins in this table were derived were manually validated.

4.3. Conclusions

SILAC LC-FAIMS-MS/MS, developed in Chapter 3, was applied to SUM52 triple negative breast carcinoma cancer cells by combining data acquired from external stepping LC-FAIMS-MS/MS and internal stepping TiO_2-enriched LC-FAIMS-MS/MS. An improved understanding of how the SUM52 proteome undergoes combinatorial PTM regulation from hyperactive FGFR2 signalling through to the cancer phenotype was probed. Differential combinatorial PTM regulation upon FGFR kinase inhibition by two separate inhibitor drugs, AZD4547 and SU5402, demonstrated the relevance of highly orchestrated multi-PTM based regulation. FGFR inhibition yielded the reduced abundance of cytoskeletal proteins as well as an upregulation in the abundance of ribosomal proteins. From the global proteomic viewpoint, this suggested a change in cellular behaviour away from cytoskeletal development for cell migration towards increased protein synthesis upon inhibition of the FGFR signalling pathway. Analysis of changes in the combinatorial PTM status of proteins, stimulated by FGFR inhibition, elucidated drug-responsive modifications to proteins relating to mitosis, protein synthesis machinery and cytoskeletal dynamics, with all three biological processes representing characterised endpoints of FGFR signalling. Overall, 498 multi-PTM peptides that were significantly modified by FGFR inhibition were identified, 422 of these multi-PTM peptides were downregulated upon inhibition of the FGFR kinase signalling pathway. The development of actin lamellipodia is a key mechanism for cell migration, involving numerous actin-related proteins and signal transducers downstream of FGFR signalling. This work revealed changes in the combinatorial PTM status of many such proteins upon the inhibition of FGFR stimulation. For example, FGFR inhibition stimulated changes in the multi-PTM status of myosin and tropomyosin

isoforms, demonstrating the relevance of multiple novel PTMs in regulating their behaviour in the context of lamellipodia formation, stimulated by upstream FGFR signalling. Further, the F-actin destabilising protein: cofilin, underwent a previously unidentified N-terminal acetylation together with a negative regulatory serine-3 phosphorylation in SUM52 cells. FGFR inhibition abrogated both modifications, implicating their relevance in stimulating lamellipodia formation for cancer cell migration in FGFR amplified cell lines. The cofilin-sequestering protein: 14-3-3, was identified with phosphorylation, N-terminal, and lysine ε-amine acetylation across multiple protein isoforms in SUM52 cells, with these combinatorial PTMs downregulated upon inhibition of FGFR signalling. These findings implicate 14-3-3 regulation at an isoform and multi-PTM proteoform level in sequestering cofilin to promote F-actin stabilisation for lamellipodia development. FGFR signal transduction proteins including cortactin, Vav, Rac and GIT1 link cytoskeleton regulatory proteins to growth factor mediated stimulation. Novel multi-PTM sites were identified across these proteins, with such modifications sensitive to FGFR inhibition, again demonstrating the relevance of combinatorial PTM regulation to the SUM52 cell migratory phenotype. These findings demonstrate the potential of LC-FAIMS-MS/MS in identifying novel PTMs that that may be regulating protein systems in an orchestrated, combinatorial manner by enabling the proteins that they decorate to access different functional proteoforms to stimulate the cancer phenotype.

5. Multi-Modal Mass Spectrometry Investigation of S-Glutathionylation PTM in Cyanobacteria

5.1. Overview

Many PTMs have been identified in ancient organisms, with crosstalk mechanisms being conserved across evolution[22]. As techniques for the better characterisation of PTM variable proteoforms are developed, their fundamental role in the regulation of biochemistry will inform our understanding of the evolutionary development of such crucial biological processes as photosynthesis. Cyanobacteria represent the first organisms to develop oxygenic photosynthesis. Their ability to regulate their light harvesting protein machinery in response to environmental change may be a crucial component of their evolutionary success, leading to the great oxidation event and its profound consequences on biology[273]. The phycobilisome light harvesting protein complex is composed of stacked hexamers of phycocyanin protruding from three core allophycocyanin hexamer subunits[416] (Figure 5.1). To determine whether PTMs are present within the phycobilisome; a multi-modal mass spectrometry investigation was performed. This encompassed native, native top-down, LC-MS/MS and LC-FAIMS-MS/MS mass spectrometry techniques to focus on the identification and characterisation of a novel S-glutathionylation PTM on the phycobilisome. The data acquired from native, native top-down and LC-MS/MS experiments suggested a role of S-glutathionylation in the protection of the phycobilisome from the elevated levels of reactive oxygen species that are generated as by-products of photosynthesis. An LC-FAIMS-MS/MS workflow, used to enrich for S-glutathionylated tryptic peptides by exploiting their unique physiochemical properties, revealed the upregulation of S-glutathionylated proteoforms upon exposure of cyanobacteria to increased light intensity and hydrogen peroxide, with six of these S-glutathionylated proteoforms not previously reported. Such findings suggested more, yet uncharacterised roles for proteoform S-

glutathionylation within cyanobacteria, as ancient organisms from which important biological processes evolved. Further, the identification of phosphorylation occurrence concurrently with S-glutathionylation suggested positive PTM-crosstalk as an important framework for metabolic regulation of photosynthesis and downstream carbon fixing proteins within cyanobacteria.

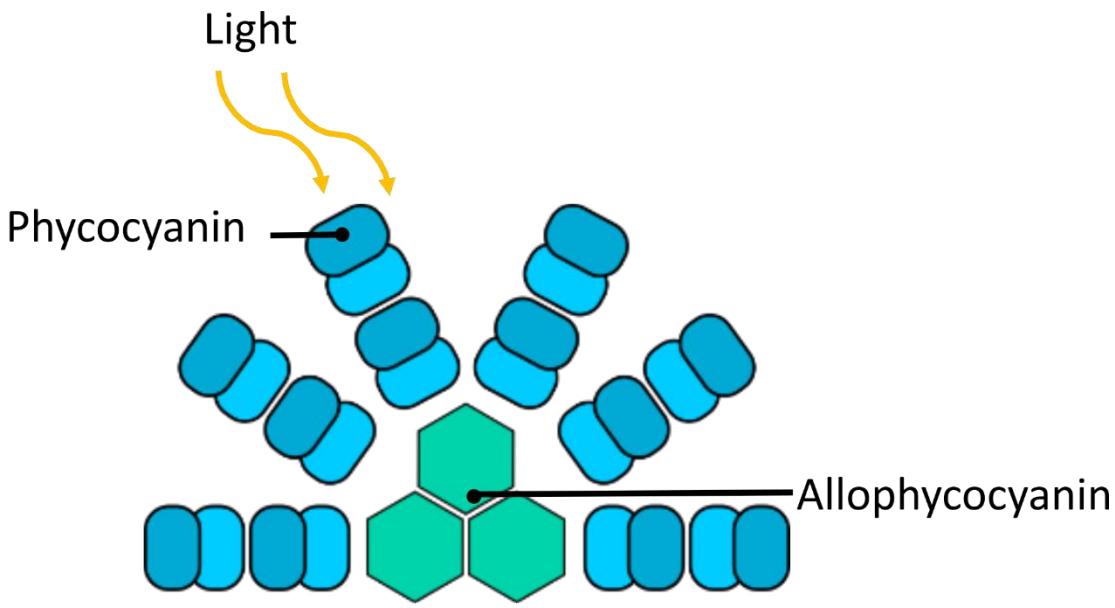


Figure 5.1 | Structure of cyanobacteria light harvesting protein complex: phycobilisome. Blue subunits each represent phycocyanin hexamer, which combine to form multiple dodecameric structures that protrude out from the allophycocyanin core. Green hexameric shapes each represent allophycocyanin hexamers at the phycobilisome core.

5.2. Results & Discussion

5.2.1. Native Mass Spectrometry Identified a Novel +305 Da Modification of the Phycobilisome Light Harvesting Protein Complex

All data was acquired by K. R. Adoni, with the exception of data acquired for Figure 5.4 in Section 5.2.1.1, which was acquired by Dr J. Bellamy-Carter. This data was processed, analysed and Figure 5.4 created by K. R. Adoni. All proteins derived from native mass spectra peaks are reported with their theoretical mass, experimental mass (with standard deviation error) and mass deviation from theoretical mass in Appendix Section 8.6

5.2.1.1. Native mass spectrometry of Arthrospira maxima revealed a +305 Da mass shift on the C-phycocyanin dimer

Native mass spectrometry was regularly performed on *Arthrospira maxima* cyanobacteria for the purposes of quality control and to ensure that cells were producing phycobilisomal proteins in abundance, such that the culture could be used for phycobilisome-based experimentation. The native mass spectrometry workflow involved growing cells in culture media, followed by thermal cell lysis (repeated freeze-thaw cycles), phycobilisome disassembly and buffer exchange into 100 mM ammonium acetate, before analysis via native mass spectrometry (Figure 5.2). The freeze-thaw cycle was performed using dH_2O (as described in Materials and Methods Section 2.3.4.1), as the low ionic strength of water facilitates phycobilisome dissociation[417–419]. Further, phycobilisome proteins represent ~50-60 % of cyanobacteria's total water-soluble protein content[420]. As such, native mass spectrometry was used to identify both phycocyanin and allophycocyanin, an example of such a native mass spectrum for this

workflow is provided below (Figure 5.3). A charge state distribution was observed corresponding to the allophycocyanin hexamer ($\alpha_3\beta_3$) of molecular weight 107,359 Da[313]. Another charge state distribution at lower *m/z* correlated with the C-phycocyanin dimer ($\alpha\beta$), (37,467 Da). Both calculated masses are consistent with previously published data[313]. Often, as well as the peaks identified and characterised in Figure 5.3, peaks corresponding to C-phycocyanin hexamer were also observed from the above workflow. Overall, the data highlight the relative ease in which allophycocyanin and phycocyanin can be analysed by native MS due to their high abundance within cyanobacteria[313].

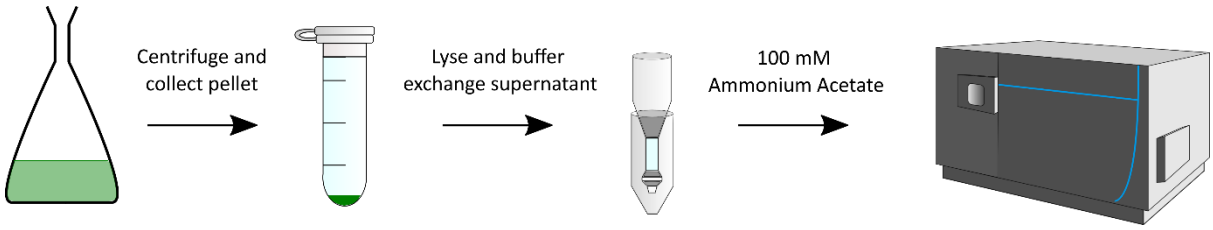


Figure 5.2 | Native mass spectrometry workflow of cyanobacteria from sample culturing to MS analysis. Samples were grown and cultured with relevant media before centrifugation and lysis of cell pellet. This was followed by buffer exchange of lysate into 100 mM ammonium acetate and analysis by native mass spectrometry using an Orbitrap Q-Exactive HF mass spectrometer.

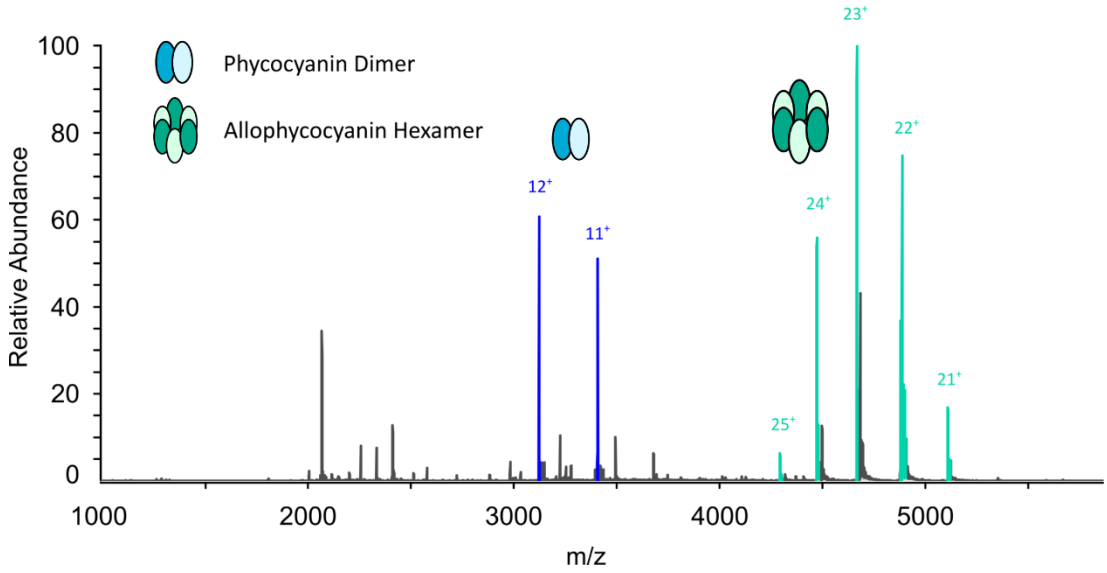


Figure 5.3 | Full native mass spectrum of *Arthrospira maxima* at a mass range of 1000 – 6000 *m/z*. Peaks corresponding to C-phycocyanin dimer are labelled in blue, and peaks corresponding to allophycocyanin hexamer are labelled in green, together with schematics to represent the multi-subunit structure of the protein complexes. α-phycocyanin subunit coloured dark blue, β-phycocyanin subunit coloured pale blue. α-allophycocyanin subunit coloured dark green, β-allophycocyanin subunit coloured pale green.

Over the course of 26 days of *Arthrospira maxima* cyanobacteria growth, native mass spectrometry data revealed the appearance and subsequent intensification of peaks adjacent to the 3123 *m/z* and 3407 *m/z* peaks, corresponding to the 12^+ and 11^+ charge states respectively, of the C-phycocyanin dimer (Figure 5.4). These peaks were identified at 3148 *m/z* and 3434 *m/z*, for the 12^+ and 11^+ charge states, respectively. No independent variable was knowingly altered throughout this period. These newly identified peaks corresponded to a mass shift of +305 Da on the *Arthrospira maxima* C-phycocyanin dimer. Further, the relative intensification of these +305 Da phycocyanin peaks over the course of weeks relative to their unmodified proteoforms were detected, despite no known independent changes in growth conditions over this period (Figure 5.4). 26 days after the initial native mass spectrum was acquired, the C-phycocyanin dimer had almost completely been converted into its +305 Da mass shifted proteoform

complex (~85 % of the 12^+ charge state corresponded to the +305 Da modified dimer). The stimulation of this mass shift remains to be characterised. Post-translational modification as the culprit for this mass difference was considered. The S-glutathionylation PTM, with a mass of 305 Da, has been shown to protect cysteine sulfhydryl groups of proteins from oxidative damage[285]. Further, as a light harvesting protein complex, the cyanobacteria phycobilisome is vulnerable to such damage from the generation and activity of reactive oxygen species (ROS) in its immediate vicinity, as has been reported with thylakoid proteins in plants[421]. Complimentary mass spectrometry techniques were used to investigate the possible S-glutathionylation of the phycobilisome, as well as to characterise the relevance of this PTM in the context of protection against oxidative damage.

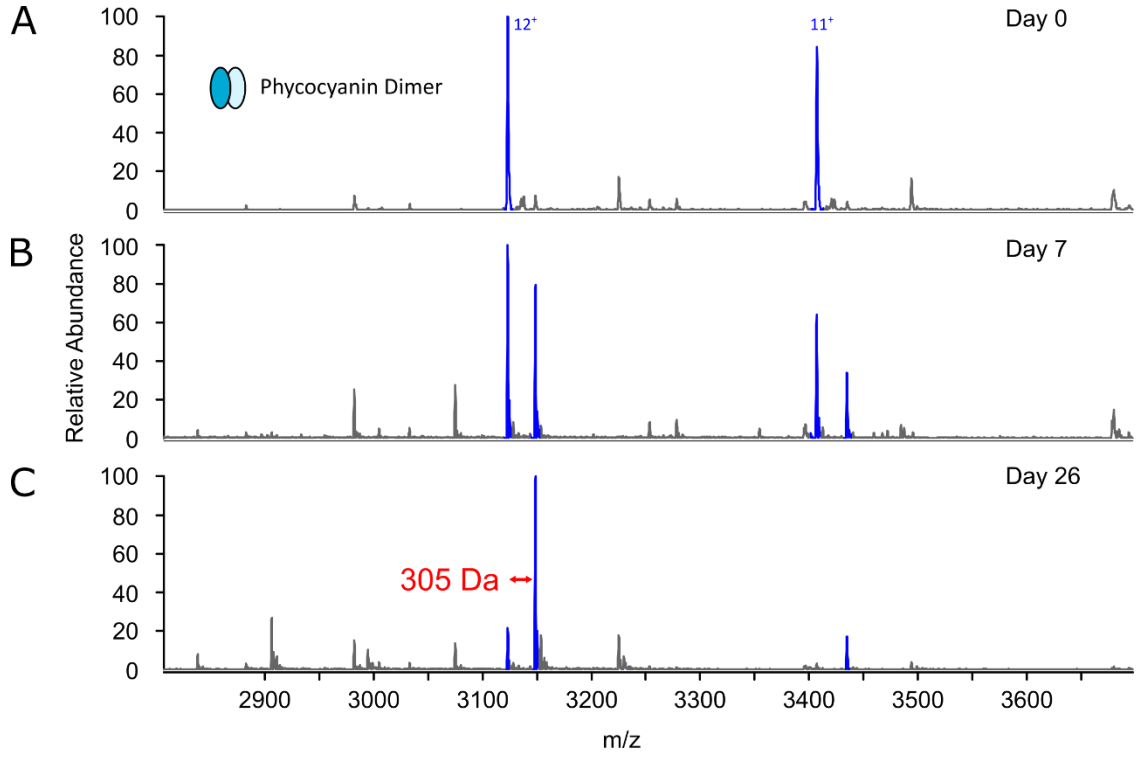


Figure 5.4 | Zoomed in (2800 -3700 *m/z*) native mass spectra revealed appearance and subsequent progression of peaks adjacent to the 3123 *m/z* peak and 3407 *m/z* peaks, corresponding to the 12^+ and 11^+ charge states respectively, of the C-phycocyanin dimer. Native mass spectrum of the C-phycocyanin dimer of *Arthrospira maxima*. (A) Native mass spectrum of sample harvested on Day 0. (B) Native mass spectrum of the same sample, harvested 1 week later on Day 7. (C) Native mass spectrum of the same sample; harvested a further 19 days on from (B), on Day 26. Peaks: 3148 *m/z* and 3434 *m/z*, for the 12^+ and 11^+ charge states of C-phycocyanin dimer, respectively; corresponded to a mass shift of +305 Da, located on the C-phycocyanin dimer of the phycobilisome light-harvesting protein complex. These peaks showed increased intensity over the course of weeks. *Data acquired by Dr J. Bellamy-Carter.*

5.2.1.2. Native mass spectrometry revealed +305 Da mass shift occurred at the biologically active C-phycocyanin hexamer level

Upon the identification of a +305 Da mass shift on the C-phycocyanin dimer, a closely related species of cyanobacteria: *Arthrospira platensis* was investigated using the same workflow as before. The native mass spectrum for *Arthrospira platensis* contained abundant peaks corresponding to C-phycocyanin in its dimer and hexamer forms (Figure 5.5a). The same mass shift of +305 Da on peaks corresponding to the C-phycocyanin dimer were identified in this species as that of *Arthrospira maxima* (Figure 5.5b). C-

phycocyanin exists within the phycobilisome light harvesting protein complex as stacked hexamers that ultimately form dodecameric structures which protrudes out of the complex for light absorption (Figure 5.1). Importantly, the identification of this +305 Da mass shift on the hexamer represented the occurrence of this modification in the C-phycocyanin functional state, implicating biological relevance to this modification. Further to this, the +305 Da mass shift was also observed at the hexamer level of C-phycocyanin as 1× +305 Da (= +305 Da) and 2× +305 Da (= +610 Da) mass shifts at decremental levels of relative abundance (Figure 5.5c). This pattern was consistent with the hypothesis of the mass shift specifically occurring on one of either the α- or β-subunit, assuming the detection of C-phycocyanin with 3× +305 Da (= +915 Da) was limited by instrument dynamic range.

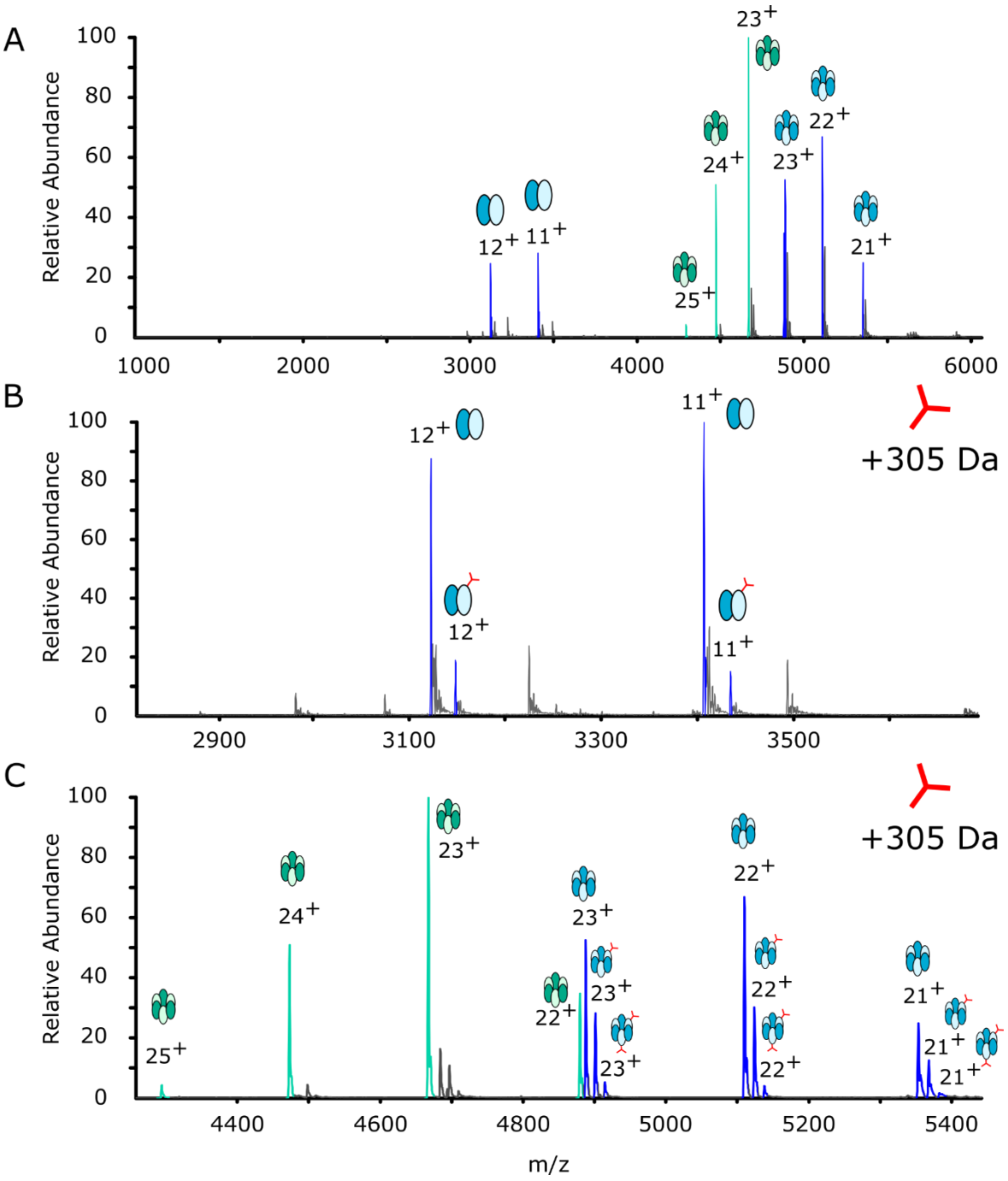


Figure 5.5 | Native mass spectrum of *Arthrospira platensis*. (A) Complete mass spectrum using mass range: 1000 – 6000 *m/z*. (B) Zoomed in at 2800 – 3700 *m/z* to show C-phycocyanin dimer. (C) Zoomed in at 4300 – 5400 *m/z* to show hexamer species of C-phycocyanin and allophycocyanin. +305 Da mass shift represented by red icon. All cultures taken at 0930 hours on day of harvest.

5.2.2. Native Top-Down Mass Spectrometry of C-Phycocyanin Dimer Localised the +305 Da Mass Shift onto the β-subunit

Whilst applied for the identification of the +305 Da mass shift on the C-phycocyanin dimer and hexamer; native mass spectrometry alone could not localise the mass shift onto the α- or β-subunit. To address this, the C-phycocyanin dimer 11^+ peak of *m/z* 3434 (including the +305 Da mass shift from its theoretical mass) was isolated and subjected to higher energy collision induced dissociation (HCD) to dissociate the α- and β-subunits. Relative to the theoretical masses of the α- and β-phycocyanin subunits, the +305 Da mass shift was localised specifically onto the β-subunit of the C-phycocyanin dimer (Figure 5.6), correlating with the previous identification of no more than one +305 Da mass shift on the C-phycocyanin dimer. Thus far, native, and native top-down mass spectrometry data suggested the existence of an alternative proteoform of the C-phycocyanin β-subunit, with hexamer proteoform complex variation depending on the number of observed +305 Da mass shifted β-subunits within the biologically active hexamer complex. The different proteoform complexes of C-phycocyanin hexamer ranged from the most abundant non-modified version (with no +305 Da mass shifted β-subunit identified), to one +305 Da β-subunit, and two β-subunits out of three occupying this mass shifted proteoform. The relative abundances of these proteoform complexes decreased, respectively. Further, this mass shift was not observed with allophycocyanin, suggesting its exclusive modification of C-phycocyanin. To truly characterise the proteoform variability of the C-phycocyanin hexamer, its confirmation as an S-glutathione PTM containing proteoform, as well as the site localisation of the modification, was necessary. For this purpose, bottom-up mass spectrometry was utilised.

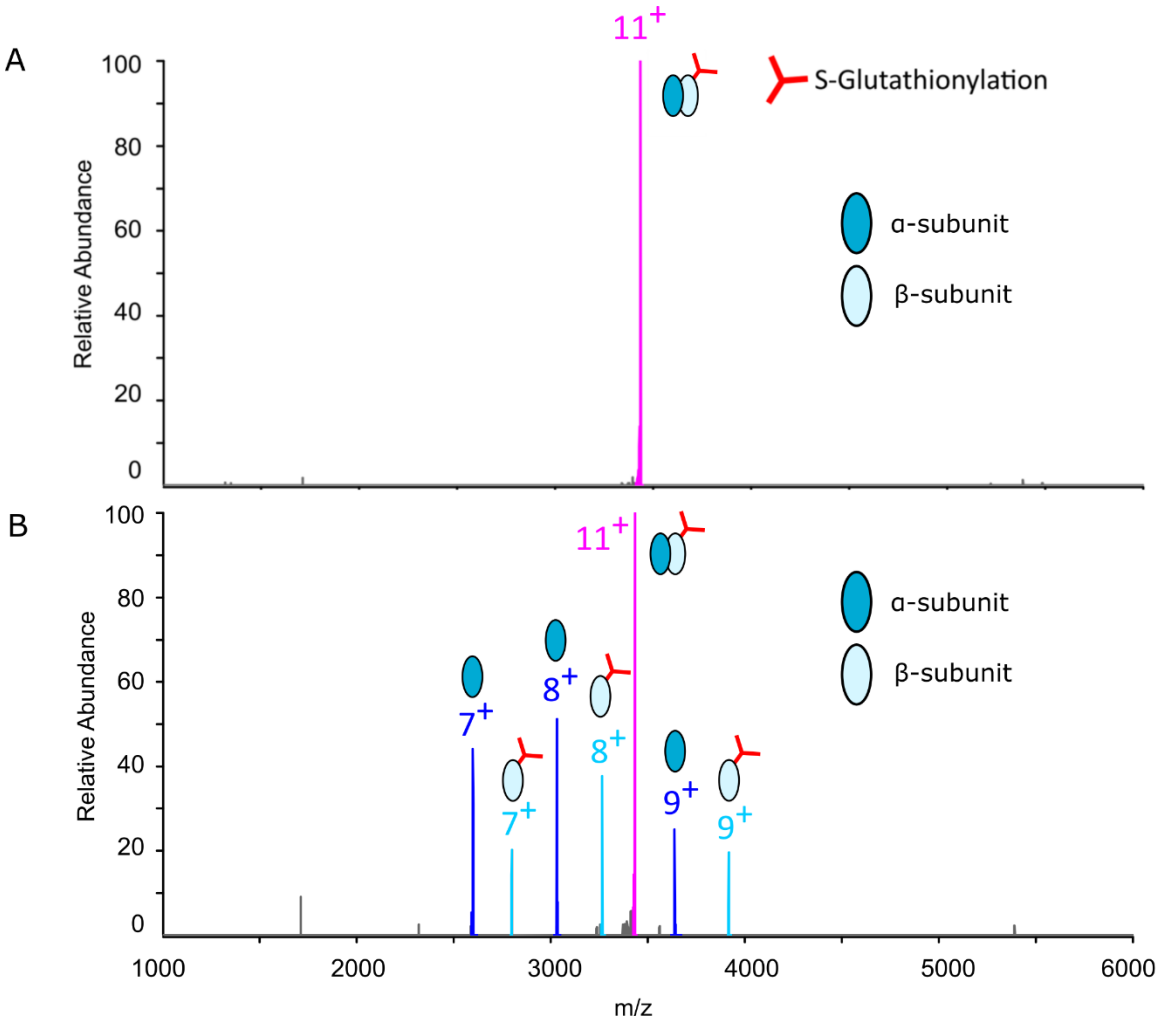


Figure 5.6 | +305 Da mass shift localised onto β-subunit of C-phycocyanin. (A) 3434 *m/z* peak corresponds to the 11^+ charge state of the C-phycocyanin dimer. Isolation spectrum (B) HCD dissociation of the C-phycocyanin dimer localised the +305 Da mass shift onto the β-subunit of C-phycocyanin.

5.2.3. Bottom-Up Mass Spectrometry Characterised +305 Da Mass Shift as a Novel S-Glutathionylation PTM on Cysteine-109 of the C-Phycocyanin β-Subunit

As seen in Figure 5.4c, native mass spectrometry revealed an almost complete conversion of the C-phycocyanin dimer towards the +305 Da mass shifted proteoform. As such, the extracted protein lysate from this sample was investigated using bottom-up

mass spectrometry (see Materials and Methods Section 2.3.5.2). As discussed above, a mass shift of +305 Da corresponds to the S-glutathionylation post-translational modification of cysteine. Within the *Arthrospira platensis* C-phycocyanin β-subunit, four cysteine residues are present. Of these, cysteine-82 and cysteine-153 are bound to phycocyanobilin 1 and phycocyanobilin 2 chromophores, respectively[422].

C-phycocyanin α-subunit

MKTPLTEAVSIADSQGRFLSSTEIQVAFGRFRQAKAGLEAAKAL
TSKADSLISGAAQAVYNKFPYTTQMQGPNYAADQRGKDKCARDI
GYYLRMVTYCLIAGGTGPMDEYLIAGIDEINRTFELSPSWYIEA
LKYIKANHGLSGDAATEANSYLDYAINALS

C-phycocyanin β-subunit

MFDAFTKVVSQADTRGEMLSTAQIDALSQMVAESNKRLDAVNRI
TSNASTIVSNAARSLFAEQPQLIAPGGNAYTSRRMAACLRDMEI
ILRYVTYAVFAGDASVLEDRCLNGLRETYLALGTPGSSVAVGVG
KMKEAALAIVNDPAGITPGDCSALASEIASYFDRACAAVS

Figure 5.7 | Amino acid sequences for C-phycocyanin α- and β-subunits, respectively.

Thus, cysteine-109 and cysteine-168 were considered as viable candidate amino acids for S-glutathionylation. This PTM was included in the search database before processing the bottom-up mass spectrometry data. Consequently, a peptide spectral match (PSM) containing the S-glutathionylation PTM at cysteine-109 of peptide amino acid residues 109-114 corresponding to amino acids: CLNGLR of the C-phycocyanin β-subunit was identified, with a false discovery rate (FDR) that satisfied the high-confidence threshold (FDR = 0.01). To ensure that the PSM had not been erroneously mismatched to the S-glutathionylation PTM, the MS/MS spectrum was manually analysed (Figure 5.8a). Characteristic peaks corresponding to HCD fragmentation of the peptide, as well as

glutathione, were assigned (Figure 5.8b). The identification of S-glutathionylation on one specific cysteine residue corroborated the previous native and native top-down mass spectrometry analysis of C-phycocyanin, whereby only one +305 Da mass shift was detected per β-subunit. This multi-modal mass spectrometry approach therefore suggested specific modification of cysteine-109 on the C-phycocyanin β-subunit.

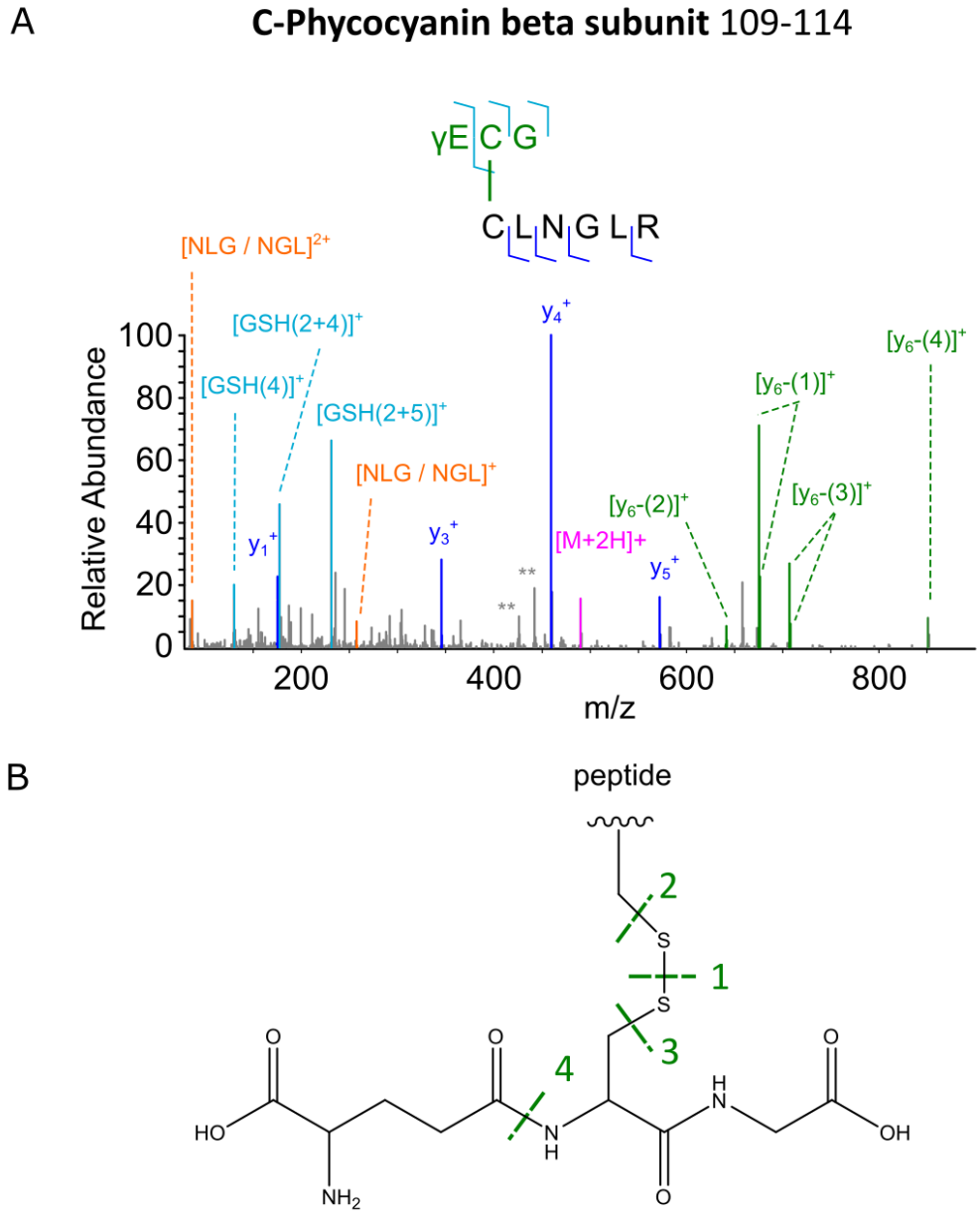


Figure 5.8 | Bottom-up mass spectrometry localised the S-glutathione PTM to cysteine-109 of the C-phycocyanin β-subunit. (A) Annotated MS/MS spectrum of precursor peptide ion: CLNGLR (residues 109-114) with HCD fragmentation. Four of the five *y*-ions of the peptide 109-114 series are annotated in blue. Peptide bond cleavages of glutathione are annotated in cyan, intra-molecular fragmentation products from cleavage at the disulfide bond and nearby bonds are annotated in green, internal fragments are annotated in orange, precursor ion is annotated in pink, ** represents neutral loss fragment ions. (B) Chemical structure of S-glutathione covalently bound to a peptide. Dashed green lines represent fragmentation of bond by HCD, corresponding to fragment ions identified in tandem MS spectrum with associated number to indicate fragmented bond(s) for the generation of the fragment ion peak.

In an evolutionary context, the phycobilisome light harvesting complex represents the first oxygenic photosynthesising protein machinery of life[272]. As such, the identification of an ancient S-glutathionylated proteoform could be associated with a mechanistically relevant role in photosynthesis. Sequence alignment analysis of amino acid residues 100-117 of *Arthrospira platensis* C-phycocyanin β-subunit to 16 homologues from various cyanobacterial species was performed to assess the level of conservation of this S-glutathionylatable cysteine residue (Table 5.1). The analysis revealed the high level of conservation of the S-glutathionylatable cysteine-109 and its flanking amino acid residue across species, suggesting a biologically relevant mechanistic role for this PTM in an evolutionarily important manner. However, to fully characterise the significance of this proteoform, investigation into the stimulation of C-phycocyanin β-subunit cysteine-109 S-glutathionylation by an independent variable was necessary.

Table 5.1 | Sequence alignment of *Arthrospira platensis* C-phycocyanin β-subunit sequence: 100 – 117 (including S-glutathionylatable cysteine-109) against 16 homologues from different cyanobacteria species. Alignment performed using BLAST sequence alignment tool[423]. Alignment illustrates almost complete conservation of cysteine-109 across all species included in the search, as well as flanking amino acids (9 amino acids on the N-terminal side and 8 amino acids on the C-terminal side). Red box locates conserved cysteine-109 of S-glutathionylated *Arthrospira platensis* C-phycocyanin β-subunit across other species homologues.

Species	***Sequence (with amino acid position)***		
Arthrospira platensis	100	GDASVLEDRCLNGLRETY	117
Synechocystis sp. (strain PCC 6803)	100	GDASVLEDRCLNGLRETY	117
Nostoc sp. (strain PCC 7120)	101	GDASVLDDRCLNGLRETY	118
Agmenellum quadruplicatum	100	GDASVLNDRCLNGLRETY	117
Synechococcus sp. (strain PCC 7002)	101	GDASILDDRCLNGLRETY	118
Mastigocladus laminosus	101	GDASILDDRCLNGLRETY	118
Cyanidium caldarium	100	GDSSVLDDRCLNGLRETY	117
Aglaothamnion neglectum	100	GDSSVLDDRCLNGLRETY	117
Porphyra purpurea	100	GDSSVLDDRCLNGLRETY	117
Neoporphyra haitanensis	100	GDSSVLDDRCLNGLRETY	117
Neopyropia yezoensis	100	GDSSVLDDRCLNGLRETY	117
Cyanophora paradoxa	100	GDSSVLDDRCLNGLRETY	117
Thermosynechococcus vestitus	100	GDSSVLDDRCLNGLRETY	117
Microchaete diplosiphon	100	GDASVLDDRCLNGLKETY	117
Galdieria sulphuraria	100	GDSSILDDRCLNGLRETY	117
Rhodella violacea	100	GDSSVLDDRCLNGLKETY	117
Aphanizomenon flos-aquae	85	GDASVLDDR----RETY	97

5.2.4. Native Mass Spectrometry for the Investigation of Stimulation of S-Glutathione PTM on C-Phycocyanin β-subunit

5.2.4.1. S-Glutathionylation of C-phycocyanin β-subunit was not stimulated by high light intensity

Whilst S-glutathionylation of C-phycocyanin had not previously been reported, examples of other phycobiliprotein S-glutathionylation such as: phycobiliprotein ApcE, phycobilisome 7.8 kDa linker polypeptide, and phycobilisome rod-core linker polypeptide CpcG have been reported[301]. Added to this, many other photosynthetic proteins including: Ribulose Bisphosphate Carboxylase (RuBisCo), Photosystem I reaction centre subunit II psaD, and Photosystem I reaction centre subunit PsaK have been reported in their S-glutathionylated state[300,301]. Also, numerous downstream metabolic proteins of photosynthesis and respiration have been reported to undergo S-glutathionylation. All of these proteins require protection from the oxidative damage that can occur due to reactive oxygen species that are generated as a consequence of photosynthesis, and the same is true of mitochondrial proteins involved in the electron transport chain of respiration[424]. Consequently, their S-glutathionylation protects the nucleophilic cysteine sulfhydryl groups from damaging oxidative reactions with ROS. As such, to assess the S-glutathionylation status of C-phycocyanin under conditions that favoured photosynthesis, *Arthrospira platensis* was exposed to 150 mA of white light for three hours before cells were lysed and the phycobilisome was analysed by native mass spectrometry. It was hypothesised that such conditions would boost photosynthetic activity, and consequently elevate ROS levels within the cells. This could stimulate S-glutathionylation of C-phycocyanin β cysteine-109 to protect the protein against oxidative damage. However, native mass spectrometry revealed no increase in the S-

glutathionylated proteoform of the C-phycocyanin dimer (Figure 5.9). Surprisingly, whilst low abundance peaks corresponding to S-glutathionylated C-phycocyanin dimer were observed in the control condition (exposed to 30 mA light for three hours), these peaks were eradicated upon exposure of the cyanobacteria to high light intensity (150 mA for 3 h). Previous work has attributed exposure to high light intensity to the photo-induced dissociation and degradation of the phycobilisome[425,426], and this may have contributed to the complete loss of glutathionylated C-phycocyanin, with the PTM perhaps acting as a signal for protein degradation. In this scenario, high availability of light energy ensures that chlorophyll within the thylakoid is sufficient to carry out light harvesting processes without the need for the complimentary phycobiliproteins. Thus, whilst the native mass spectrometry data contradicts the hypothesis that increased light intensity stimulates S-glutathionylation of the phycobilisome complex, further work must be carried out to ensure that the loss of peaks corresponding to the S-glutathionylated proteoform are not just a consequence of the disassembly and degradation of the phycobilisome, possibly in response to its S-glutathionylation via a signalling mechanism.

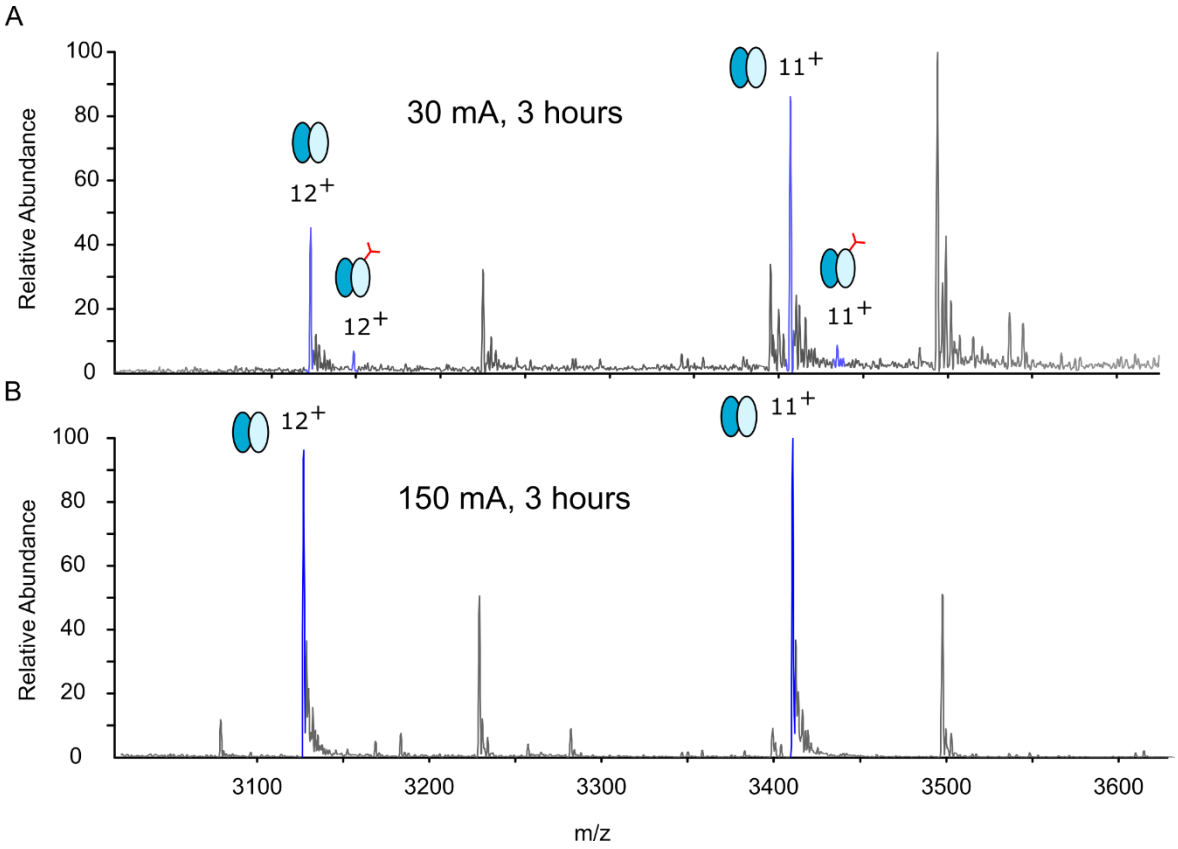


Figure 5.9 | Native mass spectra of control (30 mA) (A) and exposure to high light intensity (150 mA) (B) of C-phycocyanin dimer from *Arthrospira platensis*. Spectra zoomed in at mass range 3000 – 3600 Da. Exposure to high light did not stimulate increased S-glutathionylation PTM on β-subunit of C-phycocyanin. Red icon attached to dimer schematic represents S-glutathionylation.

5.2.4.2. S-Glutathionylation of C-phycocyanin β-subunit Cysteine-109 was stimulated by exposure to hydrogen peroxide

Whilst exposure of cyanobacteria to high light intensity did not generate peaks corresponding to S-glutathionylated C-phycocyanin in the native mass spectra, the findings suggested the necessity for a more nuanced approach to characterising this PTM on the phycobilisome. Increased photosynthesis, though increasing ROS abundance and possibly promoting S-glutathionylation, may concomitantly facilitate the breakdown and degradation of the phycobilisome, impeding the identification of the PTM on the

complex's subunits. As such, 100 µM hydrogen peroxide (H_2O_2) was administered to the cyanobacteria culture, to increase ROS concentration without increasing photosynthetic activity (and subsequent phycobilisome breakdown). Following this, cells were lysed and the phycobilisome was analysed by native mass spectrometry. Upon exposure to 100 µM hydrogen peroxide, native mass spectrometry revealed the loss of peaks corresponding to the non-glutathionylated C-phycocyanin dimer (at *m/z* 3123 and 3407 corresponding to the 12^+ and 11^+ peaks respectively), which contributed ~70 % of the total C-phycocyanin dimer abundance in the control condition. Thus, C-phycocyanin dimer was exclusively detected as an S-glutathionylated proteoform dimer (Figure 5.10). The data corroborated the working hypothesis that S-glutathionylation of cysteine-109 of the C-phycocyanin β-subunit was stimulated by the increased abundance of ROS, with this PTM protecting the cysteine sulfhydryl group from oxidative damage. Whilst novel in its identification, this cysteine-109 PTM may also be a response to hydrogen peroxide in multiple different contexts, with H_2O_2 known to play a significant role as a secondary signalling molecule for biological processes including but not limited to: biotic and abiotic stress[427], development[428], apoptosis[429] and pathogenesis[430]. However, the specific increase in S-glutathionylated C-phycocyanin in response to increased ROS (hydrogen peroxide) and the identification of other phycobiliprotein S-glutathionylation events in previous work[300,301], together with the requirement of the phycobilisome to protect itself in an oxidative environment such as photosynthesis, suggests this to be a novel regulatory mechanism of the cyanobacteria phycobilisome complex in response to ROS (Figure 5.11). Further work must be performed to characterise the specific signalling context of this PTM. For example, whether S-glutathionylation both protects the protein and signals the complex for ordered disassembly and subsequent degradation.

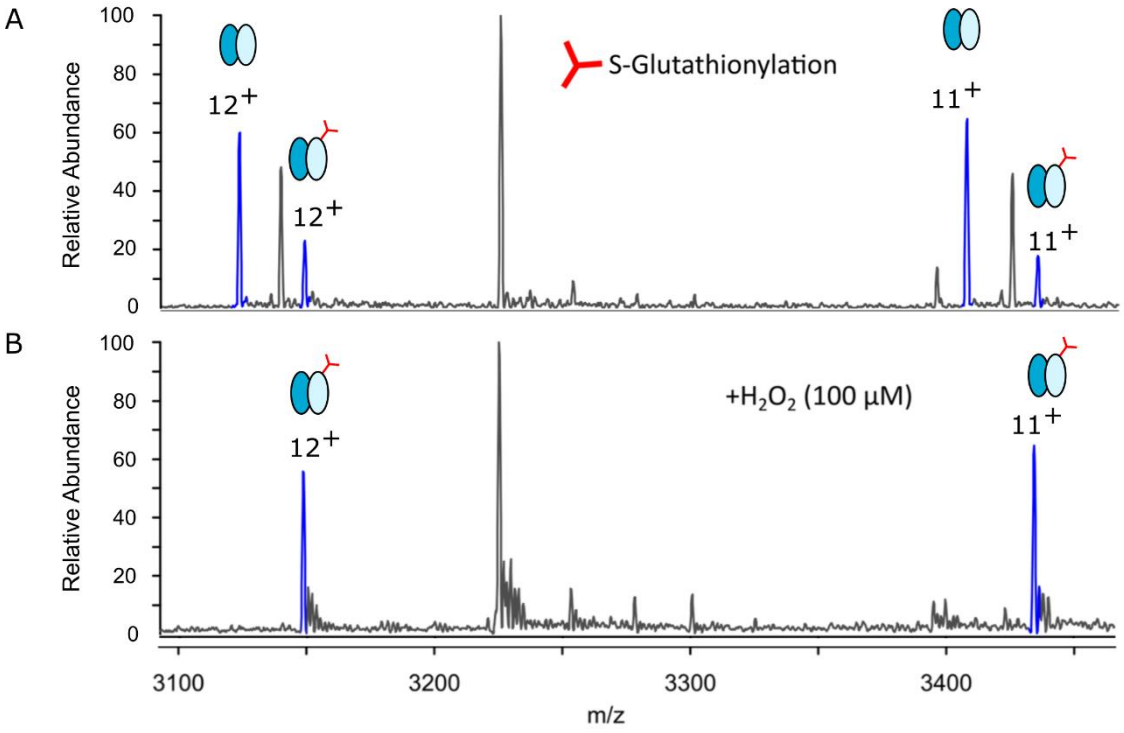


Figure 5.10 | Native mass spectra of *Arthrospira platensis* control (A) and exposure of *Arthrospira platensis* to 100 µM H_2O_2 for 3 h (B). Spectra zoomed in at mass range 3100 - 3450 *m/z*. Exposure to H_2O_2 stimulated S-glutathionylation of C-phycocyanin dimer whilst decreasing the abundance of the non-modified proteoform. Red icon attached to dimer schematic represents S-glutathionylation PTM. Both conditions were subjected to 30 mA of light.

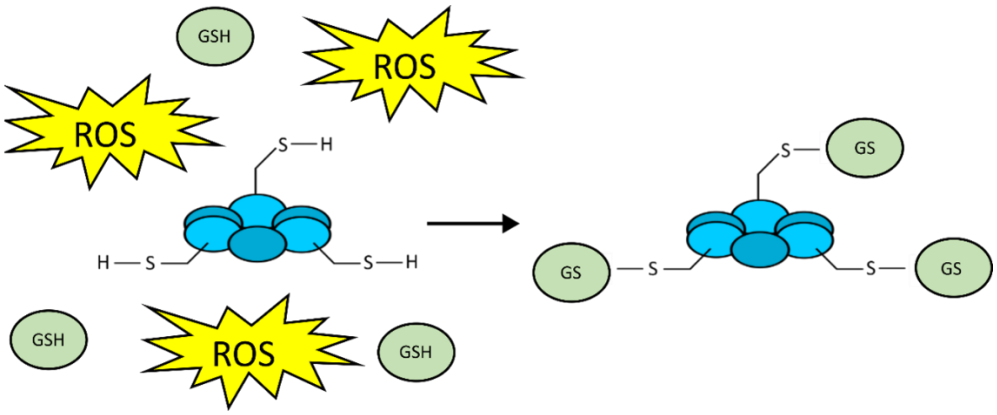


Figure 5.11 | Working model for S-glutathionylation of C-phycocyanin upon exposure to reactive oxygen species. GSH represents reduced glutathione, S-H represents sulfhydryl groups of cysteine residues in C-phycocyanin β-subunit. ROS represents reactive oxygen species in the vicinity of the phycobilisome. Light blue subunits represent β-phycocyanin, dark blue subunits represent α-phycocyanin. Specifically, hydrogen peroxide was used as the ROS stimulator of S-glutathionylation in this work.

5.2.5. LC-FAIMS-MS/MS for the Investigation of S-Glutathionylation in *Arthrospira platensis*

Thus far, a multi-modal mass spectrometry approach had been applied to identify and characterise a novel S-glutathionylation PTM on the β-subunit of the C-phycocyanin protein within the multi-protein light harvesting phycobilisome complex. Then, this modified S-glutathionylated proteoform was characterised as potentially evolutionarily relevant due to the level of conservation of the modified cysteine residue and its flanking amino acids across cyanobacteria species. Following this, stimulation of this modified proteoform relative to the unmodified version was probed with exposure to high light intensity and hydrogen peroxide as a reactive oxygen species, with H_2O_2 facilitating elevated levels of S-glutathionylation on the C-phycocyanin β-subunit of the phycobilisome. As previously discussed, proteins relating to photosynthesis, respiration and other metabolic processes have been identified as glutathionylated proteoforms (Section 5.2.4.1). However, the complete characterisation of S-glutathionylation as a contributor to global proteoform variability has been limited by insufficient techniques for its accurate identification and characterisation. Considering the physiochemical impact of S-glutathionylation of a peptide in a bottom-up mass spectrometry context, the covalent addition of this tripeptide onto a trypsin digested peptide (via a cysteine amino acid residue) was predicted to impart unique characteristics on these peptides relating to their ion mobility. In some ways, the S-glutathionylated peptide could be considered as a crosslinked peptide entity, whereby the glutathione tripeptide and the tryptic peptide are crosslinked via a disulfide bond, leading to changes in the collision cross sectional area of the peptide. As such, LC-FAIMS-MS/MS was applied to exploit these unique properties of S-glutathionylated peptides compared to other tryptic peptides within the

lysed cyanobacterial proteome sample, for more efficient S-glutathionylated precursor peptide ion identification and fragmentation. The LC-FAIMS-MS/MS workflow developed in Chapter 3 and applied in Chapter 4 was used for this purpose (for specific workflow parameters, see Materials and Methods Section 2.3.5).

5.2.5.1. LC-FAIMS-MS/MS for the investigation of high light intensity stimulated proteome S-glutathionylation in Arthrospira platensis

First, *Arthrospira platensis* was exposed to 150 mA light intensity for 3 hours before cell lysis and analysis by LC-FAIMS-MS/MS. For the corresponding control condition, *Arthrospira platensis* was exposed to 30 mA light intensity for 3 hours before analysis using the same methodology. No S-glutathionylated peptides were identified in at least 2/3 biological replicates within the control sample, whereas 26 S-glutathionylated peptide spectral matches (PSMs) (Figure 5.12a), corresponding to 6 peptides (Figure 5.12b), each of which belonged to a separate protein (Figure 5.12c), were identified upon high light exposure. Of the six S-glutathionylated proteins that were identified (Table 5.2), three were found to contain transferase activity, with some of these protein linked to photosynthesis and respiration[431–433]. The specific S-glutathionylation of 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase and Glycerol-3-phosphate acyltransferase as metabolism related proteins was indicative of the relevance of this PTM to photosynthesis and respiration and their downstream metabolism proteins. PolyA polymerase and Glycerol-3-phosphate acyltransferase have, to our knowledge, not previously been reported as S-glutathionylated proteoforms. Further, the detection of 26 PSMs upon exposure of the cyanobacteria to high light intensity, compared to no PSM identifications in the control, suggested a role for light in stimulating S-glutathionylated

proteoforms. The S-glutathionylated proteoform of the C-phycocyanin β-subunit was not identified in the dataset, despite being characterised by native mass spectrometry and detected with standard LC-MS/MS.

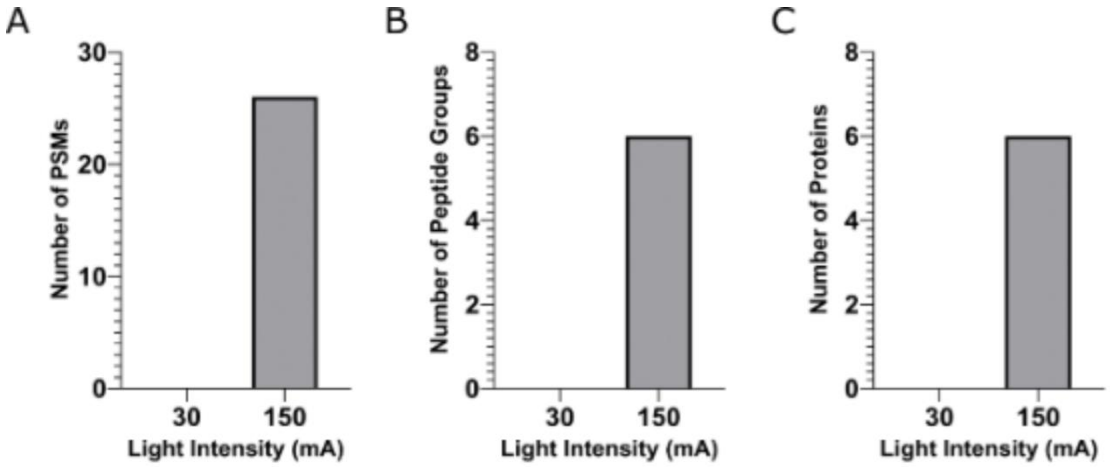


Figure 5.12 | Number of S-Glutathionylated PSMs (A), Peptide Groups (B) and Proteins (C) identified at 30 mA vs 150 mA light intensity. S-glutathionylation was stimulated by exposure to high light intensity. All PSMs identified in at least 2 of 3 biological replicates

Table 5.2 | S-Glutathionylated proteins after exposure to high light intensity (150 mA, 3 h). *Protein Function information from UniProt*[315]

Protein Description	**Protein Accession**	**Sequence Location**	**Modification(s)**	**Protein Function**
2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase	D4ZNL7	95-119	1xPhospho [T24(100)]; 1xGlutathione [C13(100)]	Catalyses 2-C-Methyl-D-erythritol 4-phosphate (MEP) and cytosine triphosphate (CTP) to 4-diphosphocytidyl-2-C-methyl-D-erythritol (CDPME) and inorganic pyrophosphate (PPi)[431]

Glycerol-3-phosphate acyltransferase	D4ZW09	38-55	1xAcetyl [R14]; 1xMethyl [K18(100)]; 1xPhospho [S4(95)]; 1xGlutathione [C16(100)]	Catalyses the conversion of glycerol-3-phosphate and long chain acyl-CoA into lysophosphatidic acid[433]
Methyltransferase 21 domain-containing protein	D5A3R0	10-27	1xMethyl [R18(100)]; 1xPhospho [S17(93.9)]; 1xGlutathione [C13(100)]	Ambiguous functionality (based on metagenomic analysis of *Arthrospira platensis).* Involved in methyl-group transfer.
PolyA polymerase	D5A120	1-29	1xPhospho [T/S/Y/H]; 1xGlutathione [C6(100)]	Catalyses the adenylation of RNA-3'OH to generate polyadenylate tail of newly synthesised pre-messenger RNA[432].
Uncharacterized protein	D4ZUV0	189-232	2xMethyl [R15(100); R20(100)]; 1xPhospho [S19(100)]; 1xGlutathione [C12(100)]	-
Uncharacterized protein	D4ZS41	189-226	1xAcetyl [K27(100)]; 2xPhospho [T/S]; 1xGlutathione [C20(100)]	-

5.2.5.2. FAIMS enriches S-Glutathionylated peptides between compensation voltages: -40.0 - -50.0 V

The large size of the S-glutathionylated PTM, and the unique ion mobility properties that it was predicted to impart on a peptide, suggested that a specific range of FAIMS CVs would enrich for S-glutathionylated peptides. As such, we next investigated identifications of these PSMs at each CV from the data generated by the above experiment (LC-FAIMS-MS/MS for investigation of stimulation of S-glutathionylation with high light intensity, Section 5.2.5.1). CVs: -40.0 V, -45.0 V and 50.0 V enriched for S-glutathionylated PSMs, identifying 93 % of total S-glutathionylation containing PSMs. Compensation voltages: -55.0 V and -60.0 V identified just 5.9 % and 1.0 % of the total identified S-glutathionylated PSMs, respectively. For this analysis, all PSM identifications were considered, regardless of the number of replicates they were identified in. This data suggested the preferential enrichment of S-glutathionylation between -40.0 V and -50.0 V in an LC-FAIMS-MS/MS workflow.

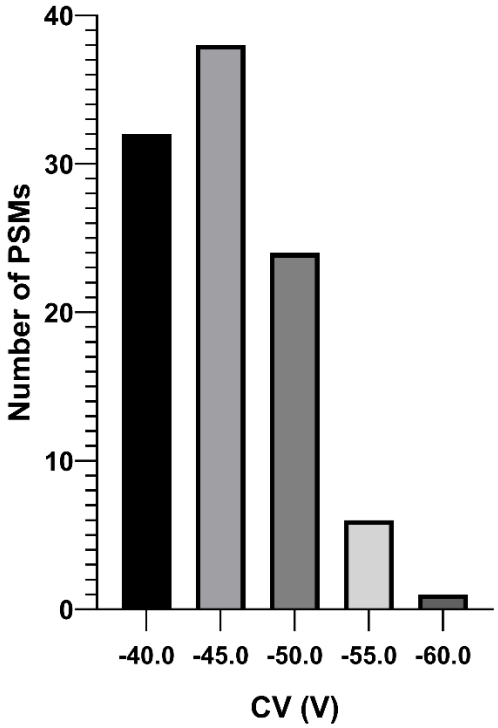


Figure 5.13 | Number of S-glutathionylated peptides vs Compensation Voltage (CV) at which they were identified. CVs: -40.0 V, -45.0 V and -50.0 V yielded 93 % of the total identified S-glutathionylated peptides.

5.2.5.3. Arthrospira platensis protein S-glutathionylation was increased upon stimulation with hydrogen peroxide

Having identified the light intensity stimulated S-glutathionylated proteoform of several proteins with LC-FAIMS-MS/MS, we next sought to investigate the role of hydrogen peroxide as a reactive oxygen species. H_2O_2 had already been identified as a stimulator of C-phycocyanin S-glutathionylation, possibly as a means of protecting the phycobilisome from ROS mediated degradation (Section 5.2.4.2). To this end, *Arthrospira platensis* was exposed to 100 μM hydrogen peroxide and incubated for 3 hours. 50 S-glutathionylated PSMs were identified upon hydrogen peroxide stimulation (present in at least 2 replicates) (Figure 5.14a), translating to 13 peptides (Figure 5.14b),

and 11 proteins (Figure 5.14c). In the control condition, just 1 S-glutathionylated PSM was detected (putative helicase, accession: D5A1E1). D-fructose 1,6-bisphosphatase class 2/sedoheptulose 1,7-bisphosphatase, fructose-1,6-bisphosphate aldolase, glycosyl transferase and sulfate transport system permease protein CysT were all S-glutathionylated upon treatment with H_2O_2. These proteins correspond to the metabolism of carbon fixation processes that start at photosynthesis, again suggesting the importance of S-glutathionylation in this domain (Table 5.3). Further, glycosyl transferase, reverse transcriptase, VWFA domain containing protein and sulfate transport system permease protein CysT have, to our knowledge, not previously been reported in their S-glutathionylated state. As was identified with exposure to high light, this work illustrates the increased S-glutathionylation of cyanobacterial proteins upon hydrogen peroxide exposure. Of note, the S-glutathionylated proteoform of the C-phycocyanin β-subunit was again not identified, perhaps due to the short sequence length of the modified peptide reducing its compatibility with FAIMS based enrichment.

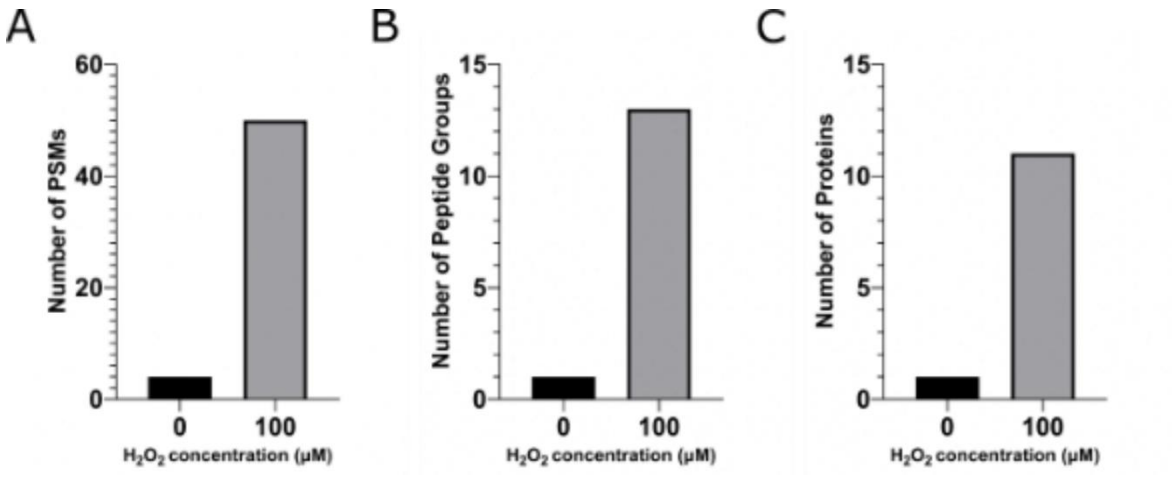


Figure 5.14 | Number of S-Glutathionylated PSMs (A), Peptide Groups (B) and Proteins (C) identified at 100 µM H_2O_2 versus 0 µM H_2O_2 (3-hour incubation). S-glutathionylation was stimulated by exposure to H_2O_2.

Table 5.3 | S-Glutathionylated proteins stimulated by exposure to H_2O_2. All peptides of proteins identified were found in 2 out of 2 biological replicates. *Protein Function information from UniProt*[315]

Protein Description	Protein Accession	Sequence Location	Modification(s)	Protein Function
D-fructose 1,6-bisphosphatase class 2/sedoheptulose 1,7-bisphosphatase	D4ZR96	150-164	1xGlutathione [C12(100)]	Catalyses the hydrolysis of fructose-1,6-bisphosphate and sedoheptulose-1,7-bisphosphate into fructose-6-phosphate and sedoheptulose-7-phosphate respectively
Fructose-1,6-bisphosphate aldolase	D4ZYP4	124-168	1xGlutathione [C22(100)]	Splits fructose-1,6-bisphosphate into dihydroxyacetone phosphate (DHAP) and glyceraldehyde-3-phosphate (G3P)

Probable protein phosphatase	D5A1Y6	436-455	1xAcetyl [R20]; 1xPhospho [S19(100)]; 1xGlutathione [C6(100)]	Likely involved in the dephosphorylation of substrate proteins
Putative glycosyl transferase	D4ZW46	150-174	1xAcetyl [K23(99.6)]; 1xPhospho [S9(89.2)]; 1xGlutathione [C14(100)]	Catalyses the transfer of a sugar moiety from a nucleotide sugar onto a glycosyl acceptor molecule
Reverse transcriptase homolog	D4ZQN2	454-480	2xMethyl [K23(100); R27(100)]; 1xPhospho [S17(99.8)]; 1xGlutathione [C25(100)]	DNA polymerase enzyme that transcribes single stranded RNA into DNA
Sulfate transport system permease protein CysT	D5A1U3	248-288	2xPhospho [T/S/Y]; 1xGlutathione [C29(100)]	Part of the ABC transporter complex involved in sulfate/thiosulfate import into cell
Uncharacterized protein	D5A143	30-56	1xMethyl [K9(94.9)]; 2xPhospho [Y6(100); Y24(100)]; 1xGlutathione [C13(100)]	-
Uncharacterized protein	D4ZRX2	198-214	1xPhospho [S15(100)]; 1xAcetyl [K3(100)]; 1xGlutathione [C4(100)]	-
Uncharacterized protein	D4ZR32	133-161	1xPhospho [H16(87.7)]; 2xAcetyl [K18(100); R];	-

			1xGlutathione [C8(97.2)]	
Uncharacterized protein	D4ZUE5	169-205	1xAcetyl [K37]; 2xPhospho [T10; T15]; 1xGlutathione [C24]	-
Uncharacterized protein	D4ZRX7	158-177	1xGlutathione [C2(100)]	-
Uncharacterized protein	D4ZRZ1	110-141	1xPhospho [S/T/Y]; 1xGlutathione [C21(100)]	-
VWFA domain-containing protein	D4ZUM0	134-158	2xPhospho [S5(94.5); H17(99.6)]; 1xGlutathione [C23(100)]	VWFA domain involved in cell adhesion. Usually, extracellular

Despite the above findings from the LC-FAIMS-MS/MS experiments, the low translation of high confidence (FDR = 0.01) peptides (identified in at least one biological replicate) to validated peptides (identified in at least 2 biological replicates) was lower than expected in both experiments. For example, in the LC-FAIMS-MS/MS investigation of the impact of high light intensity on cyanobacterial global protein S-glutathionylation; 55 high confidence S-glutathionylated peptide groups were identified after exposure of cyanobacteria to 150 mA light intensity, compared to just 7 in the control. However, of the 55 high light intensity S-glutathionylated peptide groups; only six were identified in at least two of the three biological replicates, deriving from six proteins. None of the seven control S-glutathionylated peptide groups were present in at

least two biological replicates. This trend was also demonstrated in the LC-FAIMS-MS/MS investigation of cyanobacteria hydrogen peroxide exposure on protein S-glutathionylation, whereby 69 high confidence peptides were identified upon hydrogen peroxide exposure, translating to 13 peptides that were present across both biological replicates. S-glutathionylation, like other oxidative modifications, is a labile PTM[434], which may help explain the inconsistency across replicates that led to such few validated peptide identifications (i.e. in two or more replicates) compared to high confidence peptide identifications. However, adjustment of the methods used in online liquid chromatography and mass spectrometry may help overcome this hurdle. Specifically, investigation of the advantages of biasing for more hydrophobic peptides via more time spent at relatively high organic mobile phase may preferentially elute these larger, bulkier S-glutathionylated peptide ions. Within the mass spectrometry duty cycle, specific glutathione signature peaks could be used to identify an S-glutathionylated peptide ion, before more of this species is preferentially isolated and fragmented for MS/MS analysis. From a fragmentation perspective, softer techniques such as ETD or UVPD[83,435] may improve precursor ion identification via preservation of the S-glutathionylation PTM.

5.2.6. LC-FAIMS-MS/MS Revealed Possible PTM Crosstalk Between Phosphorylation and S-Glutathionylation PTMs

As discussed in the Introduction, PTM crosstalk remains one of the bottlenecks to accurately characterise signalling mechanisms across biochemistry[318]. LC-FAIMS-MS/MS, developed in Chapter 3 and implemented in Chapter 4, provided an up to 6-fold improvement in the identification of candidates for positive PTM crosstalk for such

PTMs as: phosphorylation, acetylation, methylation, HexNAc, nitrosylation, deamidation and ubiquitination[115]. Of the six validated S-glutathionylated peptide groups that were identified in the high light intensity experiment; all six also contained a phosphorylation site in the same peptide group. As such, these peptides could be considered as candidates for phosphorylation-glutathionylation positive PTM crosstalk. Further to this, of the 13 S-glutathionylated peptide groups identified with cyanobacteria exposure to H_2O_2; 10 contained an additional phosphorylation site, again inferring a possible crosstalking relationship between the two PTMs. Having used LC-FAIMS-MS/MS to acquire 180 S-glutathionylated high confidence (FDR = 0.01) peptide groups (present in at least 1 replicate), analysis revealed phosphorylation to be a common co-occurring modification. 84 % of the validated S-glutathionylated peptides in Section 5.2.5 were also phosphorylated. This data supplemented previous investigations into such a positive crosstalking mechanism: 172 proteins were found to contain both phosphorylation and oxidised thiols in the *Chlamydomonas reinhardtii* microalgae species in a sequential enrichment LC-MS/MS experiment[436]. Further, the phagocyte NADPH oxidase p47phox was shown to require sequential phosphorylation and S-glutathionylation for its sustained ROS generation during pathogenesis[437]. Our data corroborated these reports, implying coordinated positive PTM crosstalk between phosphorylation and S-glutathionylation within cyanobacteria.

5.3. Conclusions

In this work, a multi-modal mass spectrometry approach was applied to identify and characterise a novel S-glutathionylation PTM on the phycobilisome light harvest protein complex. Native mass spectrometry revealed a novel modification on the cyanobacteria C-phycocyanin protein of the phycobilisome light harvesting complex. Native top-down mass spectrometry localised this +305 Da modification onto the β-subunit of C-phycocyanin, with LC-MS/MS site localising the modification onto cysteine-109 and assigning it as S-glutathionylation with manual annotation of the MS/MS spectrum. Previous work has reported S-glutathionylation as a protective PTM against oxidative damage of proteins by reactive oxygen species during photosynthesis[438]. In this vein, cyanobacteria were exposed to high light intensity, and whilst native mass spectrometry did not reveal increased S-glutathionylation of phycocyanin, exposure of cyanobacteria to hydrogen peroxide stimulated the conversion of C-phycocyanin into its S-glutathionylated proteoform, as identified by native mass spectrometry. These data suggest a novel role for the S-glutathione PTM on phycocyanin, perhaps to protect the protein when exposed to abnormally high levels of reactive oxygen species. Further, LC-FAIMS-MS/MS demonstrated an increase in protein S-glutathionylation upon exposure of cyanobacteria to high light intensity and hydrogen peroxide, often with phosphorylation co-occurring in the identified peptides- as candidates for positive PTM crosstalk. Many of these validated S-glutathionylated proteins were found to be involved in metabolic processes related to photosynthesis or respiration, as electron transport chain utilising processes that generate reactive oxygen species as by-products. Together, these findings highlight the importance of proteoform regulation of the cyanobacteria light-harvesting protein complex and downstream metabolic proteins, with S-

glutathionylation possibly playing a key role in regulating and protecting proteins from oxidative damage.

6. Conclusions and Future Directions

The work presented in this thesis describes mass spectrometry-based approaches towards the improved identification and characterisation of protein post-translational modifications (PTM), particularly in the context of PTM crosstalk leading to proteoform variability. PTMs are dynamic regulators of gene transcribed proteins, providing a post-biosynthesis infrastructure for protein regulation and thus enabling cells to adapt to environmental change in real-time[439]. Further to this, an appreciation of the combinatorial impact of PTMs that occur concurrently on a protein, as a proteoform, has become a priority in understanding the variable behaviour of the protein[38,150]. As the field of biochemistry strives to acquire complete characterisation of all PTM-decorated proteoforms of every gene coded protein, the onus has fallen on the shoulders of mass spectrometry to meet this challenging demand, as described by the Human Proteome Project[2,8]. To this end, rapid advances in sample pre-enrichment[148,440], liquid chromatography[153,441], ion mobility[115,189,190,442], precursor ion activation[82,88,90,435], mass spectrometry instrumentation including sensitivity, resolution, scan speed and dynamic range[443] and post-acquisition data processing[194,196,207,213] have primed the field of mass spectrometry towards this target.

In Chapter 3, High Field Asymmetric Waveform Ion Mobility Spectrometry (FAIMS) was incorporated into an LC-MS/MS bottom-up proteomics workflow for the development of a novel technique for improved identification of multiple PTM containing peptides as candidates for positive PTM crosstalk, in HeLa S3 human cervical cancer cells[115]. Previous work has demonstrated the utility of LC-FAIMS-MS/MS for improved peptide and protein identification for greater proteome coverage in high-throughput proteomics[133,186], and in improving the identification of multiply charged, multi-phosphorylated peptides[111,320]. Based off this data, we hypothesised the improved

detection of multiple-PTM containing peptides on account of their preferential transmission thought the FAIMS device at specific compensation voltages. As such, despite the low stoichiometry of these peptides relative to unmodified and mono-modified peptides, LC-FAIMS-MS/MS enhanced multi-PTM peptide identifications by up to 6-fold, with 40 % of the identified candidate crosstalk sites not previously reported. Further, the exclusive identification of specific PTM-containing peptidoforms enabled the order of PTM crosstalking to be investigated, with the occurrence of one PTM possibly required for the second PTM, such as with FACT complex subunit SSRP1 phospho-priming and Polyadenylate Binding Protein 2 (PABP2) N-terminal acetylation as a prerequisite for two methylations, with further work required to validate these findings.

Whilst this workflow can be applied across standard high-throughput proteomics experiments for a deeper characterisation of PTMs, improvements can still be targeted for greater efficiency of the technique. For example, the internal stepping LC-FAIMS-MS/MS workflow demonstrated a 2-fold improvement in multi-PTM peptide identifications on the same timescale as standard LC-MS/MS, despite data from external stepping and internal stepping LC-FAIMS-MS/MS suggesting the suboptimal choice of cycling compensation voltages used (-90.0 V, as one of four compensation voltages used, did not identify a single validated multi-PTM containing peptide). Further, analysis of specific dual-PTM combinations suggested their differential preference for specific compensation voltages. Thus, the user should tailor the specific CVs within the FAIMS internal stepping workflow to the PTM-crosstalk of interest. As well as improved rates of PTM-identification, PTM site localisation for each PTM is required to truly characterise the significance of a specific modification on a protein. For MS/MS

fragmentation, the linear ion trap mass analyser of the Orbitrap Eclipse Tribrid mass spectrometer was chosen on account of its superior scan speed for a faster duty cycle, translating to improved throughput. But its relatively poor mass resolution compromised the confidence in PTM site localisations within the peptide, especially in cases where multiple modifiable PTM sites were present on the same tryptic peptide. To this end, previous literature has warned of the perils of PTM misassignment from ion trap based tandem MS acquisition[66,217]. Thus, it is suggested that the user carefully consider the compromise of throughput vs PTM site localisation accuracy, before choosing between the ion trap and Orbitrap for tandem MS acquisition (when using Orbitrap Tribrid instrumentation).

As with the development of any analytical technique, its applicability must be probed against biologically relevant case studies to demonstrate its utility. As such, the LC-FAIMS-MS/MS technique that was developed in Chapter 3 was then used for the investigation of positive PTM crosstalk within human SUM52 triple negative breast carcinoma cancer cells. Fibroblast Growth Factor Receptor 2 (FGFR2) gene amplification of the SUM52 cell line results in the amplification of signalling pathways including RAS/MAP/ERK/RSK and PI3K/PDK1/AKT, stimulated by hyperactive FGFR2 tyrosine kinase activity[252]. Terminal targets of these signal transduction pathways promote gene expression corresponding to upregulated proliferation, metabolism, cell motility and survival[177,350]. A SILAC LC-FAIMS-MS/MS approach, with FAIMS internal stepping of titanium dioxide phosphoenriched SUM52 cell lysate and FAIMS external stepping of non-enriched SUM52 cell lysate, was used to probe positive PTM crosstalk that may contribute towards the cancer phenotype, using FGFR inhibitor molecules AZD4547 and SU5402 as two separate SILAC conditions. The

detection of 498 site localised multi-PTM containing peptides, whose abundance was significantly modified by inhibition of FGFR tyrosine kinase, demonstrated the importance of positive PTM crosstalk in SUM52 cells. Further, proteoform characterisation of proteins involved in actin dynamics for lamellipodia formation, as a mechanism for cell migration, suggested their highly orchestrated combinatorial PTM regulation, stimulated by FGFR. For example, peptidoform identification of cofilin peptide sequence 1 – 13 suggested the requirement for a previously unidentified methionine cleavage and N-terminal acetylation for protein sequestering serine-3 phosphorylation, to promote actin polymerisation. Thus, LC-FAIMS-MS/MS for the detection of multiple PTM containing peptides proved highly compatible with the high-throughput proteomics-based investigation of cancer cells.

Though tandem MS analysis was performed in the Orbitrap to ensure higher accuracy of PTM site localisation, as appose to the linear ion trap as in Chapter 3, the associated reduction in throughput compared to ion trap tandem MS acquisition was somewhat offset by more appropriate FAIMS compensation voltages. For example, internal stepping involved cycling the FAIMS device through -45.0 V, -52.5 V, -60.0 V and -67.5 V, based on data analysis that was performed in Chapter 3. However, without prior PTM-enrichment of samples that underwent external stepping LC-FAIMS-MS/MS, the methodology was reliant on Data-Dependant Acquisition (DDA) to identify low stoichiometry multi-PTM containing peptides. In the context of PTM characterisation, one could argue that Data Independent Acquisition (DIA) would have been a more suitable technique. Whilst DDA fragments the most abundant precursor ions of an LC eluted ion packet (with dynamic exclusion to prevent the repeated acquisition of MS/MS spectra pertaining to the same precursor peptide), DIA fragments precursor ions by

scanning across *m/z* bins, potentially increasing the likelihood that lower stoichiometry multiple PTM containing peptides would be chosen for MS/MS acquisition. This viewpoint is supplemented by successful applications of DIA in PTM based bottom-up proteomics[138–140,444], with ever improving DIA data-processing software driving this methodology to the forefront of PTM based proteomics[445], however issues around search space requirements leading to increased false positive identifications must be considered. Another limitation of LC-FAIMS-MS/MS involves the necessary digestion of the protein into peptides. Oftentimes, the tertiary and quaternary structure of proteins and protein complexes facilitates PTM crosstalk, as modifications that are far away in primary sequence may be in close proximity in three-dimensional space. As such, the crosstalking relationship between PTMs that were localised across different peptides could not be characterised with bottom-up proteomics. A multi-modal mass spectrometry investigation, that combines LC-FAIMS-MS/MS and native/top-down proteomics workflows could prove powerful in this regard, with the bottom-up data informing further characterisation of specific proteins for top-down analysis.

Indeed, in Chapter 5, multi-modal mass spectrometry was applied to investigate the evolutionarily relevant cyanobacteria light harvesting protein complex- the phycobilisome. Here, native, and native top-down mass spectrometry identified a novel S-glutathionylation PTM onto the β-subunit of the C-phycocyanin hexameric protein complex within the phycobilisome. LC-MS/MS then complimented these findings with site localisation of this S-glutathione PTM onto cysteine-109 of C-phycocyanin, with sequence conservation analysis highlighting the potential relevance of this modification in an evolutionary context. All 17 homologues that were investigated across cyanobacterial species contained this cysteine, with well conserved flanking amino

acids. Native mass spectrometry analysis after incubation of cyanobacteria with reactive oxygen species (hydrogen peroxide) suggested this modification to be a biologically relevant, evolutionarily conserved mechanism for the protection of the phycobilisome against reactive oxygen species (ROS) induced damage, with photosynthesis known to generate abundant ROS[438]. As well, LC-FAIMS-MS/MS was used to enrich the global cyanobacterial proteome for S-glutathionylated species, without prior sample enrichment or fractionation. An upregulated S-glutathionylome was identified upon stimulation of cyanobacteria with high light intensity, and with hydrogen peroxide. Also of note, the concurrent identification of phosphorylation in 84 % of validated S-glutathionylated peptides suggested a positive crosstalking relationship between these PTMs in cyanobacteria, as previously reported[436]. However, just 11 % of the 180 high confidence S-glutathionylated peptides that were identified upon high light or hydrogen peroxide exposure were present across at least two biological replicates, indicating the low reproducibility of these identifications. HCD cleavage of disulfide bonds can occur upon MS/MS fragmentation[446], resulting in the loss of the glutathione moiety, with associated neutral loss of pyroglutamic acid[447]. As such, the exploitation of more PTM-friendly ion activation techniques, to prevent the cleavage of labile PTMs, may be worth considering as a future direction of this work. For example, neutral loss triggered EThcD may prove fruitful for superior glutathione identification, with neutral loss scanning at Δ129 Da. This would also add valuable information in probing S-glutathionylation-phosphorylation positive crosstalk, with neutral loss scanning at Δ98/Δ80 Da upon phosphodiester bond cleavage for phosphate release[88]. In a more general LC-FAIMS-MS/MS context, neutral loss triggered EThcD could improve multi-PTM peptide identification and site localisation for more informative results.

To conclude, LC-FAIMS-MS/MS as a technique for the identification of multiple PTM containing peptides has shown success in contributing to the ultimate objective of complete proteoform characterisation within the field of protein biochemistry. However, the limitations of LC-FAIMS-MS/MS, as discussed above, suggests that multiple complimentary mass spectrometry techniques, including native, native top-down and middle down proteomics are required to truly understand the dynamic behaviour of proteins. With instrumentation continuing to develop, and computational processing efficiency converting more spectra to tangible peptide/protein hits, this will inevitably lead to a better understanding of how multiple post-translational modifications orchestrate protein regulation.

7. References

(1) Lander, E. S.; Linton, L. M.; Birren, B.; Nusbaum, C.; Zody, M. C.; Baldwin, J.; Devon, K.; Dewar, K.; Doyle, M.; FitzHugh, W.; Funke, R.; Gage, D.; Harris, K.; Heaford, A.; Howland, J.; Kann, L.; Lehoczky, J.; LeVine, R.; McEwan, P.; McKernan, K.; Meldrim, J.; Mesirov, J. P.; Miranda, C.; Morris, W.; Naylor, J.; Raymond, C.; Rosetti, M.; Santos, R.; Sheridan, A.; Sougnez, C.; Stange-Thomann, N.; Stojanovic, N.; Subramanian, A.; Wyman, D.; Rogers, J.; Sulston, J.; Ainscough, R.; Beck, S.; Bentley, D.; Burton, J.; Clee, C.; Carter, N.; Coulson, A.; Deadman, R.; Deloukas, P.; Dunham, A.; Dunham, I.; Durbin, R.; French, L.; Grafham, D.; Gregory, S.; Hubbard, T.; Humphray, S.; Hunt, A.; Jones, M.; Lloyd, C.; McMurray, A.; Matthews, L.; Mercer, S.; Milne, S.; Mullikin, J. C.; Mungall, A.; Plumb, R.; Ross, M.; Shownkeen, R.; Sims, S.; Waterston, R. H.; Wilson, R. K.; Hillier, L. W.; McPherson, J. D.; Marra, M. A.; Mardis, E. R.; Fulton, L. A.; Chinwalla, A. T.; Pepin, K. H.; Gish, W. R.; Chissoe, S. L.; Wendl, M. C.; Delehaunty, K. D.; Miner, T. L.; Delehaunty, A.; Kramer, J. B.; Cook, L. L.; Fulton, R. S.; Johnson, D. L.; Minx, P. J.; Clifton, S. W.; Hawkins, T.; Branscomb, E.; Predki, P.; Richardson, P.; Wenning, S.; Slezak, T.; Doggett, N.; Cheng, J.-F.; Olsen, A.; Lucas, S.; Elkin, C.; Uberbacher, E.; Frazier, M.; Gibbs, R. A.; Muzny, D. M.; Scherer, S. E.; Bouck, J. B.; Sodergren, E. J.; Worley, K. C.; Rives, C. M.; Gorrell, J. H.; Metzker, M. L.; Naylor, S. L.; Kucherlapati, R. S.; Nelson, D. L.; Weinstock, G. M.; Sakaki, Y.; Fujiyama, A.; Hattori, M.; Yada, T.; Toyoda, A.; Itoh, T.; Kawagoe, C.; Watanabe, H.; Totoki, Y.; Taylor, T.; Weissenbach, J.; Heilig, R.; Saurin, W.; Artiguenave, F.; Brottier, P.; Bruls, T.; Pelletier, E.; Robert, C.; Wincker, P.; Rosenthal, A.; Platzer, M.; Nyakatura, G.; Taudien, S.; Rump, A.; Smith, D. R.; Doucette-Stamm, L.; Rubenfield, M.; Weinstock, K.; Lee, H. M.; Dubois, J.; Yang, H.; Yu, J.; Wang, J.; Huang, G.; Gu, J.; Hood, L.; Rowen, L.; Madan, A.; Qin, S.; Davis, R. W.; Federspiel, N. A.; Abola, A. P.; Proctor, M. J.; Roe, B. A.; Chen, F.; Pan, H.; Ramser, J.; Lehrach, H.; Reinhardt, R.; McCombie, W. R.; de la Bastide, M.; Dedhia, N.; Blöcker, H.; Hornischer, K.; Nordsiek, G.; Agarwala, R.; Aravind, L.; Bailey, J. A.; Bateman, A.; Batzoglou, S.; Birney, E.; Bork, P.; Brown, D. G.; Burge, C. B.; Cerutti, L.; Chen, H.-C.; Church, D.; Clamp, M.; Copley, R. R.; Doerks, T.; Eddy, S. R.; Eichler, E. E.; Furey, T. S.; Galagan, J.; Gilbert, J. G. R.; Harmon, C.; Hayashizaki, Y.; Haussler, D.; Hermjakob, H.; Hokamp, K.; Jang, W.; Johnson, L. S.; Jones, T. A.; Kasif, S.; Kaspryzk, A.; Kennedy, S.; Kent, W. J.; Kitts, P.; Koonin, E. V.; Korf, I.; Kulp, D.; Lancet, D.; Lowe, T. M.; McLysaght, A.; Mikkelsen, T.; Moran, J. V.; Mulder, N.; Pollara, V. J.; Ponting, C. P.; Schuler, G.; Schultz, J.; Slater, G.; Smit, A. F. A.; Stupka, E.; Szustakowki, J.; Thierry-Mieg, D.; Thierry-Mieg, J.; Wagner, L.; Wallis, J.; Wheeler, R.; Williams, A.; Wolf, Y. I.; Wolfe, K. H.; Yang, S.-P.; Yeh, R.-F.; Collins, F.; Guyer, M. S.; Peterson, J.; Felsenfeld, A.; Wetterstrand, K. A.; Myers, R. M.; Schmutz, J.; Dickson, M.; Grimwood, J.; Cox, D. R.; Olson, M. V.; Kaul, R.; Raymond, C.; Shimizu, N.; Kawasaki, K.; Minoshima, S.; Evans, G. A.; Athanasiou, M.; Schultz, R.; Patrinos, A.; Morgan, M. J.; International Human Genome Sequencing Consortium; Whitehead Institute for Biomedical Research, C. for G. R.; The Sanger Centre:; Washington University Genome Sequencing Center; US DOE Joint Genome Institute:; Baylor College of Medicine Human Genome Sequencing Center:; RIKEN Genomic Sciences Center:; Genoscope and CNRS UMR-8030:;

Department of Genome Analysis, I. of M. B.; GTC Sequencing Center:; Beijing Genomics Institute/Human Genome Center:; Multimegabase Sequencing Center, T. I. for S. B.; Stanford Genome Technology Center:; University of Oklahoma's Advanced Center for Genome Technology:; Max Planck Institute for Molecular Genetics:; Cold Spring Harbor Laboratory, L. A. H. G. C.; GBF—German Research Centre for Biotechnology:; *Genome Analysis Group (listed in alphabetical order, also includes individuals listed under other headings):; Scientific management: National Human Genome Research Institute, U. N. I. of H.; Stanford Human Genome Center:; University of Washington Genome Center:; Department of Molecular Biology, K. U. S. of M.; University of Texas Southwestern Medical Center at Dallas:; Office of Science, U. D. of E.; The Wellcome Trust: Initial Sequencing and Analysis of the Human Genome. *Nature* **2001**, *409* (6822), 860–921. https://doi.org/10.1038/35057062.

(2) Legrain, P.; Aebersold, R.; Archakov, A.; Bairoch, A.; Bala, K.; Beretta, L.; Bergeron, J.; Borchers, C. H.; Corthals, G. L.; Costello, C. E.; Deutsch, E. W.; Domon, B.; Hancock, W.; He, F.; Hochstrasser, D.; Marko-Varga, G.; Salekdeh, G. H.; Sechi, S.; Snyder, M.; Srivastava, S.; Uhlén, M.; Wu, C. H.; Yamamoto, T.; Paik, Y.-K.; Omenn, G. S. The Human Proteome Project: Current State and Future Direction. *Mol. Cell. Proteomics* **2011**, *10* (7). https://doi.org/10.1074/mcp.M111.009993.

(3) Sánchez-Baracaldo, P.; Cardona, T. On the Origin of Oxygenic Photosynthesis and Cyanobacteria. *New Phytol.* **2020**, *225* (4), 1440–1446. https://doi.org/10.1111/nph.16249.

(4) Berg, J. M.; Tymoczko, J. L.; Stryer, L. *Biochemistry*, 7th ed., International ed.; W.H. Freeman: Basingstoke, 2012.

(5) Ramazi, S.; Zahiri, J. Post-Translational Modifications in Proteins: Resources, Tools and Prediction Methods. *Database* **2021**, *2021*, baab012. https://doi.org/10.1093/database/baab012.

(6) Holzer, H.; Heinrich, P. C. Control of Proteolysis. *Annu. Rev. Biochem.* **1980**, *49* (1), 63–91. https://doi.org/10.1146/annurev.bi.49.070180.000431.

(7) Goth, C. K.; Vakhrushev, S. Y.; Joshi, H. J.; Clausen, H.; Schjoldager, K. T. Fine-Tuning Limited Proteolysis: A Major Role for Regulated Site-Specific O-Glycosylation. *Trends Biochem. Sci.* **2018**, *43* (4), 269–284. https://doi.org/10.1016/j.tibs.2018.02.005.

(8) Mann, M.; Jensen, O. N. Proteomic Analysis of Post-Translational Modifications. *Nat. Biotechnol.* **2003**, *21* (3), 255–261. https://doi.org/10.1038/nbt0303-255.

(9) Wang, Y.-C.; Peterson, S. E.; Loring, J. F. Protein Post-Translational Modifications and Regulation of Pluripotency in Human Stem Cells. *Cell Res.* **2014**, *24* (2), 143–160. https://doi.org/10.1038/cr.2013.151.

(10) Dhanasekaran, N.; Tsim, S.-T.; Dermott, J. M.; Onesime, D. Regulation of Cell Proliferation by G Proteins. *Oncogene* **1998**, *17* (11), 1383–1394. https://doi.org/10.1038/sj.onc.1202242.

(11) Basson, M. A. Signaling in Cell Differentiation and Morphogenesis. *Cold Spring Harb. Perspect. Biol.* **2012**, *4* (6), a008151. https://doi.org/10.1101/cshperspect.a008151.

(12) Jaqaman, K.; Grinstein, S. Regulation from within: The Cytoskeleton in Transmembrane Signaling. *Trends Cell Biol.* **2012**, *22* (10), 515–526. https://doi.org/10.1016/j.tcb.2012.07.006.

(13) Kurucz, E.; Zettervall, C.-J.; Sinka, R.; Vilmos, P.; Pivarcsi, A.; Ekengren, S.; Hegedüs, Z.; Ando, I.; Hultmark, D. Hemese, a Hemocyte-Specific Transmembrane Protein, Affects the Cellular Immune Response in Drosophila. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (5), 2622–2627. https://doi.org/10.1073/pnas.0436940100.
(14) Schneider, P.; Tschopp, J. Apoptosis Induced by Death Receptors. In *Pharmacochemistry Library*; Gulini, U., Gianella, M., Quaglia, W., Marucci, G., Eds.; Receptor Chemistry towards the Third Millennium; Elsevier, 2000; Vol. 31, pp 281–286. https://doi.org/10.1016/S0165-7208(00)80030-6.
(15) Feinberg, A. P.; Vogelstein, B. Hypomethylation Distinguishes Genes of Some Human Cancers from Their Normal Counterparts. *Nature* **1983**, *301* (5895), 89–92. https://doi.org/10.1038/301089a0.
(16) Schaffert, L.-N.; Carter, W. G. Do Post-Translational Modifications Influence Protein Aggregation in Neurodegenerative Diseases: A Systematic Review. *Brain Sci.* **2020**, *10* (4), 232. https://doi.org/10.3390/brainsci10040232.
(17) Wu, Z.; Huang, R.; Yuan, L. Crosstalk of Intracellular Post-Translational Modifications in Cancer. *Arch. Biochem. Biophys.* **2019**, *676*, 108138. https://doi.org/10.1016/j.abb.2019.108138.
(18) Torres, M. P.; Dewhurst, H.; Sundararaman, N. Proteome-Wide Structural Analysis of PTM Hotspots Reveals Regulatory Elements Predicted to Impact Biological Function and Disease *. *Mol. Cell. Proteomics* **2016**, *15* (11), 3513–3528. https://doi.org/10.1074/mcp.M116.062331.
(19) Chatterjee, B.; Thakur, S. S. Investigation of Post-Translational Modifications in Type 2 Diabetes. *Clin. Proteomics* **2018**, *15*, 32. https://doi.org/10.1186/s12014-018-9208-y.
(20) Khoury, G. A.; Baliban, R. C.; Floudas, C. A. Proteome-Wide Post-Translational Modification Statistics: Frequency Analysis and Curation of the Swiss-Prot Database. *Sci. Rep.* **2011**, *1* (1), 90. https://doi.org/10.1038/srep00090.
(21) Huang, K.-Y.; Lee, T.-Y.; Kao, H.-J.; Ma, C.-T.; Lee, C.-C.; Lin, T.-H.; Chang, W.-C.; Huang, H.-D. DbPTM in 2019: Exploring Disease Association and Cross-Talk of Post-Translational Modifications. *Nucleic Acids Res.* **2019**, *47* (Database issue), D298–D308. https://doi.org/10.1093/nar/gky1074.
(22) Beltrao, P.; Bork, P.; Krogan, N. J.; van Noort, V. Evolution and Functional Cross-Talk of Protein Post-Translational Modifications. *Mol. Syst. Biol.* **2013**, *9*, 714. https://doi.org/10.1002/msb.201304521.
(23) Bradley, D.; Beltrao, P. Evolution of Protein Kinase Substrate Recognition at the Active Site. *PLOS Biol.* **2019**, *17* (6), e3000341. https://doi.org/10.1371/journal.pbio.3000341.
(24) Macek, B.; Forchhammer, K.; Hardouin, J.; Weber-Ban, E.; Grangeasse, C.; Mijakovic, I. Protein Post-Translational Modifications in Bacteria. *Nat. Rev. Microbiol.* **2019**, *17* (11), 651–664. https://doi.org/10.1038/s41579-019-0243-0.
(25) Pan, Z.; Wang, B.; Zhang, Y.; Wang, Y.; Ullah, S.; Jian, R.; Liu, Z.; Xue, Y. DbPSP: A Curated Database for Protein Phosphorylation Sites in Prokaryotes. *Database* **2015**, *2015*, bav031. https://doi.org/10.1093/database/bav031.
(26) Stancik, I. A.; Šestak, M. S.; Ji, B.; Axelson-Fisk, M.; Franjevic, D.; Jers, C.; Domazet-Lošo, T.; Mijakovic, I. Serine/Threonine Protein Kinases from Bacteria, Archaea and Eukarya Share a Common Evolutionary Origin Deeply Rooted in the Tree of Life. *J. Mol. Biol.* **2018**, *430* (1), 27–32. https://doi.org/10.1016/j.jmb.2017.11.004.

(27) *Protein Data Bank | Nucleic Acids Research | Oxford Academic*. https://academic.oup.com/nar/article/28/1/235/2384399 (accessed 2021-10-11).
(28) Humphrey, S. J.; James, D. E.; Mann, M. Protein Phosphorylation: A Major Switch Mechanism for Metabolic Regulation. *Trends Endocrinol. Metab.* **2015**, *26* (12), 676–687. https://doi.org/10.1016/j.tem.2015.09.013.
(29) Drazic, A.; Myklebust, L. M.; Ree, R.; Arnesen, T. The World of Protein Acetylation. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2016**, *1864* (10), 1372–1401. https://doi.org/10.1016/j.bbapap.2016.06.007.
(30) Ohtake, F.; Tsuchiya, H. The Emerging Complexity of Ubiquitin Architecture. *J. Biochem. (Tokyo)* **2017**, *161* (2), 125–133. https://doi.org/10.1093/jb/mvw088.
(31) Afjehi-Sadat, L.; Garcia, B. A. Comprehending Dynamic Protein Methylation with Mass Spectrometry. *Curr. Opin. Chem. Biol.* **2013**, *17* (1), 12–19. https://doi.org/10.1016/j.cbpa.2012.12.023.
(32) Yan, A.; Lennarz, W. J. Unraveling the Mechanism of Protein N-Glycosylation *. *J. Biol. Chem.* **2005**, *280* (5), 3121–3124. https://doi.org/10.1074/jbc.R400036200.
(33) Celen, A. B.; Sahin, U. Sumoylation on Its 25th Anniversary: Mechanisms, Pathology, and Emerging Concepts. *FEBS J.* **2020**, *287* (15), 3110–3140. https://doi.org/10.1111/febs.15319.
(34) Blanc, M.; David, F.; Abrami, L.; Migliozzi, D.; Armand, F.; Bürgi, J.; van der Goot, F. G. SwissPalm: Protein Palmitoylation Database. *F1000Research* **2015**, *4*, 261. https://doi.org/10.12688/f1000research.6464.1.
(35) Udenwobele, D. I.; Su, R.-C.; Good, S. V.; Ball, T. B.; Varma Shrivastav, S.; Shrivastav, A. Myristoylation: An Important Protein Modification in the Immune Response. *Front. Immunol.* **2017**, *8*.
(36) T. Shantha Raju. Prenylation of Proteins. In *Co and Post-Translational Modifications of Therapeutic Antibodies and Proteins*; John Wiley & Sons, Ltd, 2019; pp 177–181. https://doi.org/10.1002/9781119053354.ch14.
(37) Monigatti, F.; Hekking, B.; Steen, H. Protein Sulfation Analysis—A Primer. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2006**, *1764* (12), 1904–1913. https://doi.org/10.1016/j.bbapap.2006.07.002.
(38) Venne, A. S.; Kollipara, L.; Zahedi, R. P. The next Level of Complexity: Crosstalk of Posttranslational Modifications. *Proteomics* **2014**, *14* (4–5), 513–524. https://doi.org/10.1002/pmic.201300344.
(39) Wang, Z.; Gucek, M.; Hart, G. W. Cross-Talk between GlcNAcylation and Phosphorylation: Site-Specific Phosphorylation Dynamics in Response to Globally Elevated O-GlcNAc. *Proc. Natl. Acad. Sci.* **2008**, *105* (37), 13793–13798. https://doi.org/10.1073/pnas.0806216105.
(40) Zeidan, Q.; Hart, G. W. The Intersections between O-GlcNAcylation and Phosphorylation: Implications for Multiple Signaling Pathways. *J. Cell Sci.* **2010**, *123* (Pt 1), 13–22. https://doi.org/10.1242/jcs.053678.
(41) Laarse, S. A. M. van der; Leney, A. C.; Heck, A. J. R. Crosstalk between Phosphorylation and O-GlcNAcylation: Friend or Foe. *FEBS J.* **2018**, *285* (17), 3152–3167. https://doi.org/10.1111/febs.14491.
(42) Beurel, E.; Grieco, S. F.; Jope, R. S. Glycogen Synthase Kinase-3 (GSK3): Regulation, Actions, and Diseases. *Pharmacol. Ther.* **2015**, *148*, 114–131. https://doi.org/10.1016/j.pharmthera.2014.11.016.
(43) Lo, W. S.; Trievel, R. C.; Rojas, J. R.; Duggan, L.; Hsu, J. Y.; Allis, C. D.; Marmorstein, R.; Berger, S. L. Phosphorylation of Serine 10 in Histone H3 Is

Functionally Linked in Vitro and in Vivo to Gcn5-Mediated Acetylation at Lysine 14. *Mol. Cell* **2000**, *5* (6), 917–926. https://doi.org/10.1016/s1097-2765(00)80257-9.

(44) Swaney, D. L.; Beltrao, P.; Starita, L.; Guo, A.; Rush, J.; Fields, S.; Krogan, N. J.; Villén, J. Global Analysis of Phosphorylation and Ubiquitylation Cross-Talk in Protein Degradation. *Nat. Methods* **2013**, *10* (7), 676–682. https://doi.org/10.1038/nmeth.2519.

(45) Wu, R.-C.; Feng, Q.; Lonard, D. M.; O'Malley, B. W. SRC-3 Coactivator Functional Lifetime Is Regulated by a Phospho-Dependent Ubiquitin Time Clock. *Cell* **2007**, *129* (6), 1125–1140. https://doi.org/10.1016/j.cell.2007.04.039.

(46) Daujat, S.; Bauer, U.-M.; Shah, V.; Turner, B.; Berger, S.; Kouzarides, T. Crosstalk between CARM1 Methylation and CBP Acetylation on Histone H3. *Curr. Biol. CB* **2002**, *12* (24), 2090–2097. https://doi.org/10.1016/s0960-9822(02)01387-8.

(47) Liu, N.; Yang, R.; Shi, Y.; Chen, L.; Liu, Y.; Wang, Z.; Liu, S.; Ouyang, L.; Wang, H.; Lai, W.; Mao, C.; Wang, M.; Cheng, Y.; Liu, S.; Wang, X.; Zhou, H.; Cao, Y.; Xiao, D.; Tao, Y. The Cross-Talk between Methylation and Phosphorylation in Lymphoid-Specific Helicase Drives Cancer Stem-like Properties. *Signal Transduct. Target. Ther.* **2020**, *5* (1), 1–14. https://doi.org/10.1038/s41392-020-00249-w.

(48) Zhang, C.; Gao, S.; Molascon, A. J.; Wang, Z.; Gorovsky, M. A.; Liu, Y.; Andrews, P. C. Bioinformatic and Proteomic Analysis of Bulk Histones Reveals PTM Crosstalk and Chromatin Features. *J. Proteome Res.* **2014**, *13* (7), 3330–3337. https://doi.org/10.1021/pr5001829.

(49) Smith, L. M.; Kelleher, N. L. Proteoform: A Single Term Describing Protein Complexity. *Nat. Methods* **2013**, *10* (3), 186–187. https://doi.org/10.1038/nmeth.2369.

(50) Pruitt, K. D.; Tatusova, T.; Maglott, D. R. NCBI Reference Sequences (RefSeq): A Curated Non-Redundant Sequence Database of Genomes, Transcripts and Proteins. *Nucleic Acids Res.* **2007**, *35* (Database issue), D61-65. https://doi.org/10.1093/nar/gkl842.

(51) Schlüter, H.; Apweiler, R.; Holzhütter, H.-G.; Jungblut, P. R. Finding One's Way in Proteomics: A Protein Species Nomenclature. *Chem. Cent. J.* **2009**, *3*, 11. https://doi.org/10.1186/1752-153X-3-11.

(52) Horita, H.; Law, A.; Hong, S.; Middleton, K. A Simple Toolset to Identify Endogenous Post-Translational Modifications for a Target Protein: A Snapshot of the EGFR Signaling Pathway. *Biosci. Rep.* **2017**, *37* (4). https://doi.org/10.1042/BSR20170919.

(53) Leney, A. C.; El Atmioui, D.; Wu, W.; Ovaa, H.; Heck, A. J. R. Elucidating Crosstalk Mechanisms between Phosphorylation and O-GlcNAcylation. *Proc. Natl. Acad. Sci.* **2017**, *114* (35), E7255–E7261. https://doi.org/10.1073/pnas.1620529114.

(54) Sharif, S.; Shi, J.; Ruijtenbeek, R.; Pieters, R. J. Study of Cross Talk between Phosphatases and OGA on a ZO-3-Derived Peptide. *Amino Acids* **2019**, *51* (4), 739–743. https://doi.org/10.1007/s00726-019-02699-1.

(55) Shi, J.; Tomašič, T.; Sharif, S.; Brouwer, A. J.; Anderluh, M.; Ruijtenbeek, R.; Pieters, R. J. Peptide Microarray Analysis of the Cross-Talk between O-GlcNAcylation and Tyrosine Phosphorylation. *FEBS Lett.* **2017**, *591* (13), 1872–1883. https://doi.org/10.1002/1873-3468.12708.

(56) Hoffmann, E. de; Stroobant, V. *Mass Spectrometry: Principles and Applications*; John Wiley & Sons, 2007.

(57) Rohner, T. C.; Lion, N.; Girault, H. H. Electrochemical and Theoretical Aspects of Electrospray Ionisation. *Phys. Chem. Chem. Phys.* **2004**, *6* (12), 3056–3068. https://doi.org/10.1039/B316836K.
(58) Taylor, G. R. Disintegration of Water Drops in an Electric Field. *Proc. R. Soc. Math. Phys. Eng. Sci.* **1964**. https://doi.org/10.1098/rspa.1964.0151.
(59) Konermann, L.; Ahadi, E.; Rodriguez, A. D.; Vahidi, S. Unraveling the Mechanism of Electrospray Ionization. *Anal. Chem.* **2013**, *85* (1), 2–9. https://doi.org/10.1021/ac302789c.
(60) Beveridge, R.; Covill, S.; Pacholarz, K. J.; Kalapothakis, J. M. D.; MacPhee, C. E.; Barran, P. E. A Mass-Spectrometry-Based Framework To Define the Extent of Disorder in Proteins. *Anal. Chem.* **2014**, *86* (22), 10979–10991. https://doi.org/10.1021/ac5027435.
(61) Silva, A. M. N.; Vitorino, R.; Domingues, M. R. M.; Spickett, C. M.; Domingues, P. Post-Translational Modifications and Mass Spectrometry Detection. *Free Radic. Biol. Med.* **2013**, *65*, 925–941. https://doi.org/10.1016/j.freeradbiomed.2013.08.184.
(62) Brodbelt, J. S. Ion Activation Methods for Peptides and Proteins. *Anal. Chem.* **2016**, *88* (1), 30–51. https://doi.org/10.1021/acs.analchem.5b04563.
(63) McLafferty, F. W.; Bryce, T. A. Metastable-Ion Characteristics: Characterization of Isomeric Molecules. *Chem. Commun. Lond.* **1967**, No. 23, 1215–1217. https://doi.org/10.1039/C19670001215.
(64) Tsaprailis, G.; Nair, H.; Somogyi, Á.; Wysocki, V. H.; Zhong, W.; Futrell, J. H.; Summerfield, S. G.; Gaskell, S. J. Influence of Secondary Structure on the Fragmentation of Protonated Peptides. *J. Am. Chem. Soc.* **1999**, *121* (22), 5142–5154. https://doi.org/10.1021/ja982980h.
(65) McCormack, A. L.; Somogyi, A.; Dongré, A. R.; Wysocki, V. H. Fragmentation of Protonated Peptides: Surface-Induced Dissociation in Conjunction with a Quantum Mechanical Approach. *Anal. Chem.* **1993**, *65* (20), 2859–2872. https://doi.org/10.1021/ac00068a024.
(66) Kim, M.-S.; Zhong, J.; Pandey, A. Common Errors in Mass Spectrometry-Based Analysis of Post-Translational Modifications. *Proteomics* **2016**, *16* (5), 700–714. https://doi.org/10.1002/pmic.201500355.
(67) Mitchell Wells, J.; McLuckey, S. A. Collision-Induced Dissociation (CID) of Peptides and Proteins. In *Methods in Enzymology*; Biological Mass Spectrometry; Academic Press, 2005; Vol. 402, pp 148–185. https://doi.org/10.1016/S0076-6879(05)02005-7.
(68) Hauser, A.; Penkert, M.; Hackenberger, C. P. R. Chemical Approaches to Investigate Labile Peptide and Protein Phosphorylation. *Acc. Chem. Res.* **2017**, *50* (8), 1883–1893. https://doi.org/10.1021/acs.accounts.7b00170.
(69) Martin, D. B.; Eng, J. K.; Nesvizhskii, A. I.; Gemmill, A.; Aebersold, R. Investigation of Neutral Loss during Collision Induced Dissociation of Peptide Ions. *Anal. Chem.* **2005**, *77* (15), 4870–4882. https://doi.org/10.1021/ac050701k.
(70) Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process. *J. Am. Chem. Soc.* **1998**, *120* (13), 3265–3266. https://doi.org/10.1021/ja973478k.
(71) Jones, A. W.; Cooper, H. J. Dissociation Techniques in Mass Spectrometry-Based Proteomics. *Analyst* **2011**, *136* (17), 3419–3429. https://doi.org/10.1039/C0AN01011A.

(72) Simons, J. Mechanisms for S–S and N–Cα Bond Cleavage in Peptide ECD and ETD Mass Spectrometry. *Chem. Phys. Lett.* **2010**, *484* (4), 81–95. https://doi.org/10.1016/j.cplett.2009.10.062.
(73) Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. Peptide and Protein Sequence Analysis by Electron Transfer Dissociation Mass Spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (26), 9528–9533. https://doi.org/10.1073/pnas.0402700101.
(74) Stensballe, A.; Jensen, O. N.; Olsen, J. V.; Haselmann, K. F.; Zubarev, R. A. Electron Capture Dissociation of Singly and Multiply Phosphorylated Peptides. *Rapid Commun. Mass Spectrom. RCM* **2000**, *14* (19), 1793–1800. https://doi.org/10.1002/1097-0231(20001015)14:19<1793::AID-RCM95>3.0.CO;2-Q.
(75) Guan, Z. Identification and Localization of the Fatty Acid Modification in Ghrelin by Electron Capture Dissociation. *J. Am. Soc. Mass Spectrom.* **2002**, *13* (12), 1443–1447. https://doi.org/10.1016/S1044-0305(02)00707-9.
(76) Cooper, H. J. Investigation of the Presence of b Ions in Electron Capture Dissociation Mass Spectra. *J. Am. Soc. Mass Spectrom.* **2005**, *16* (12), 1932–1940. https://doi.org/10.1016/j.jasms.2005.07.014.
(77) Cooper, H. J.; Tatham, M. H.; Jaffray, E.; Heath, J. K.; Lam, T. T.; Marshall, A. G.; Hay, R. T. Fourier Transform Ion Cyclotron Resonance Mass Spectrometry for the Analysis of Small Ubiquitin-like Modifier (SUMO) Modification: Identification of Lysines in RanBP2 and SUMO Targeted for Modification during the E3 AutoSUMOylation Reaction. *Anal. Chem.* **2005**, *77* (19), 6310–6319. https://doi.org/10.1021/ac058019d.
(78) Chi, A.; Bai, D. L.; Geer, L. Y.; Shabanowitz, J.; Hunt, D. F. Analysis of Intact Proteins on a Chromatographic Time Scale by Electron Transfer Dissociation Tandem Mass Spectrometry. *Int. J. Mass Spectrom.* **2007**, *259* (1–3), 197–203. https://doi.org/10.1016/j.ijms.2006.09.030.
(79) Riley, N. M.; Sikora, J. W.; Seckler, H. S.; Greer, J. B.; Fellers, R. T.; LeDuc, R. D.; Westphall, M. S.; Thomas, P. M.; Kelleher, N. L.; Coon, J. J. The Value of Activated Ion Electron Transfer Dissociation for High-Throughput Top-Down Characterization of Intact Proteins. *Anal. Chem.* **2018**, *90* (14), 8553–8560. https://doi.org/10.1021/acs.analchem.8b01638.
(80) Antoine, R.; Dugourd, P. Visible and Ultraviolet Spectroscopy of Gas Phase Protein Ions. *Phys. Chem. Chem. Phys. PCCP* **2011**, *13* (37), 16494–16509. https://doi.org/10.1039/c1cp21531k.
(81) Brodbelt, J. S.; Morrison, L. J.; Santos, I. Ultraviolet Photodissociation Mass Spectrometry for Analysis of Biological Molecules. *Chem. Rev.* **2020**, *120* (7), 3328–3380. https://doi.org/10.1021/acs.chemrev.9b00440.
(82) Halim, M. A.; MacAleese, L.; Lemoine, J.; Antoine, R.; Dugourd, P.; Girod, M. Ultraviolet, Infrared, and High-Low Energy Photodissociation of Post-Translationally Modified Peptides. *J. Am. Soc. Mass Spectrom.* **2018**, *29* (2), 270–283. https://doi.org/10.1007/s13361-017-1794-9.
(83) Greer, S. M.; Sidoli, S.; Coradin, M.; Jespersen, M. S.; Schwämmle, V.; Jensen, O. N.; Garcia, B. A.; Brodbelt, J. S. Extensive Characterization of Heavily Modified Histone Tails by 193 Nm Ultraviolet Photodissociation Mass Spectrometry via a Middle-Down Strategy. *Anal. Chem.* **2018**, *90* (17), 10425–10433. https://doi.org/10.1021/acs.analchem.8b02320.

(84) Greer, S. M.; Brodbelt, J. S. Top-Down Characterization of Heavily Modified Histones Using 193 Nm Ultraviolet Photodissociation Mass Spectrometry. *J. Proteome Res.* **2018**, *17* (3), 1138–1145. https://doi.org/10.1021/acs.jproteome.7b00801.
(85) Reiding, K. R.; Bondt, A.; Franc, V.; Heck, A. J. R. The Benefits of Hybrid Fragmentation Methods for Glycoproteomics. *TrAC Trends Anal. Chem.* **2018**, *108*, 260–268. https://doi.org/10.1016/j.trac.2018.09.007.
(86) Swaney, D. L.; McAlister, G. C.; Wirtala, M.; Schwartz, J. C.; Syka, J. E. P.; Coon, J. J. Supplemental Activation Method for High-Efficiency Electron-Transfer Dissociation of Doubly Protonated Peptide Precursors. *Anal. Chem.* **2007**, *79* (2), 477–485. https://doi.org/10.1021/ac061457f.
(87) Frese, C. K.; Altelaar, A. F. M.; van den Toorn, H.; Nolting, D.; Griep-Raming, J.; Heck, A. J. R.; Mohammed, S. Toward Full Peptide Sequence Coverage by Dual Fragmentation Combining Electron-Transfer and Higher-Energy Collision Dissociation Tandem Mass Spectrometry. *Anal. Chem.* **2012**, *84* (22), 9668–9673. https://doi.org/10.1021/ac3025366.
(88) Frese, C. K.; Zhou, H.; Taus, T.; Altelaar, A. F. M.; Mechtler, K.; Heck, A. J. R.; Mohammed, S. Unambiguous Phosphosite Localization Using Electron-Transfer/Higher-Energy Collision Dissociation (EThcD). *J. Proteome Res.* **2013**, *12* (3), 1520–1525. https://doi.org/10.1021/pr301130k.
(89) Beausoleil, S. A.; Jedrychowski, M.; Schwartz, D.; Elias, J. E.; Villén, J.; Li, J.; Cohn, M. A.; Cantley, L. C.; Gygi, S. P. Large-Scale Characterization of HeLa Cell Nuclear Phosphoproteins. *Proc. Natl. Acad. Sci.* **2004**, *101* (33), 12130–12135. https://doi.org/10.1073/pnas.0404720101.
(90) Papanastasiou, D.; Kounadis, D.; Lekkas, A.; Orfanopoulos, I.; Mpozatzidis, A.; Smyrnakis, A.; Panagiotopoulos, E.; Kosmopoulou, M.; Reinhardt-Szyba, M.; Fort, K.; Makarov, A.; Zubarev, R. A. The Omnitrap Platform: A Versatile Segmented Linear Ion Trap for Multidimensional Multiple-Stage Tandem Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2022**, *33* (10), 1990–2007. https://doi.org/10.1021/jasms.2c00214.
(91) Zubarev, R. A.; Makarov, A. Orbitrap Mass Spectrometry. *Anal. Chem.* **2013**, *85* (11), 5288–5296. https://doi.org/10.1021/ac4001223.
(92) Savaryn, J. P.; Toby, T. K.; Kelleher, N. L. A Researcher's Guide to Mass Spectrometry-Based Proteomics. *PROTEOMICS* **2016**, *16* (18), 2435–2443. https://doi.org/10.1002/pmic.201600113.
(93) Wolfgang, P.; Helmut, S. Apparatus for Separating Charged Particles of Different Specific Charges. US2939952A, June 7, 1960.
(94) *Quadrupole Mass Analyzer - an overview | ScienceDirect Topics*. https://www.sciencedirect.com/topics/medicine-and-dentistry/quadrupole-mass-analyzer (accessed 2022-12-01).
(95) *Discover the Advantage of High-Resolution Quadrupole Mass Spectrometry*. AZoM.com. https://www.azom.com/article.aspx?ArticleID=16511 (accessed 2022-12-01).
(96) *Mass Spectrometry: Principles and Applications, 3rd Edition | Wiley*. Wiley.com. https://www.wiley.com/en-gb/Mass+Spectrometry%3A+Principles+and+Applications%2C+3rd+Edition-p-9780470033104 (accessed 2022-11-25).

(97) *Introduction to Mass Spectrometry: Instrumentation, Applications, and Strategies for Data Interpretation, 4th Edition | Wiley*. Wiley.com. https://www.wiley.com/en-gb/Introduction+to+Mass+Spectrometry%3A+Instrumentation%2C+Applications%2C+and+Strategies+for+Data+Interpretation%2C+4th+Edition-p-9780470516348 (accessed 2022-12-01).
(98) Schwartz, J. C.; Senko, M. W.; Syka, J. E. P. A Two-Dimensional Quadrupole Ion Trap Mass Spectrometer. *J. Am. Soc. Mass Spectrom.* **2002**, *13* (6), 659–669. https://doi.org/10.1016/S1044-0305(02)00384-7.
(99) Jabbour, R. E.; Snyder, A. P. 14 - Mass Spectrometry-Based Proteomics Techniques for Biological Identification**This Chapter Has Been Written by Two Employees of the US Army Research, Development and Engineering Command (RDECOM), Operated by the Edgewood Chemical Biological Center for the US Department of Defense-Army. In *Biological Identification*; Schaudies, R. P., Ed.; Woodhead Publishing, 2014; pp 370–430. https://doi.org/10.1533/9780857099167.4.370.
(100) Kaufmann, A.; Teale, P. Capabilities and Limitations of High-Resolution Mass Spectrometry (HRMS): Time-of-Flight and Orbitrap™; 2016; pp 93–139. https://doi.org/10.1002/9781118696781.ch3.
(101) Kendra. *Advantages of Time-of-Flight Mass Spectrometry Over Quadrupole MS*. TOFWERK. https://www.tofwerk.com/advantages-time-of-flight-mass-spectrometry-over-quadrupole-ms/ (accessed 2022-12-01).
(102) Hopfgartner, G.; Varesio, E.; Tschäppät, V.; Grivet, C.; Bourgogne, E.; Leuthold, L. A. Triple Quadrupole Linear Ion Trap Mass Spectrometer for the Analysis of Small Molecules and Macromolecules. *J. Mass Spectrom.* **2004**, *39* (8), 845–855. https://doi.org/10.1002/jms.659.
(103) Beaudette, P.; Bateman, K. P. Discovery Stage Pharmacokinetics Using Dried Blood Spots. *J. Chromatogr. B* **2004**, *809* (1), 153–158. https://doi.org/10.1016/j.jchromb.2004.06.018.
(104) Anderson, L.; Hunter, C. L. Quantitative Mass Spectrometric Multiple Reaction Monitoring Assays for Major Plasma Proteins. *Mol. Cell. Proteomics MCP* **2006**, *5* (4), 573–588. https://doi.org/10.1074/mcp.M500331-MCP200.
(105) Moreras-Marti, A.; Fox-Powell, M.; Stueeken, E.; Di Rocco, T.; Galloway, T.; Osinski, G. R.; Cousins, C. R.; Zerkle, A. L. Quadruple Sulfur Isotope Biosignatures from Terrestrial Mars Analogue Systems. *Geochim. Cosmochim. Acta* **2021**, *308*, 157–172. https://doi.org/10.1016/j.gca.2021.06.007.
(106) Li, C.; Chu, S.; Tan, S.; Yin, X.; Jiang, Y.; Dai, X.; Gong, X.; Fang, X.; Tian, D. Towards Higher Sensitivity of Mass Spectrometry: A Perspective From the Mass Analyzers. *Front. Chem.* **2021**, *9*.
(107) Ganten, D.; Ruckpaul, K. *Encyclopedic Reference of Genomics and Proteomics in Molecular Medicine*; Ganten, D., Ruckpaul, K., Eds.; Springer: Berlin [u.a.], 2006.
(108) Senko, M. W.; Remes, P. M.; Canterbury, J. D.; Mathur, R.; Song, Q.; Eliuk, S. M.; Mullen, C.; Earley, L.; Hardman, M.; Blethrow, J. D.; Bui, H.; Specht, A.; Lange, O.; Denisov, E.; Makarov, A.; Horning, S.; Zabrouskov, V. Novel Parallelized Quadrupole/Linear Ion Trap/Orbitrap Tribrid Mass Spectrometer Improving Proteome Coverage and Peptide Identification Rates. *Anal. Chem.* **2013**, *85* (24), 11710–11714. https://doi.org/10.1021/ac403115c.
(109) Yu, Q.; Paulo, J. A.; Naverrete-Perea, J.; McAlister, G. C.; Canterbury, J. D.; Bailey, D. J.; Robitaille, A. M.; Huguet, R.; Zabrouskov, V.; Gygi, S. P.; Schweppe, D. K. Benchmarking the Orbitrap Tribrid Eclipse for Next Generation Multiplexed

Proteomics. *Anal. Chem.* **2020**, *92* (9), 6478–6485. https://doi.org/10.1021/acs.analchem.9b05685.
(110) Winter, D. L.; Wilkins, M. R.; Donald, W. A. Differential Ion Mobility–Mass Spectrometry for Detailed Analysis of the Proteome. *Trends Biotechnol.* **2019**, *37* (2), 198–213. https://doi.org/10.1016/j.tibtech.2018.07.018.
(111) Zhao, H.; Creese, A. J.; Cooper, H. J. Online LC-FAIMS-MS/MS for the Analysis of Phosphorylation in Proteins. In *Phospho-Proteomics: Methods and Protocols*; von Stechow, L., Ed.; Methods in Molecular Biology; Springer: New York, NY, 2016; pp 241–250. https://doi.org/10.1007/978-1-4939-3049-4_16.
(112) Adoni, K. R.; Cunningham, D. L.; Heath, J. K.; Leney, A. C. FAIMS Enhances the Detection of PTM Crosstalk Sites. *J. Proteome Res.* **2022**, acs.jproteome.1c00721. https://doi.org/10.1021/acs.jproteome.1c00721.
(113) Gomez, J. D.; Ridgeway, M. E.; Park, M. A.; Fritz, K. S. Utilizing Ion Mobility to Identify Isobaric Post-Translational Modifications: Resolving Acrolein and Propionyl Lysine Adducts by TIMS Mass Spectrometry. *Int. J. Ion Mobil. Spectrom. Off. Publ. Int. Soc. Ion Mobil. Spectrom.* **2018**, *21* (3), 65–69. https://doi.org/10.1007/s12127-018-0237-z.
(114) Delafield, D. G.; Lu, G.; Kaminsky, C. J.; Li, L. High-End Ion Mobility Mass Spectrometry: A Current Review of Analytical Capacity in Omics Applications and Structural Investigations. *TrAC Trends Anal. Chem.* **2022**, *157*, 116761. https://doi.org/10.1016/j.trac.2022.116761.
(115) Adoni, K. R.; Cunningham, D. L.; Heath, J. K.; Leney, A. C. FAIMS Enhances the Detection of PTM Crosstalk Sites. *J. Proteome Res.* **2022**, *21* (4), 930–939. https://doi.org/10.1021/acs.jproteome.1c00721.
(116) Lanucara, F.; Holman, S. W.; Gray, C. J.; Eyers, C. E. The Power of Ion Mobility-Mass Spectrometry for Structural Characterization and the Study of Conformational Dynamics. *Nat. Chem.* **2014**, *6* (4), 281–294. https://doi.org/10.1038/nchem.1889.
(117) Bohrer, B. C.; Merenbloom, S. I.; Koeniger, S. L.; Hilderbrand, A. E.; Clemmer, D. E. Biomolecule Analysis by Ion Mobility Spectrometry. *Annu. Rev. Anal. Chem. Palo Alto Calif* **2008**, *1*, 293–327. https://doi.org/10.1146/annurev.anchem.1.031207.113001.
(118) Cumeras, R.; Figueras, E.; Davis, C. E.; Baumbach, J. I.; Gràcia, I. Review on Ion Mobility Spectrometry. Part 1: Current Instrumentation. *Analyst* **2015**, *140* (5), 1376–1390. https://doi.org/10.1039/C4AN01100G.
(119) Giles, K.; Pringle, S. D.; Worthington, K. R.; Little, D.; Wildgoose, J. L.; Bateman, R. H. Applications of a Travelling Wave-Based Radio-Frequency-Only Stacked Ring Ion Guide. *Rapid Commun. Mass Spectrom. RCM* **2004**, *18* (20), 2401–2414. https://doi.org/10.1002/rcm.1641.
(120) Guevremont, R. High-Field Asymmetric Waveform Ion Mobility Spectrometry: A New Tool for Mass Spectrometry. *J. Chromatogr. A* **2004**, *1058* (1–2), 3–19.
(121) Giles, K.; Wildgoose, J.; Pringle, S.; Garside, J.; Carney, P.; Nixon, P.; Langridge, D. DESIGN AND UTILITY OF A MULTI-PASS CYCLIC ION MOBILITY SEPARATOR; 2014; p 1.
(122) May, J. C.; McLean, J. A. The Influence of Drift Gas Composition on the Separation Mechanism in Traveling Wave Ion Mobility Spectrometry: Insight from Electrodynamic Simulations. *Int. J. Ion Mobil. Spectrom. Off. Publ. Int. Soc. Ion Mobil. Spectrom.* **2003**, *16* (2), 85–94. https://doi.org/10.1007/s12127-013-0123-7.

(123) Meier, F.; Beck, S.; Grassl, N.; Lubeck, M.; Park, M. A.; Raether, O.; Mann, M. Parallel Accumulation–Serial Fragmentation (PASEF): Multiplying Sequencing Speed and Sensitivity by Synchronized Scans in a Trapped Ion Mobility Device. *J. Proteome Res.* **2015**, *14* (12), 5378–5387. https://doi.org/10.1021/acs.jproteome.5b00932.
(124) Giles, K.; Ujma, J.; Wildgoose, J.; Pringle, S.; Richardson, K.; Langridge, D.; Green, M. A Cyclic Ion Mobility-Mass Spectrometry System. *Anal. Chem.* **2019**, *91* (13), 8564–8573. https://doi.org/10.1021/acs.analchem.9b01838.
(125) Shvartsburg, A. A. *Differential Ion Mobility Spectrometry: Nonlinear Ion Transport and Fundamentals of FAIMS*; CRC Press: Boca Raton, 2008. https://doi.org/10.1201/9781420051070.
(126) Buryakov, I. A.; Krylov, E. V.; Nazarov, E. G.; Rasulev, U. Kh. A New Method of Separation of Multi-Atomic Ions by Mobility at Atmospheric Pressure Using a High-Frequency Amplitude-Asymmetric Strong Electric Field. *Int. J. Mass Spectrom. Ion Process.* **1993**, *128* (3), 143–148. https://doi.org/10.1016/0168-1176(93)87062-W.
(127) Shvartsburg, A. A.; Haris, A.; Andrzejewski, R.; Entwistle, A.; Giles, R. Differential Ion Mobility Separations in the Low-Pressure Regime. *Anal. Chem.* **2018**, *90* (1), 936–943. https://doi.org/10.1021/acs.analchem.7b03925.
(128) Toutoungi, D. D. UltraFAIMS: A New Dimension in Mass Spectrometry. 15.
(129) Barnett, D. A.; Ells, B.; Guevremont, R.; Purves, R. W. Separation of Leucine and Isoleucine by Electrospray Ionization-High Field Asymmetric Waveform Ion Mobility Spectrometry-Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **1999**, *10* (12), 1279–1284. https://doi.org/10.1016/S1044-0305(99)00098-7.
(130) Krylov, E. V. A Method of Reducing Diffusion Losses in a Drift Spectrometer. *Tech. Phys.* **1999**, *44* (1), 113–116. https://doi.org/10.1134/1.1259263.
(131) Swearingen, K. E.; Hoopmann, M. R.; Johnson, R. S.; Saleem, R. A.; Aitchison, J. D.; Moritz, R. L. Nanospray FAIMS Fractionation Provides Significant Increases in Proteome Coverage of Unfractionated Complex Protein Digests. *Mol. Cell. Proteomics MCP* **2012**, *11* (4), M111.014985. https://doi.org/10.1074/mcp.M111.014985.
(132) Thermo Fisher Scientific. *FAIMS Pro™ Interface*. Thermo Fisher. https://www.thermofisher.com/order/catalog/product/FMS02-10001 (accessed 2022-11-29).
(133) Hebert, A. S.; Prasad, S.; Belford, M. W.; Bailey, D. J.; McAlister, G. C.; Abbatiello, S. E.; Huguet, R.; Wouters, E. R.; Dunyach, J.-J.; Brademan, D. R.; Westphall, M. S.; Coon, J. J. Comprehensive Single-Shot Proteomics with FAIMS on a Hybrid Orbitrap Mass Spectrometer. *Anal. Chem.* **2018**, *90* (15), 9529–9537. https://doi.org/10.1021/acs.analchem.8b02233.
(134) Wolters, D. A.; Washburn, M. P.; Yates, J. R. An Automated Multidimensional Protein Identification Technology for Shotgun Proteomics. *Anal. Chem.* **2001**, *73* (23), 5683–5690. https://doi.org/10.1021/ac010617e.
(135) Eng, J. K.; McCormack, A. L.; Yates, J. R. An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. *J. Am. Soc. Mass Spectrom.* **1994**, *5* (11), 976–989. https://doi.org/10.1016/1044-0305(94)80016-2.
(136) Hoffman, M. D.; Kast, J. Mass Spectrometric Characterization of Lipid-Modified Peptides for the Analysis of Acylated Proteins. *J. Mass Spectrom. JMS* **2006**, *41* (2), 229–241. https://doi.org/10.1002/jms.981.

(137) Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M.-C.; Yates, J. R. Protein Analysis by Shotgun/Bottom-up Proteomics. *Chem. Rev.* **2013**, *113* (4), 2343–2394. https://doi.org/10.1021/cr3003533.
(138) Fernández-Costa, C.; Martínez-Bartolomé, S.; McClatchy, D. B.; Saviola, A. J.; Yu, N.-K.; Yates, J. R. I. Impact of the Identification Strategy on the Reproducibility of the DDA and DIA Results. *J. Proteome Res.* **2020**, *19* (8), 3153–3161. https://doi.org/10.1021/acs.jproteome.0c00153.
(139) Yang, Y.; Qiao, L. Data-Independent Acquisition Proteomics Methods for Analyzing Post-Translational Modifications. *PROTEOMICS n/a* (n/a), 2200046. https://doi.org/10.1002/pmic.202200046.
(140) Hansen, F. M.; Tanzer, M. C.; Brüning, F.; Bludau, I.; Stafford, C.; Schulman, B. A.; Robles, M. S.; Karayel, O.; Mann, M. Data-Independent Acquisition Method for Ubiquitinome Analysis Reveals Regulation of Circadian Biology. *Nat. Commun.* **2021**, *12* (1), 254. https://doi.org/10.1038/s41467-020-20509-1.
(141) Hornbeck, P. V.; Zhang, B.; Murray, B.; Kornhauser, J. M.; Latham, V.; Skrzypek, E. PhosphoSitePlus, 2014: Mutations, PTMs and Recalibrations. *Nucleic Acids Res.* **2015**, *43* (Database issue), D512-520. https://doi.org/10.1093/nar/gku1267.
(142) Huang, H.; Arighi, C. N.; Ross, K. E.; Ren, J.; Li, G.; Chen, S.-C.; Wang, Q.; Cowart, J.; Vijay-Shanker, K.; Wu, C. H. IPTMnet: An Integrated Resource for Protein Post-Translational Modification Network Discovery. *Nucleic Acids Res.* **2018**, *46* (D1), D542–D550. https://doi.org/10.1093/nar/gkx1104.
(143) Adamczyk, M.; Gebler, J. C.; Wu, J. Selective Analysis of Phosphopeptides within a Protein Mixture by Chemical Modification, Reversible Biotinylation and Mass Spectrometry. *Rapid Commun. Mass Spectrom. RCM* **2001**, *15* (16), 1481–1488. https://doi.org/10.1002/rcm.394.
(144) Kim, S. C.; Sprung, R.; Chen, Y.; Xu, Y.; Ball, H.; Pei, J.; Cheng, T.; Kho, Y.; Xiao, H.; Xiao, L.; Grishin, N. V.; White, M.; Yang, X.-J.; Zhao, Y. Substrate and Functional Diversity of Lysine Acetylation Revealed by a Proteomics Survey. *Mol. Cell* **2006**, *23* (4), 607–618. https://doi.org/10.1016/j.molcel.2006.06.026.
(145) Ong, S.-E.; Mittler, G.; Mann, M. Identifying and Quantifying in Vivo Methylation Sites by Heavy Methyl SILAC. *Nat. Methods* **2004**, *1* (2), 119–126. https://doi.org/10.1038/nmeth715.
(146) Fulzele, A.; Bennett, E. J. Ubiquitin DiGLY Proteomics as an Approach to Identify and Quantify the Ubiquitin-Modified Proteome. *Methods Mol. Biol. Clifton NJ* **2018**, *1844*, 363–384. https://doi.org/10.1007/978-1-4939-8706-1_23.
(147) Brandi, J.; Noberini, R.; Bonaldi, T.; Cecconi, D. Advances in Enrichment Methods for Mass Spectrometry-Based Proteomics Analysis of Post-Translational Modifications. *J. Chromatogr. A* **2022**, *1678*, 463352. https://doi.org/10.1016/j.chroma.2022.463352.
(148) Mertins, P.; Qiao, J. W.; Patel, J.; Udeshi, N. D.; Clauser, K. R.; Mani, D. R.; Burgess, M. W.; Gillette, M. A.; Jaffe, J. D.; Carr, S. A. Integrated Proteomic Analysis of Post-Translational Modifications by Serial Enrichment. *Nat. Methods* **2013**, *10* (7), 634–637. https://doi.org/10.1038/nmeth.2518.
(149) Thingholm, T. E.; Jensen, O. N.; Robinson, P. J.; Larsen, M. R. SIMAC (Sequential Elution from IMAC), a Phosphoproteomics Strategy for the Rapid Separation of Monophosphorylated from Multiply Phosphorylated Peptides *. *Mol. Cell. Proteomics* **2008**, *7* (4), 661–671. https://doi.org/10.1074/mcp.M700362-MCP200.

(150) Korkuć, P.; Walther, D. Towards Understanding the Crosstalk between Protein Post-Translational Modifications: Homo- and Heterotypic PTM Pair Distances on Protein Surfaces Are Not Random. *Proteins Struct. Funct. Bioinforma.* **2017**, *85* (1), 78–92. https://doi.org/10.1002/prot.25200.
(151) Boone, C.; Adamec, J. 10 - Top-Down Proteomics. In *Proteomic Profiling and Analytical Chemistry (Second Edition)*; Ciborowski, P., Silberring, J., Eds.; Elsevier: Boston, 2016; pp 175–191. https://doi.org/10.1016/B978-0-444-63688-1.00010-0.
(152) Tian, Z.; Zhao, R.; Tolić, N.; Moore, R. J.; Stenoien, D. L.; Robinson, E. W.; Smith, R. D.; Paša-Tolić, L. Two-Dimensional Liquid Chromatography System for Online Top-down Mass Spectrometry. *PROTEOMICS* **2010**, *10* (20), 3610–3620. https://doi.org/10.1002/pmic.201000367.
(153) Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, J. R. Direct Analysis of Protein Complexes Using Mass Spectrometry. *Nat. Biotechnol.* **1999**, *17* (7), 676–682. https://doi.org/10.1038/10890.
(154) McDonald, W. H.; Ohi, R.; Miyamoto, D. T.; Mitchison, T. J.; Yates, J. R. Comparison of Three Directly Coupled HPLC MS/MS Strategies for Identification of Proteins from Complex Mixtures: Single-Dimension LC-MS/MS, 2-Phase MudPIT, and 3-Phase MudPIT. *Int. J. Mass Spectrom.* **2002**, *219* (1), 245–251. https://doi.org/10.1016/S1387-3806(02)00563-8.
(155) Ankney, J. A.; Muneer, A.; Chen, X. Relative and Absolute Quantitation in Mass Spectrometry-Based Proteomics. *Annu. Rev. Anal. Chem. Palo Alto Calif* **2018**, *11* (1), 49–77. https://doi.org/10.1146/annurev-anchem-061516-045357.
(156) Matic, I.; Jaffray, E. G.; Oxenham, S. K.; Groves, M. J.; Barratt, C. L. R.; Tauro, S.; Stanley-Wall, N. R.; Hay, R. T. Absolute SILAC-Compatible Expression Strain Allows Sumo-2 Copy Number Determination in Clinical Samples. *J. Proteome Res.* **2011**, *10* (10), 4869–4875. https://doi.org/10.1021/pr2004715.
(157) MacLeod, A. K.; Zang, T.; Riches, Z.; Henderson, C. J.; Wolf, C. R.; Huang, J. T.-J. A Targeted in Vivo SILAC Approach for Quantification of Drug Metabolism Enzymes: Regulation by the Constitutive Androstane Receptor. *J. Proteome Res.* **2014**, *13* (2), 866–874. https://doi.org/10.1021/pr400897t.
(158) Phillips, N. J.; Steichen, C. T.; Schilling, B.; Post, D. M. B.; Niles, R. K.; Bair, T. B.; Falsetta, M. L.; Apicella, M. A.; Gibson, B. W. Proteomic Analysis of Neisseria Gonorrhoeae Biofilms Shows Shift to Anaerobic Respiration and Changes in Nutrient Transport and Outermembrane Proteins. *PLOS ONE* **2012**, *7* (6), e38303. https://doi.org/10.1371/journal.pone.0038303.
(159) Old, W. M.; Meyer-Arendt, K.; Aveline-Wolf, L.; Pierce, K. G.; Mendoza, A.; Sevinsky, J. R.; Resing, K. A.; Ahn, N. G. Comparison of Label-Free Methods for Quantifying Human Proteins by Shotgun Proteomics. *Mol. Cell. Proteomics MCP* **2005**, *4* (10), 1487–1502. https://doi.org/10.1074/mcp.M500084-MCP200.
(160) Liu, H.; Sadygov, R. G.; Yates, J. R. A Model for Random Sampling and Estimation of Relative Protein Abundance in Shotgun Proteomics. *Anal. Chem.* **2004**, *76* (14), 4193–4201. https://doi.org/10.1021/ac0498563.
(161) Zhang, Y.; Wen, Z.; Washburn, M. P.; Florens, L. Effect of Dynamic Exclusion Duration on Spectral Count Based Quantitative Proteomics. *Anal. Chem.* **2009**, *81* (15), 6317–6326. https://doi.org/10.1021/ac9004887.
(162) Lu, P.; Vogel, C.; Wang, R.; Yao, X.; Marcotte, E. M. Absolute Protein Expression Profiling Estimates the Relative Contributions of Transcriptional and Translational Regulation. *Nat. Biotechnol.* **2007**, *25* (1), 117–124. https://doi.org/10.1038/nbt1270.

(163) Zybailov, B.; Mosley, A. L.; Sardiu, M. E.; Coleman, M. K.; Florens, L.; Washburn, M. P. Statistical Analysis of Membrane Proteome Expression Changes in Saccharomyces Cerevisiae. *J. Proteome Res.* **2006**, *5* (9), 2339–2347. https://doi.org/10.1021/pr060161n.
(164) Nahnsen, S.; Bielow, C.; Reinert, K.; Kohlbacher, O. Tools for Label-Free Peptide Quantification. *Mol. Cell. Proteomics MCP* **2013**, *12* (3), 549–556. https://doi.org/10.1074/mcp.R112.025163.
(165) Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. Quantitative Analysis of Complex Protein Mixtures Using Isotope-Coded Affinity Tags. *Nat. Biotechnol.* **1999**, *17* (10), 994–999. https://doi.org/10.1038/13690.
(166) Werner, T.; Becher, I.; Sweetman, G.; Doce, C.; Savitski, M. M.; Bantscheff, M. High-Resolution Enabled TMT 8-Plexing. *Anal. Chem.* **2012**, *84* (16), 7188–7194. https://doi.org/10.1021/ac301553x.
(167) Yu, K.; Wang, Z.; Wu, Z.; Tan, H.; Mishra, A.; Peng, J. High-Throughput Profiling of ProteomeModificationsProteomesand PosttranslationalModificationsposttranslational Modifications by 16-Plex TMT Labeling and Mass SpectrometryMassspectrometry. In *Quantitative Methods in Proteomics*; Marcus, K., Eisenacher, M., Sitek, B., Eds.; Methods in Molecular Biology; Springer US: New York, NY, 2021; pp 205–224. https://doi.org/10.1007/978-1-0716-1024-4_15.
(168) Ow, S. Y.; Salim, M.; Noirel, J.; Evans, C.; Rehman, I.; Wright, P. C. ITRAQ Underestimation in Simple and Complex Mixtures: "The Good, the Bad and the Ugly." *J. Proteome Res.* **2009**, *8* (11), 5347–5355. https://doi.org/10.1021/pr900634c.
(169) Karp, N. A.; Huber, W.; Sadowski, P. G.; Charles, P. D.; Hester, S. V.; Lilley, K. S. Addressing Accuracy and Precision Issues in ITRAQ Quantitation. *Mol. Cell. Proteomics MCP* **2010**, *9* (9), 1885–1897. https://doi.org/10.1074/mcp.M900628-MCP200.
(170) Koehler, C. J.; Strozynski, M.; Kozielski, F.; Treumann, A.; Thiede, B. Isobaric Peptide Termini Labeling for MS/MS-Based Quantitative Proteomics. *J. Proteome Res.* **2009**, *8* (9), 4333–4341. https://doi.org/10.1021/pr900425n.
(171) Hsu, J.-L.; Huang, S.-Y.; Chow, N.-H.; Chen, S.-H. Stable-Isotope Dimethyl Labeling for Quantitative Proteomics. *Anal. Chem.* **2003**, *75* (24), 6843–6852. https://doi.org/10.1021/ac0348625.
(172) Chen, X.; Smith, L. M.; Bradbury, E. M. Site-Specific Mass Tagging with Stable Isotopes in Proteins for Accurate and Efficient Protein Identification. *Anal. Chem.* **2000**, *72* (6), 1134–1143. https://doi.org/10.1021/ac9911600.
(173) Merrill, A. E.; Hebert, A. S.; MacGilvray, M. E.; Rose, C. M.; Bailey, D. J.; Bradley, J. C.; Wood, W. W.; El Masri, M.; Westphall, M. S.; Gasch, A. P.; Coon, J. J. NeuCode Labels for Relative Protein Quantification. *Mol. Cell. Proteomics MCP* **2014**, *13* (9), 2503–2512. https://doi.org/10.1074/mcp.M114.040287.
(174) Gk, P.; Ea, V.; Dj, B.; Cm, R.; Ms, W.; As, H.; J, Y.; Jj, C. Neucode Labels for Multiplexed, Absolute Protein Quantification. *Anal. Chem.* **2016**, *88* (6). https://doi.org/10.1021/acs.analchem.5b04773.
(175) Wolf-Yadlin, A.; Hautaniemi, S.; Lauffenburger, D. A.; White, F. M. Multiple Reaction Monitoring for Robust Quantitative Proteomic Analysis of Cellular Signaling Networks. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (14), 5860–5865. https://doi.org/10.1073/pnas.0608638104.

(176) Cuomo, A.; Soldi, M.; Bonaldi, T. SILAC-Based Quantitative Strategies for Accurate Histone Posttranslational Modification Profiling Across Multiple Biological Samples. *Methods Mol. Biol. Clifton NJ* **2017**, *1528*, 97–119. https://doi.org/10.1007/978-1-4939-6630-1_7.
(177) Cunningham, D. L.; Sarhan, A. R.; Creese, A. J.; Larkins, K. P. B.; Zhao, H.; Ferguson, H. R.; Brookes, K.; Marusiak, A. A.; Cooper, H. J.; Heath, J. K. Differential Responses to Kinase Inhibition in FGFR2-Addicted Triple Negative Breast Cancer Cells: A Quantitative Phosphoproteomics Study. *Sci. Rep.* **2020**, *10* (1), 7950. https://doi.org/10.1038/s41598-020-64534-y.
(178) Duan, J.; J. Gaffrey, M.; Qian, W.-J. Quantitative Proteomic Characterization of Redox-Dependent Post-Translational Modifications on Protein Cysteines. *Mol. Biosyst.* **2017**, *13* (5), 816–829. https://doi.org/10.1039/C6MB00861E.
(179) Bremang, M.; Cuomo, A.; Maria Agresta, A.; Stugiewicz, M.; Spadotto, V.; Bonaldi, T. Mass Spectrometry-Based Identification and Characterisation of Lysine and Arginine Methylation in the Human Proteome. *Mol. Biosyst.* **2013**, *9* (9), 2231–2247. https://doi.org/10.1039/C3MB00009E.
(180) Anania, V. G.; Bustos, D. J.; Lill, J. R.; Kirkpatrick, D. S.; Coscoy, L. A Novel Peptide-Based SILAC Method to Identify the Posttranslational Modifications Provides Evidence for Unconventional Ubiquitination in the ER-Associated Degradation Pathway. *Int. J. Proteomics* **2013**, *2013*, 1–12. https://doi.org/10.1155/2013/857918.
(181) Creese, A. J.; Cooper, H. J. Separation and Identification of Isomeric Glycopeptides by High Field Asymmetric Waveform Ion Mobility Spectrometry. *Anal. Chem.* **2012**, *84* (5), 2597–2601. https://doi.org/10.1021/ac203321y.
(182) Purves, R. W.; Guevremont, R. Electrospray Ionization High-Field Asymmetric Waveform Ion Mobility Spectrometry-Mass Spectrometry. *Anal. Chem.* **1999**, *71* (13), 2346–2357. https://doi.org/10.1021/ac981380y.
(183) Cooper, H. J. To What Extent Is FAIMS Beneficial in the Analysis of Proteins? *J. Am. Soc. Mass Spectrom.* **2016**, *27*, 566–577. https://doi.org/10.1007/s13361-015-1326-4.
(184) Guevremont, R.; Barnett, D. A.; Purves, R. W.; Vandermey, J. Analysis of a Tryptic Digest of Pig Hemoglobin Using ESI-FAIMS-MS. *Anal. Chem.* **2000**, *72* (19), 4577–4584. https://doi.org/10.1021/ac0000271.
(185) Barnett, D. A.; Ding, L.; Ells, B.; Purves, R. W.; Guevremont, R. Tandem Mass Spectra of Tryptic Peptides at Signal-to-Background Ratios Approaching Unity Using Electrospray Ionization High-Field Asymmetric Waveform Ion Mobility Spectrometry/Hybrid Quadrupole Time-of-Flight Mass Spectrometry. *Rapid Commun. Mass Spectrom. RCM* **2002**, *16* (7), 676–680. https://doi.org/10.1002/rcm.621.
(186) Venne, K.; Bonneil, E.; Eng, K.; Thibault, P. Improvement in Peptide Detection for Proteomics Analyses Using NanoLC−MS and High-Field Asymmetry Waveform Ion Mobility Mass Spectrometry. *Anal. Chem.* **2005**, *77* (7), 2176–2186. https://doi.org/10.1021/ac048410j.
(187) Canterbury, J. D.; Yi, X.; Hoopmann, M. R.; MacCoss, M. J. Assessing the Dynamic Range and Peak Capacity of Nanoflow LC-FAIMS-MS on an Ion Trap Mass Spectrometer for Proteomics. *Anal. Chem.* **2008**, *80* (18), 6888–6897. https://doi.org/10.1021/ac8004988.

(188) Creese, A. J.; Shimwell, N. J.; Larkins, K. P. B.; Heath, J. K.; Cooper, H. J. Probing the Complementarity of FAIMS and Strong Cation Exchange Chromatography in Shotgun Proteomics. *J. Am. Soc. Mass Spectrom.* **2013**, *24* (3), 431–443. https://doi.org/10.1007/s13361-012-0544-2.
(189) Shvartsburg, A. A.; Zheng, Y.; Smith, R. D.; Kelleher, N. L. Separation of Variant Methylated Histone Tails by Differential Ion Mobility. *Anal. Chem.* **2012**, *84* (15), 6317–6320. https://doi.org/10.1021/ac301541r.
(190) Pathak, P.; Baird, M. A.; Shvartsburg, A. A. High-Resolution Ion Mobility Separations of Isomeric Glycoforms with Variations on the Peptide and Glycan Levels. *J. Am. Soc. Mass Spectrom.* **2020**, *31* (7), 1603–1609. https://doi.org/10.1021/jasms.0c00183.
(191) Baird, M. A.; Shvartsburg, A. A. Localization of Post-Translational Modifications in Peptide Mixtures via High-Resolution Differential Ion Mobility Separations Followed by Electron Transfer Dissociation. *J. Am. Soc. Mass Spectrom.* **2016**, *27* (12), 2064–2070. https://doi.org/10.1007/s13361-016-1498-6.
(192) Cifani, P.; Kentsis, A. Towards Comprehensive and Quantitative Proteomics for Diagnosis and Therapy of Human Disease. *PROTEOMICS* **2017**, *17* (1–2), 1600079. https://doi.org/10.1002/pmic.201600079.
(193) Chen, C.; Hou, J.; Tanner, J. J.; Cheng, J. Bioinformatics Methods for Mass Spectrometry-Based Proteomics Data Analysis. *Int. J. Mol. Sci.* **2020**, *21* (8), 2873. https://doi.org/10.3390/ijms21082873.
(194) Tyanova, S.; Temu, T.; Cox, J. The MaxQuant Computational Platform for Mass Spectrometry-Based Shotgun Proteomics. *Nat. Protoc.* **2016**, *11* (12), 2301–2319. https://doi.org/10.1038/nprot.2016.136.
(195) Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. Probability-Based Protein Identification by Searching Sequence Databases Using Mass Spectrometry Data. *Electrophoresis* **1999**, *20* (18), 3551–3567. https://doi.org/10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2.
(196) Brodbelt, J. S.; Russell, D. H. Focus on the 20-Year Anniversary of SEQUEST. *J. Am. Soc. Mass Spectrom.* **2015**, *26* (11), 1797–1798. https://doi.org/10.1007/s13361-015-1264-1.
(197) Craig, R.; Beavis, R. C. TANDEM: Matching Proteins with Tandem Mass Spectra. *Bioinformatics* **2004**, *20* (9), 1466–1467. https://doi.org/10.1093/bioinformatics/bth092.
(198) Jeong, K.; Kim, S.; Pevzner, P. A. UniNovo: A Universal Tool for de Novo Peptide Sequencing. *Bioinformatics* **2013**, *29* (16), 1953–1962. https://doi.org/10.1093/bioinformatics/btt338.
(199) Frank, A.; Pevzner, P. PepNovo: De Novo Peptide Sequencing via Probabilistic Network Modeling. *Anal. Chem.* **2005**, *77* (4), 964–973. https://doi.org/10.1021/ac048788h.
(200) Ma, B.; Zhang, K.; Hendrie, C.; Liang, C.; Li, M.; Doherty-Kirby, A.; Lajoie, G. PEAKS: Powerful Software for Peptide de Novo Sequencing by Tandem Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2003**, *17* (20), 2337–2342. https://doi.org/10.1002/rcm.1196.
(201) Ramsbottom, K. A.; Prakash, A.; Riverol, Y. P.; Camacho, O. M.; Martin, M.-J.; Vizcaíno, J. A.; Deutsch, E. W.; Jones, A. R. Method for Independent Estimation of the False Localization Rate for Phosphoproteomics. *J. Proteome Res.* **2022**, *21* (7), 1603–1615. https://doi.org/10.1021/acs.jproteome.1c00827.

(202) Dupree, E. J.; Jayathirtha, M.; Yorkey, H.; Mihasan, M.; Petre, B. A.; Darie, C. C. A Critical Review of Bottom-Up Proteomics: The Good, the Bad, and the Future of This Field. *Proteomes* **2020**, *8* (3), 14. https://doi.org/10.3390/proteomes8030014.
(203) Elias, J. E.; Gygi, S. P. Target-Decoy Search Strategy for Increased Confidence in Large-Scale Protein Identifications by Mass Spectrometry. *Nat. Methods* **2007**, *4* (3), 207–214. https://doi.org/10.1038/nmeth1019.
(204) Choi, H.; Nesvizhskii, A. I. Semisupervised Model-Based Validation of Peptide Identifications in Mass Spectrometry-Based Proteomics. *J. Proteome Res.* **2008**, *7* (1), 254–265. https://doi.org/10.1021/pr070542g.
(205) Chick, J. M.; Kolippakkam, D.; Nusinow, D. P.; Zhai, B.; Rad, R.; Huttlin, E. L.; Gygi, S. P. A Mass-Tolerant Database Search Identifies a Large Proportion of Unassigned Spectra in Shotgun Proteomics as Modified Peptides. *Nat. Biotechnol.* **2015**, *33* (7), 743–749. https://doi.org/10.1038/nbt.3267.
(206) Tharakan, R.; Edwards, N.; Graham, D. R. M. Data Maximization by Multipass Analysis of Protein Mass Spectra. *Proteomics* **2010**, *10* (6), 1160–1171. https://doi.org/10.1002/pmic.200900433.
(207) Bern, M.; Kil, Y. J.; Becker, C. Byonic: Advanced Peptide and Protein Identification Software. *Curr. Protoc. Bioinforma. Ed. Board Andreas Baxevanis Al* **2012**, *CHAPTER*, Unit13.20. https://doi.org/10.1002/0471250953.bi1320s40.
(208) Wang, X.; Li, Y.; Wu, Z.; Wang, H.; Tan, H.; Peng, J. JUMP: A Tag-Based Database Search Tool for Peptide Identification with High Sensitivity and Accuracy*. *Mol. Cell. Proteomics* **2014**, *13* (12), 3663–3673. https://doi.org/10.1074/mcp.O114.039586.
(209) Cifani, P.; Dhabaria, A.; Chen, Z.; Yoshimi, A.; Kawaler, E.; Abdel-Wahab, O.; Poirier, J. T.; Kentsis, A. ProteomeGenerator: A Framework for Comprehensive Proteomics Based on de Novo Transcriptome Assembly and High-Accuracy Peptide Mass Spectral Matching. *J. Proteome Res.* **2018**, *17* (11), 3681–3692. https://doi.org/10.1021/acs.jproteome.8b00295.
(210) Bugyi, F.; Szabó, D.; Szabó, G.; Révész, Á.; Pape, V. F. S.; Soltész-Katona, E.; Tóth, E.; Kovács, O.; Langó, T.; Vékey, K.; Drahos, L. Influence of Post-Translational Modifications on Protein Identification in Database Searches. *ACS Omega* **2021**, *6* (11), 7469–7477. https://doi.org/10.1021/acsomega.0c05997.
(211) Ferries, S.; Perkins, S.; Brownridge, P. J.; Campbell, A.; Eyers, P. A.; Jones, A. R.; Eyers, C. E. Evaluation of Parameters for Confident Phosphorylation Site Localization Using an Orbitrap Fusion Tribrid Mass Spectrometer. *J. Proteome Res.* **2017**, *16* (9), 3448–3459. https://doi.org/10.1021/acs.jproteome.7b00337.
(212) Tsur, D.; Tanner, S.; Zandi, E.; Bafna, V.; Pevzner, P. A. Identification of Post-Translational Modifications via Blind Search of Mass-Spectra. In *2005 IEEE Computational Systems Bioinformatics Conference (CSB'05)*; 2005; pp 157–166. https://doi.org/10.1109/CSB.2005.34.
(213) Kong, A. T.; Leprevost, F. V.; Avtonomov, D. M.; Mellacheruvu, D.; Nesvizhskii, A. I. MSFragger: Ultrafast and Comprehensive Peptide Identification in Mass Spectrometry–Based Proteomics. *Nat. Methods* **2017**, *14* (5), 513–520. https://doi.org/10.1038/nmeth.4256.
(214) Yu, F.; Teo, G. C.; Kong, A. T.; Haynes, S. E.; Avtonomov, D. M.; Geiszler, D. J.; Nesvizhskii, A. I. Identification of Modified Peptides Using Localization-Aware Open Search. *Nat. Commun.* **2020**, *11* (1), 4065. https://doi.org/10.1038/s41467-020-17921-y.

(215) Lee, D. C. H.; Jones, A. R.; Hubbard, S. J. Computational Phosphoproteomics: From Identification to Localization. *Proteomics* **2015**, *15* (5–6), 950–963. https://doi.org/10.1002/pmic.201400372.
(216) Potel, C. M.; Lemeer, S.; Heck, A. J. R. Phosphopeptide Fragmentation and Site Localization by Mass Spectrometry: An Update. *Anal. Chem.* **2019**, *91* (1), 126–141. https://doi.org/10.1021/acs.analchem.8b04746.
(217) Leijten, N. M.; Heck, A. J. R.; Lemeer, S. Histidine Phosphorylation in Human Cells; a Needle or Phantom in the Haystack? *Nat. Methods* **2022**, *19* (7), 827–828. https://doi.org/10.1038/s41592-022-01524-0.
(218) Beausoleil, S. A.; Villén, J.; Gerber, S. A.; Rush, J.; Gygi, S. P. A Probability-Based Approach for High-Throughput Protein Phosphorylation Analysis and Site Localization. *Nat. Biotechnol.* **2006**, *24* (10), 1285–1292. https://doi.org/10.1038/nbt1240.
(219) Savitski, M. M.; Lemeer, S.; Boesche, M.; Lang, M.; Mathieson, T.; Bantscheff, M.; Kuster, B. Confident Phosphorylation Site Localization Using the Mascot Delta Score. *Mol. Cell. Proteomics MCP* **2011**, *10* (2), M110.003830. https://doi.org/10.1074/mcp.M110.003830.
(220) Taus, T.; Köcher, T.; Pichler, P.; Paschke, C.; Schmidt, A.; Henrich, C.; Mechtler, K. Universal and Confident Phosphorylation Site Localization Using PhosphoRS. *J. Proteome Res.* **2011**, *10* (12), 5354–5362. https://doi.org/10.1021/pr200611n.
(221) Locard-Paulet, M.; Bouyssié, D.; Froment, C.; Burlet-Schiltz, O.; Jensen, L. J. Comparing 22 Popular Phosphoproteomics Pipelines for Peptide Identification and Site Localization. *J. Proteome Res.* **2020**, *19* (3), 1338–1345. https://doi.org/10.1021/acs.jproteome.9b00679.
(222) Baker, P. R.; Trinidad, J. C.; Chalkley, R. J. Modification Site Localization Scoring Integrated into a Search Engine. *Mol. Cell. Proteomics MCP* **2011**, *10* (7), M111.008078. https://doi.org/10.1074/mcp.M111.008078.
(223) Fermin, D.; Walmsley, S. J.; Gingras, A.-C.; Choi, H.; Nesvizhskii, A. I. LuciPHOr: Algorithm for Phosphorylation Site Localization with False Localization Rate Estimation Using Modified Target-Decoy Approach. *Mol. Cell. Proteomics MCP* **2013**, *12* (11), 3409–3419. https://doi.org/10.1074/mcp.M113.028928.
(224) Heuvel, R. H. van den; Heck, A. J. Native Protein Mass Spectrometry: From Intact Oligomers to Functional Machineries. *Curr. Opin. Chem. Biol.* **2004**, *8* (5), 519–526. https://doi.org/10.1016/j.cbpa.2004.08.006.
(225) Ruotolo, B. T.; Robinson, C. V. Aspects of Native Proteins Are Retained in Vacuum. *Curr. Opin. Chem. Biol.* **2006**, *10* (5), 402–408. https://doi.org/10.1016/j.cbpa.2006.08.020.
(226) Juraschek, R.; Dülcks, T.; Karas, M. Nanoelectrospray—More than Just a Minimized-Flow Electrospray Ionization Source. *J. Am. Soc. Mass Spectrom.* **1999**, *10* (4), 300–308. https://doi.org/10.1016/S1044-0305(98)00157-3.
(227) Susa, A. C.; Xia, Z.; Williams, E. R. Small Emitter Tips for Native Mass Spectrometry of Proteins and Protein Complexes from Nonvolatile Buffers That Mimic the Intracellular Environment. *Anal. Chem.* **2017**, *89* (5), 3116–3122. https://doi.org/10.1021/acs.analchem.6b04897.
(228) Stiving, A. Q.; VanAernum, Z. L.; Busch, F.; Harvey, S. R.; Sarni, S. H.; Wysocki, V. H. Surface-Induced Dissociation: An Effective Method for Characterization of Protein Quaternary Structure. *Anal. Chem.* **2019**, *91* (1), 190–209. https://doi.org/10.1021/acs.analchem.8b05071.

(229) Harvey, S. R.; O'Neale, C.; Schey, K. L.; Wysocki, V. H. Native Mass Spectrometry and Surface Induced Dissociation Provide Insight into the Post-Translational Modifications of Tetrameric AQP0 Isolated from Bovine Eye Lens. *Anal. Chem.* **2022**, *94* (3), 1515–1519. https://doi.org/10.1021/acs.analchem.1c04322.
(230) J. Thompson, N.; Rosati, S.; J. Rose, R.; R. Heck, A. J. The Impact of Mass Spectrometry on the Study of Intact Antibodies: From Post-Translational Modifications to Structural Analysis. *Chem. Commun.* **2013**, *49* (6), 538–548. https://doi.org/10.1039/C2CC36755F.
(231) Yan, Y.; Xing, T.; Wang, S.; Li, N. Versatile, Sensitive, and Robust Native LC–MS Platform for Intact Mass Analysis of Protein Drugs. *J. Am. Soc. Mass Spectrom.* **2020**, *31* (10), 2171–2179. https://doi.org/10.1021/jasms.0c00277.
(232) Füssl, F.; Strasser, L.; Carillo, S.; Bones, J. Native LC–MS for Capturing Quality Attributes of Biopharmaceuticals on the Intact Protein Level. *Curr. Opin. Biotechnol.* **2021**, *71*, 32–40. https://doi.org/10.1016/j.copbio.2021.05.008.
(233) Lanucara, F.; Eyers, C. E. Top-down Mass Spectrometry for the Analysis of Combinatorial Post-Translational Modifications. *Mass Spectrom. Rev.* **2013**, *32* (1), 27–42. https://doi.org/10.1002/mas.21348.
(234) Pesavento, J. J.; Yang, H.; Kelleher, N. L.; Mizzen, C. A. Certain and Progressive Methylation of Histone H4 at Lysine 20 during the Cell Cycle. *Mol. Cell. Biol.* **2008**, *28* (1), 468–486. https://doi.org/10.1128/MCB.01517-07.
(235) Thomas, C. E.; Kelleher, N. L.; Mizzen, C. A. Mass Spectrometric Characterization of Human Histone H3: A Bird's Eye View. *J. Proteome Res.* **2006**, *5* (2), 240–247. https://doi.org/10.1021/pr050266a.
(236) Liu, R.; Xia, S.; Li, H. Native Top-down Mass Spectrometry for Higher-Order Structural Characterization of Proteins and Complexes. *Mass Spectrom. Rev. n/a* (n/a), e21793. https://doi.org/10.1002/mas.21793.
(237) Tang, Y.; Zhao, W.; Chen, Y.; Zhao, Y.; Gu, W. Acetylation Is Indispensable for P53 Activation. *Cell* **2008**, *133* (4), 612–626. https://doi.org/10.1016/j.cell.2008.03.025.
(238) Feeley, K. P.; Adams, C. M.; Mitra, R.; Eischen, C. M. Mdm2 Is Required for Survival and Growth of P53-Deficient Cancer Cells. *Cancer Res.* **2017**, *77* (14), 3823–3833. https://doi.org/10.1158/0008-5472.CAN-17-0809.
(239) Liang, L.; Wang, H.; Shi, H.; Li, Z.; Yao, H.; Bu, Z.; Song, N.; Li, C.; Xiang, D.; Zhang, Y.; Wang, J.; Hu, Y.; Xu, Q.; Ma, Y.; Cheng, Z.; Wang, Y.; Zhao, S.; Qian, J.; Chen, Y.; Fang, J.-Y.; Xu, J. A Designed Peptide Targets Two Types of Modifications of P53 with Anti-Cancer Activity. *Cell Chem. Biol.* **2018**, *25* (6), 761-774.e5. https://doi.org/10.1016/j.chembiol.2018.03.010.
(240) Ji, J.; Yu, Y.; Li, Z.-L.; Chen, M.-Y.; Deng, R.; Huang, X.; Wang, G.-F.; Zhang, M.-X.; Yang, Q.; Ravichandran, S.; Feng, G.-K.; Xu, X.-L.; Yang, C.-L.; Qiu, M.-Z.; Jiao, L.; Yang, D.; Zhu, X.-F. XIAP Limits Autophagic Degradation of Sox2 and Is A Therapeutic Target in Nasopharyngeal Carcinoma Stem Cells. *Theranostics* **2018**, *8* (6), 1494–1510. https://doi.org/10.7150/thno.21717.
(241) Armelin, H. A. Pituitary Extracts and Steroid Hormones in the Control of 3T3 Cell Growth. *Proc. Natl. Acad. Sci. U. S. A.* **1973**, *70* (9), 2702–2706. https://doi.org/10.1073/pnas.70.9.2702.
(242) Burgess, W. H.; Maciag, T. The Heparin-Binding (Fibroblast) Growth Factor Family of Proteins. *Annu. Rev. Biochem.* **1989**, *58*, 575–606. https://doi.org/10.1146/annurev.bi.58.070189.003043.

(243) Wiedłocha, A.; Sørensen, V. Signaling, Internalization, and Intracellular Activity of Fibroblast Growth Factor. *Curr. Top. Microbiol. Immunol.* **2004**, *286*, 45–79. https://doi.org/10.1007/978-3-540-69494-6_3.
(244) Sheng, Z.; Lewis, J. A.; Chirico, W. J. Nuclear and Nucleolar Localization of 18-KDa Fibroblast Growth Factor-2 Is Controlled by C-Terminal Signals. *J. Biol. Chem.* **2004**, *279* (38), 40153–40160. https://doi.org/10.1074/jbc.M400123200.
(245) Ornitz, D. M.; Itoh, N. Fibroblast Growth Factors. *Genome Biol.* **2001**, *2* (3), reviews3005.1. https://doi.org/10.1186/gb-2001-2-3-reviews3005.
(246) Bafico, A.; Aaronson, S. A. Growth Factor Receptors with Tyrosine Kinase Activity. *Holl.-Frei Cancer Med. 6th Ed.* **2003**.
(247) Mohammadi, M.; Olsen, S. K.; Ibrahimi, O. A. Structural Basis for Fibroblast Growth Factor Receptor Activation. *Cytokine Growth Factor Rev.* **2005**, *16* (2), 107–137. https://doi.org/10.1016/j.cytogfr.2005.01.008.
(248) Ornitz, D. M.; Itoh, N. The Fibroblast Growth Factor Signaling Pathway. *Wiley Interdiscip. Rev. Dev. Biol.* **2015**, *4* (3), 215–266. https://doi.org/10.1002/wdev.176.
(249) Sarabipour, S.; Hristova, K. Mechanism of FGF Receptor Dimerization and Activation. *Nat. Commun.* **2016**, *7* (1), 10262. https://doi.org/10.1038/ncomms10262.
(250) Sahni, M.; Ambrosetti, D.-C.; Mansukhani, A.; Gertner, R.; Levy, D.; Basilico, C. FGF Signaling Inhibits Chondrocyte Proliferation and Regulates Bone Development through the STAT-1 Pathway. *Genes Dev.* **1999**, *13* (11), 1361–1366.
(251) Brewer, J. R.; Mazot, P.; Soriano, P. Genetic Insights into the Mechanisms of Fgf Signaling. *Genes Dev.* **2016**, *30* (7), 751–771. https://doi.org/10.1101/gad.277137.115.
(252) Eswarakumar, V. P.; Lax, I.; Schlessinger, J. Cellular Signaling by Fibroblast Growth Factor Receptors. *Cytokine Growth Factor Rev.* **2005**, *16* (2), 139–149. https://doi.org/10.1016/j.cytogfr.2005.01.001.
(253) Gotoh, N. Regulation of Growth Factor Signaling by FRS2 Family Docking/Scaffold Adaptor Proteins. *Cancer Sci.* **2008**, *99* (7), 1319–1325. https://doi.org/10.1111/j.1349-7006.2008.00840.x.
(254) Raju, R.; Palapetta, S. M.; Sandhya, V. K.; Sahu, A.; Alipoor, A.; Balakrishnan, L.; Advani, J.; George, B.; Kini, K. R.; Geetha, N. P.; Prakash, H. S.; Prasad, T. S. K.; Chang, Y.-J.; Chen, L.; Pandey, A.; Gowda, H. A Network Map of FGF-1/FGFR Signaling System. *J. Signal Transduct.* **2014**, *2014*, e962962. https://doi.org/10.1155/2014/962962.
(255) Greulich, H.; Pollock, P. M. Targeting Mutant Fibroblast Growth Factor Receptors in Cancer. *Trends Mol. Med.* **2011**, *17* (5), 283–292. https://doi.org/10.1016/j.molmed.2011.01.012.
(256) Turner, N.; Grose, R. Fibroblast Growth Factor Signalling: From Development to Cancer. *Nat. Rev. Cancer* **2010**, *10* (2), 116–129. https://doi.org/10.1038/nrc2780.
(257) Clayton, N. S.; Wilson, A. S.; Laurent, E. P.; Grose, R. P.; Carter, E. P. Fibroblast Growth Factor-Mediated Crosstalk in Cancer Etiology and Treatment. *Dev. Dyn. Off. Publ. Am. Assoc. Anat.* **2017**, *246* (7), 493–501. https://doi.org/10.1002/dvdy.24514.
(258) Lasorella, A.; Sanson, M.; Iavarone, A. FGFR-TACC Gene Fusions in Human Glioma. *Neuro-Oncol.* **2017**, *19* (4), 475–483. https://doi.org/10.1093/neuonc/now240.
(259) Parish, A.; Schwaederle, M.; Daniels, G.; Piccioni, D.; Fanta, P.; Schwab, R.; Shimabukuro, K.; Parker, B. A.; Helsten, T.; Kurzrock, R. Fibroblast Growth Factor Family Aberrations in Cancers: Clinical and Molecular Characteristics. *Cell Cycle*

Georget. Tex **2015**, *14* (13), 2121–2128. https://doi.org/10.1080/15384101.2015.1041691.

(260) Helsten, T.; Elkin, S.; Arthur, E.; Tomson, B. N.; Carter, J.; Kurzrock, R. The FGFR Landscape in Cancer: Analysis of 4,853 Tumors by Next-Generation Sequencing. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **2016**, *22* (1), 259–267. https://doi.org/10.1158/1078-0432.CCR-14-3212.

(261) Dienstmann, R.; Rodon, J.; Prat, A.; Perez-Garcia, J.; Adamo, B.; Felip, E.; Cortes, J.; Iafrate, A. J.; Nuciforo, P.; Tabernero, J. Genomic Aberrations in the FGFR Pathway: Opportunities for Targeted Therapies in Solid Tumors. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **2014**, *25* (3), 552–563. https://doi.org/10.1093/annonc/mdt419.

(262) Presta, M.; Chiodelli, P.; Giacomini, A.; Rusnati, M.; Ronca, R. Fibroblast Growth Factors (FGFs) in Cancer: FGF Traps as a New Therapeutic Approach. *Pharmacol. Ther.* **2017**, *179*, 171–187. https://doi.org/10.1016/j.pharmthera.2017.05.013.

(263) André, F.; Bachelot, T.; Campone, M.; Dalenc, F.; Perez-Garcia, J. M.; Hurvitz, S. A.; Turner, N.; Rugo, H.; Smith, J. W.; Deudon, S.; Shi, M.; Zhang, Y.; Kay, A.; Porta, D. G.; Yovine, A.; Baselga, J. Targeting FGFR with Dovitinib (TKI258): Preclinical and Clinical Data in Breast Cancer. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **2013**, *19* (13), 3693–3702. https://doi.org/10.1158/1078-0432.CCR-13-0190.

(264) Gavine, P. R.; Mooney, L.; Kilgour, E.; Thomas, A. P.; Al-Kadhimi, K.; Beck, S.; Rooney, C.; Coleman, T.; Baker, D.; Mellor, M. J.; Brooks, A. N.; Klinowska, T. AZD4547: An Orally Bioavailable, Potent, and Selective Inhibitor of the Fibroblast Growth Factor Receptor Tyrosine Kinase Family. *Cancer Res.* **2012**, *72* (8), 2045–2056. https://doi.org/10.1158/0008-5472.CAN-11-3034.

(265) Pearson, A.; Smyth, E.; Babina, I. S.; Herrera-Abreu, M. T.; Tarazona, N.; Peckitt, C.; Kilgour, E.; Smith, N. R.; Geh, C.; Rooney, C.; Cutts, R.; Campbell, J.; Ning, J.; Fenwick, K.; Swain, A.; Brown, G.; Chua, S.; Thomas, A.; Johnston, S. R. D.; Ajaz, M.; Sumpter, K.; Gillbanks, A.; Watkins, D.; Chau, I.; Popat, S.; Cunningham, D.; Turner, N. C. High-Level Clonal FGFR Amplification and Response to FGFR Inhibition in a Translational Clinical Trial. *Cancer Discov.* **2016**, *6* (8), 838–851. https://doi.org/10.1158/2159-8290.CD-15-1246.

(266) Patani, H.; Bunney, T. D.; Thiyagarajan, N.; Norman, R. A.; Ogg, D.; Breed, J.; Ashford, P.; Potterton, A.; Edwards, M.; Williams, S. V.; Thomson, G. S.; Pang, C. S. M.; Knowles, M. A.; Breeze, A. L.; Orengo, C.; Phillips, C.; Katan, M. Landscape of Activating Cancer Mutations in FGFR Kinases and Their Differential Responses to Inhibitors in Clinical Use. *Oncotarget* **2016**, *7* (17), 24252–24268. https://doi.org/10.18632/oncotarget.8132.

(267) Chell, V.; Balmanno, K.; Little, A. S.; Wilson, M.; Andrews, S.; Blockley, L.; Hampson, M.; Gavine, P. R.; Cook, S. J. Tumour Cell Responses to New Fibroblast Growth Factor Receptor Tyrosine Kinase Inhibitors and Identification of a Gatekeeper Mutation in FGFR3 as a Mechanism of Acquired Resistance. *Oncogene* **2013**, *32* (25), 3059–3070. https://doi.org/10.1038/onc.2012.319.

(268) Wang, J.; Mikse, O.; Liao, R. G.; Li, Y.; Tan, L.; Janne, P. A.; Gray, N. S.; Wong, K. -k; Hammerman, P. S. Ligand-Associated ERBB2/3 Activation Confers Acquired Resistance to FGFR Inhibition in FGFR3-Dependent Cancer Cells. *Oncogene* **2015**, *34* (17), 2167–2177. https://doi.org/10.1038/onc.2014.161.

(269) Heiser, L. M.; Sadanandam, A.; Kuo, W.-L.; Benz, S. C.; Goldstein, T. C.; Ng, S.; Gibb, W. J.; Wang, N. J.; Ziyad, S.; Tong, F.; Bayani, N.; Hu, Z.; Billig, J. I.; Dueregger, A.; Lewis, S.; Jakkula, L.; Korkola, J. E.; Durinck, S.; Pepin, F.; Guan, Y.; Purdom, E.; Neuvial, P.; Bengtsson, H.; Wood, K. W.; Smith, P. G.; Vassilev, L. T.; Hennessy, B. T.; Greshock, J.; Bachman, K. E.; Hardwicke, M. A.; Park, J. W.; Marton, L. J.; Wolf, D. M.; Collisson, E. A.; Neve, R. M.; Mills, G. B.; Speed, T. P.; Feiler, H. S.; Wooster, R. F.; Haussler, D.; Stuart, J. M.; Gray, J. W.; Spellman, P. T. Subtype and Pathway Specific Responses to Anticancer Compounds in Breast Cancer. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (8), 2724–2729. https://doi.org/10.1073/pnas.1018854108.

(270) Grossman, A. R.; Schaefer, M. R.; Chiang, G. G.; Collier, J. L. The Phycobilisome, a Light-Harvesting Complex Responsive to Environmental Conditions. *Microbiol. Rev.* **1993**, *57* (3), 725–749. https://doi.org/10.1128/mr.57.3.725-749.1993.

(271) Klotz, A. V.; Leary, J. A.; Glazer, A. N. Post-Translational Methylation of Asparaginyl Residues. Identification of Beta-71 Gamma-N-Methylasparagine in Allophycocyanin. *J. Biol. Chem.* **1986**, *261* (34), 15891–15894. https://doi.org/10.1016/S0021-9258(18)66647-5.

(272) Soo, R. M.; Hemp, J.; Parks, D. H.; Fischer, W. W.; Hugenholtz, P. On the Origins of Oxygenic Photosynthesis and Aerobic Respiration in Cyanobacteria. *Science* **2017**, *355* (6332), 1436–1440. https://doi.org/10.1126/science.aal3794.

(273) Blaustein, R. The Great Oxidation Event: Evolving Understandings of How Oxygenic Life on Earth Began. *BioScience* **2016**, *66* (3), 189–195. https://doi.org/10.1093/biosci/biv193.

(274) Kumar, V.; Maurya, P. K.; Mondal, S.; Sinha, R. P.; Singh, S. P. Chapter 6 - Photomorphogenesis in the Cyanobacterium Fremyella Diplosiphon Improves Photosynthetic Efficiency. In *Cyanobacteria*; Mishra, A. K., Tiwari, D. N., Rai, A. N., Eds.; Academic Press, 2019; pp 131–143. https://doi.org/10.1016/B978-0-12-814667-5.00006-4.

(275) Lane, N. *Transformer: The Deep Chemistry of Life and Death*, Main.; Profile Books Ltd.

(276) Turrens, J. F. Mitochondrial Formation of Reactive Oxygen Species. *J. Physiol.* **2003**, *552* (Pt 2), 335–344. https://doi.org/10.1113/jphysiol.2003.049478.

(277) Finkel, T. Signal Transduction by Reactive Oxygen Species. *J. Cell Biol.* **2011**, *194* (1), 7–15. https://doi.org/10.1083/jcb.201102095.

(278) Imlay, J. A. Pathways of Oxidative Damage. *Annu. Rev. Microbiol.* **2003**, *57*, 395–418. https://doi.org/10.1146/annurev.micro.57.030502.090938.

(279) Harman, D. Free Radical Theory of Aging. *Mutat. Res.* **1992**, *275* (3), 257–266. https://doi.org/10.1016/0921-8734(92)90030-S.

(280) Latifi, A.; Ruiz, M.; Zhang, C.-C. Oxidative Stress in Cyanobacteria. *FEMS Microbiol. Rev.* **2009**, *33* (2), 258–278. https://doi.org/10.1111/j.1574-6976.2008.00134.x.

(281) Liu, X.-G.; Zhao, J.-J.; Wu, Q.-Y. Oxidative Stress and Metal Ions Effects on the Cores of Phycobilisomes in Synechocystis Sp. PCC 6803. *FEBS Lett.* **2005**, *579* (21), 4571–4576. https://doi.org/10.1016/j.febslet.2005.07.020.

(282) Öquist, G. Light-Induced Changes in Pigment Composition of Photosynthetic Lamellae and Cell-Free Extracts from the Blue-Green Alga Anacystis Nidulans. *Physiol. Plant.* **1974**, *30* (1), 45–48. https://doi.org/10.1111/j.1399-3054.1974.tb04989.x.

(283) Lönneborg, A.; Lind, L. K.; Kalla, S. R.; Gustafsson, P.; Oquist, G. Acclimation Processes in the Light-Harvesting System of the Cyanobacterium Anacystis Nidulans Following a Light Shift from White to Red Light. *Plant Physiol.* **1985**, *78* (1), 110–114. https://doi.org/10.1104/pp.78.1.110.
(284) Mailloux, R. J.; McBride, S. L.; Harper, M.-E. Unearthing the Secrets of Mitochondrial ROS and Glutathione in Bioenergetics. *Trends Biochem. Sci.* **2013**, *38* (12), 592–602. https://doi.org/10.1016/j.tibs.2013.09.001.
(285) Dalle-Donne, I.; Rossi, R.; Colombo, G.; Giustarini, D.; Milzani, A. Protein S-Glutathionylation: A Regulatory Device from Bacteria to Humans. *Trends Biochem. Sci.* **2009**, *34* (2), 85–96. https://doi.org/10.1016/j.tibs.2008.11.002.
(286) Checconi, P.; Limongi, D.; Baldelli, S.; Ciriolo, M. R.; Nencioni, L.; Palamara, A. T. Role of Glutathionylation in Infection and Inflammation. *Nutrients* **2019**, *11* (8), 1952. https://doi.org/10.3390/nu11081952.
(287) Womersley, J. S.; Uys, J. D. Chapter Three - S-Glutathionylation and Redox Protein Signaling in Drug Addiction. In *Progress in Molecular Biology and Translational Science*; Rahman, S., Ed.; The Molecular Basis of Drug Addiction; Academic Press, 2016; Vol. 137, pp 87–121. https://doi.org/10.1016/bs.pmbts.2015.10.001.
(288) Klatt, P.; Molina, E. P.; Lamas, S. Nitric Oxide Inhibits C-Jun DNA Binding by Specifically Targeted S-Glutathionylation. *J. Biol. Chem.* **1999**, *274* (22), 15857–15864. https://doi.org/10.1074/jbc.274.22.15857.
(289) Klatt, P.; Molina, E. P.; De Lacoba, M. G.; Padilla, C. A.; Martinez-Galesteo, E.; Barcena, J. A.; Lamas, S. Redox Regulation of C-Jun DNA Binding by Reversible S-Glutathiolation. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* **1999**, *13* (12), 1481–1490. https://doi.org/10.1096/fasebj.13.12.1481.
(290) Klatt, P.; Pineda Molina, E.; Pérez-Sala, D.; Lamas, S. Novel Application of S-Nitrosoglutathione-Sepharose to Identify Proteins That Are Potential Targets for S-Nitrosoglutathione-Induced Mixed-Disulphide Formation. *Biochem. J.* **2000**, *349* (Pt 2), 567–578.
(291) Lind, C.; Gerdes, R.; Hamnell, Y.; Schuppe-Koistinen, I.; von Löwenhielm, H. B.; Holmgren, A.; Cotgreave, I. A. Identification of S-Glutathionylated Cellular Proteins during Oxidative Stress and Constitutive Metabolism by Affinity Purification and Proteomic Analysis. *Arch. Biochem. Biophys.* **2002**, *406* (2), 229–240. https://doi.org/10.1016/s0003-9861(02)00468-x.
(292) Ward, N. E.; Stewart, J. R.; Ioannides, C. G.; O'Brian, C. A. Oxidant-Induced S-Glutathiolation Inactivates Protein Kinase C-Alpha (PKC-Alpha): A Potential Mechanism of PKC Isozyme Regulation. *Biochemistry* **2000**, *39* (33), 10319–10329. https://doi.org/10.1021/bi000781g.
(293) Manevich, Y.; Feinstein, S. I.; Fisher, A. B. Activation of the Antioxidant Enzyme 1-CYS Peroxiredoxin Requires Glutathionylation Mediated by Heterodimerization with Pi GST. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (11), 3780–3785. https://doi.org/10.1073/pnas.0400181101.
(294) Rokutan, K.; Thomas, J. A.; Johnston, R. B. Phagocytosis and Stimulation of the Respiratory Burst by Phorbol Diester Initiate S-Thiolation of Specific Proteins in Macrophages. *J. Immunol.* **1991**, *147* (1), 260–264.
(295) Rokutan, K.; Johnston, R. B.; Kawai, K. Oxidative Stress Induces S-Thiolation of Specific Proteins in Cultured Gastric Mucosal Cells. *Am. J. Physiol.* **1994**, *266* (2 Pt 1), G247-254. https://doi.org/10.1152/ajpgi.1994.266.2.G247.

(296) Jahngen-Hodge, J.; Obin, M. S.; Gong, X.; Shang, F.; Nowell, T. R.; Gong, J.; Abasi, H.; Blumberg, J.; Taylor, A. Regulation of Ubiquitin-Conjugating Enzymes by Glutathione Following Oxidative Stress. *J. Biol. Chem.* **1997**, *272* (45), 28218–28226. https://doi.org/10.1074/jbc.272.45.28218.

(297) Cao, X.; Kambe, F.; Lu, X.; Kobayashi, N.; Ohmori, S.; Seo, H. Glutathionylation of Two Cysteine Residues in Paired Domain Regulates DNABinding Activity OfPax-8 *. *J. Biol. Chem.* **2005**, *280* (27), 25901–25906. https://doi.org/10.1074/jbc.M411443200.

(298) Ghezzi, P.; Casagrande, S.; Massignan, T.; Basso, M.; Bellacchio, E.; Mollica, L.; Biasini, E.; Tonelli, R.; Eberini, I.; Gianazza, E.; Dai, W. W.; Fratelli, M.; Salmona, M.; Sherry, B.; Bonetto, V. Redox Regulation of Cyclophilin A by Glutathionylation. *PROTEOMICS* **2006**, *6* (3), 817–825. https://doi.org/10.1002/pmic.200500177.

(299) Casagrande, S.; Bonetto, V.; Fratelli, M.; Gianazza, E.; Eberini, I.; Massignan, T.; Salmona, M.; Chang, G.; Holmgren, A.; Ghezzi, P. Glutathionylation of Human Thioredoxin: A Possible Crosstalk between the Glutathione and Thioredoxin Systems. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (15), 9745–9749. https://doi.org/10.1073/pnas.152168599.

(300) Gao, X.-H.; Bedhomme, M.; Veyel, D.; Zaffagnini, M.; Lemaire, S. D. Methods for Analysis of Protein Glutathionylation and Their Application to Photosynthetic Organisms. *Mol. Plant* **2009**, *2* (2), 218–235. https://doi.org/10.1093/mp/ssn072.

(301) Chardonnet, S.; Sakr, S.; Cassier-Chauvat, C.; Le Maréchal, P.; Chauvat, F.; Lemaire, S. D.; Decottignies, P. First Proteomic Study of S-Glutathionylation in Cyanobacteria. *J. Proteome Res.* **2015**, *14* (1), 59–71. https://doi.org/10.1021/pr500625a.

(302) Tamara, S.; Hoek, M.; Scheltema, R. A.; Leney, A. C.; Heck, A. J. R. A Colorful Pallet of B-Phycoerythrin Proteoforms Exposed by a Multimodal Mass Spectrometry Approach. *Chem* **2019**, *5* (5), 1302–1317. https://doi.org/10.1016/j.chempr.2019.03.006.

(303) Shteynberg, D. D.; Deutsch, E. W.; Campbell, D. S.; Hoopmann, M. R.; Kusebauch, U.; Lee, D.; Mendoza, L.; Midha, M. K.; Sun, Z.; Whetton, A. D.; Moritz, R. L. PTMProphet: Fast and Accurate Mass Modification Localization for the Trans-Proteomic Pipeline. *J. Proteome Res.* **2019**, *18* (12), 4262–4272. https://doi.org/10.1021/acs.jproteome.9b00205.

(304) Raudvere, U.; Kolberg, L.; Kuzmin, I.; Arak, T.; Adler, P.; Peterson, H.; Vilo, J. G:Profiler: A Web Server for Functional Enrichment Analysis and Conversions of Gene Lists (2019 Update). *Nucleic Acids Res.* **2019**, *47* (W1), W191–W198. https://doi.org/10.1093/nar/gkz369.

(305) Forozan, F.; Veldman, R.; Ammerman, C. A.; Parsa, N. Z.; Kallioniemi, A.; Kallioniemi, O.-P.; Ethier, S. P. Molecular Cytogenetic Analysis of 11 New Breast Cancer Cell Lines. *Br. J. Cancer* **1999**, *81* (8), 1328–1334. https://doi.org/10.1038/sj.bjc.6695007.

(306) Bradford, M. M. A Rapid and Sensitive Method for the Quantitation of Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding. *Anal. Biochem.* **1976**, *72*, 248–254. https://doi.org/10.1006/abio.1976.9999.

(307) Baker, Katie. The Role of FGF Signalling to Mcl-1 via GSK-3 in Triple Negative Breast Cancer, University of Birmingham, School of Biosciences, 2016.

(308) Jensen, L. J.; Kuhn, M.; Stark, M.; Chaffron, S.; Creevey, C.; Muller, J.; Doerks, T.; Julien, P.; Roth, A.; Simonovic, M.; Bork, P.; von Mering, C. STRING 8--a Global

View on Proteins and Their Functional Interactions in 630 Organisms. *Nucleic Acids Res.* **2009**, *37* (Database issue), D412-416. https://doi.org/10.1093/nar/gkn760.
(309) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **2003**, *13* (11), 2498–2504. https://doi.org/10.1101/gr.1239303.
(310) CCAP. *Knowledgebase – CCAP*. CCAP Knowledgbase. https://www.ccap.ac.uk/index.php/knowledgebase/ (accessed 2022-09-30).
(311) Bellamy-Carter, J.; Sound, J. K.; Leney, A. C. Probing Heavy Metal Binding to Phycobiliproteins. *FEBS J.* **2022**, *289* (15), 4646–4656. https://doi.org/10.1111/febs.16396.
(312) Bennett, A.; Bogorad, L. Complementary Chromatic Adaptation in a Filamentous Blue-Green Alga. *J. Cell Biol.* **1973**, *58* (2), 419–435. https://doi.org/10.1083/jcb.58.2.419.
(313) Sound, J. K.; Peters, A.; Bellamy-Carter, J.; Rad-Menéndez, C.; MacKechnie, K.; Green, D. H.; Leney, A. C. Rapid Cyanobacteria Species Identification with High Sensitivity Using Native Mass Spectrometry. *Anal. Chem.* **2021**, *93* (42), 14293–14299. https://doi.org/10.1021/acs.analchem.1c03412.
(314) Marty, M. T.; Baldwin, A. J.; Marklund, E. G.; Hochberg, G. K. A.; Benesch, J. L. P.; Robinson, C. V. Bayesian Deconvolution of Mass and Ion Mobility Spectra: From Binary Interactions to Polydisperse Ensembles. *Anal. Chem.* **2015**, *87* (8), 4370–4376. https://doi.org/10.1021/acs.analchem.5b00140.
(315) The UniProt Consortium. UniProt: The Universal Protein Knowledgebase. *Nucleic Acids Res.* **2017**, *45* (D1), D158–D169. https://doi.org/10.1093/nar/gkw1099.
(316) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic Local Alignment Search Tool. *J. Mol. Biol.* **1990**, *215* (3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.
(317) Tabb, D. L. The SEQUEST Family Tree. *J. Am. Soc. Mass Spectrom.* **2015**, *26* (11), 1814–1819. https://doi.org/10.1007/s13361-015-1201-3.
(318) Horita, H.; Law, A.; Middleton, K. Utilizing Optimized Tools to Investigate PTM Crosstalk: Identifying Potential PTM Crosstalk of Acetylated Mitochondrial Proteins. *Proteomes* **2018**, *6* (2). https://doi.org/10.3390/proteomes6020024.
(319) Muehlbauer, L. K.; Hebert, A. S.; Westphall, M. S.; Shishkova, E.; Coon, J. J. Global Phosphoproteome Analysis Using High-Field Asymmetric Waveform Ion Mobility Spectrometry on a Hybrid Orbitrap Mass Spectrometer. *Anal. Chem.* **2020**, *92* (24), 15959–15967. https://doi.org/10.1021/acs.analchem.0c03415.
(320) Zhao, H.; Cunningham, D. L.; Creese, A. J.; Heath, J. K.; Cooper, H. J. FAIMS and Phosphoproteomics of Fibroblast Growth Factor Signaling: Enhanced Identification of Multiply Phosphorylated Peptides. *J. Proteome Res.* **2015**, *14* (12), 5077–5087. https://doi.org/10.1021/acs.jproteome.5b00713.
(321) Shvartsburg, A. A.; Creese, A. J.; Smith, R. D.; Cooper, H. J. Separation of Peptide Isomers with Variant Modified Sites by High-Resolution Differential Ion Mobility Spectrometry. *Anal. Chem.* **2010**, *82* (19), 8327–8334. https://doi.org/10.1021/ac101878a.
(322) Thermo Scientific™ Pierce™. *Thermo Scientific™ Pierce™ HeLa Protein Digest Standard*. Thermo Scientific™ Pierce™ HeLa Protein Digest Standard. https://www.fishersci.com/shop/products/pierce-hela-protein-digest-standar-2/p-7000090#:~:text=The%20Pierce%20HeLa%20Protein%20Digest%20Standard%20i

s%20a%20lyophilized%20tryptic,development%20and%20MS%20performance%20benchmarking.
(323) Sayers, E. W.; Barrett, T.; Benson, D. A.; Bolton, E.; Bryant, S. H.; Canese, K.; Chetvernin, V.; Church, D. M.; DiCuccio, M.; Federhen, S.; Feolo, M.; Geer, L. Y.; Helmberg, W.; Kapustin, Y.; Landsman, D.; Lipman, D. J.; Lu, Z.; Madden, T. L.; Madej, T.; Maglott, D. R.; Marchler-Bauer, A.; Miller, V.; Mizrachi, I.; Ostell, J.; Panchenko, A.; Pruitt, K. D.; Schuler, G. D.; Sequeira, E.; Sherry, S. T.; Shumway, M.; Sirotkin, K.; Slotta, D.; Souvorov, A.; Starchenko, G.; Tatusova, T. A.; Wagner, L.; Wang, Y.; John Wilbur, W.; Yaschenko, E.; Ye, J. Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2010**, *38* (suppl_1), D5–D16. https://doi.org/10.1093/nar/gkp967.
(324) Kaszycki, J. L.; Shvartsburg, A. A. A Priori Intrinsic PTM Size Parameters for Predicting the Ion Mobilities of Modified Peptides. *J. Am. Soc. Mass Spectrom.* **2017**, *28* (2), 294–302. https://doi.org/10.1007/s13361-016-1553-3.
(325) Wu, S. T.; Xia, Y.-Q.; Jemal, M. High-Field Asymmetric Waveform Ion Mobility Spectrometry Coupled with Liquid Chromatography/Electrospray Ionization Tandem Mass Spectrometry (LC/ESI-FAIMS-MS/MS) Multi-Component Bioanalytical Method Development, Performance Evaluation and Demonstration of the Constancy of the Compensation Voltage with Change of Mobile Phase Composition or Flow Rate. *Rapid Commun. Mass Spectrom.* **2007**, *21* (22), 3667–3676. https://doi.org/10.1002/rcm.3264.
(326) Ahrné, E.; Müller, M.; Lisacek, F. Unrestricted Identification of Modified Proteins Using MS/MS. *PROTEOMICS* **2010**, *10* (4), 671–686. https://doi.org/10.1002/pmic.200900502.
(327) Geiszler, D. J.; Kong, A. T.; Avtonomov, D. M.; Yu, F.; Leprevost, F. da V.; Nesvizhskii, A. I. PTM-Shepherd: Analysis and Summarization of Post-Translational and Chemical Modifications From Open Search Results. *Mol. Cell. Proteomics* **2021**, *20*, 100018. https://doi.org/10.1074/mcp.TIR120.002216.
(328) Giansanti, P.; Aye, T. T.; van den Toorn, H.; Peng, M.; van Breukelen, B.; Heck, A. J. R. An Augmented Multiple-Protease-Based Human Phosphopeptide Atlas. *Cell Rep.* **2015**, *11* (11), 1834–1843. https://doi.org/10.1016/j.celrep.2015.05.029.
(329) Martens, L.; Hermjakob, H.; Jones, P.; Adamski, M.; Taylor, C.; States, D.; Gevaert, K.; Vandekerckhove, J.; Apweiler, R. PRIDE: The Proteomics Identifications Database. *PROTEOMICS* **2005**, *5* (13), 3537–3545. https://doi.org/10.1002/pmic.200401303.
(330) Saiardi, A. Protein Pyrophosphorylation: Moving Forward. *Biochem. J.* **2016**, *473* (21), 3765–3768. https://doi.org/10.1042/BCJ20160710C.
(331) Hunter, T. The Age of Crosstalk: Phosphorylation, Ubiquitination, and Beyond. *Mol. Cell* **2007**, *28* (5), 730–738. https://doi.org/10.1016/j.molcel.2007.11.019.
(332) Zhang, Y.; Zeng, L. Crosstalk between Ubiquitination and Other Post-Translational Protein Modifications in Plant Immunity. *Plant Commun.* **2020**, *1* (4), 100041. https://doi.org/10.1016/j.xplc.2020.100041.
(333) Hains, P. G.; Truscott, R. J. W. Age-Dependent Deamidation of Lifelong Proteins in the Human Lens. *Invest. Ophthalmol. Vis. Sci.* **2010**, *51* (6), 3107–3114. https://doi.org/10.1167/iovs.09-4308.
(334) Hasan, Q.; Alluri, R. V.; Rao, P.; Ahuja, Y. R. Role of Glutamine Deamidation in Neurodegenerative Diseases Associated with Triplet Repeat Expansions: A Hypothesis. *J. Mol. Neurosci. MN* **2006**, *29* (1), 29–33.

(335) Moss, C. X.; Matthews, S. P.; Lamont, D. J.; Watts, C. Asparagine Deamidation Perturbs Antigen Presentation on Class II Major Histocompatibility Complex Molecules*. *J. Biol. Chem.* **2005**, *280* (18), 18498–18503. https://doi.org/10.1074/jbc.M501241200.
(336) Curnis, F.; Longhi, R.; Crippa, L.; Cattaneo, A.; Dondossola, E.; Bachi, A.; Corti, A. Spontaneous Formation of L-Isoaspartate and Gain of Function in Fibronectin. *J. Biol. Chem.* **2006**, *281* (47), 36466–36476. https://doi.org/10.1074/jbc.M604812200.
(337) York, W. S.; Mazumder, R.; Ranzinger, R.; Edwards, N.; Kahsay, R.; Aoki-Kinoshita, K. F.; Campbell, M. P.; Cummings, R. D.; Feizi, T.; Martin, M.; Natale, D. A.; Packer, N. H.; Woods, R. J.; Agarwal, G.; Arpinar, S.; Bhat, S.; Blake, J.; Castro, L. J. G.; Fochtman, B.; Gildersleeve, J.; Goldman, R.; Holmes, X.; Jain, V.; Kulkarni, S.; Mahadik, R.; Mehta, A.; Mousavi, R.; Nakarakommula, S.; Navelkar, R.; Pattabiraman, N.; Pierce, M. J.; Ross, K.; Vasudev, P.; Vora, J.; Williamson, T.; Zhang, W. GlyGen: Computational and Informatics Resources for Glycoscience. *Glycobiology* **2020**, *30* (2), 72–73. https://doi.org/10.1093/glycob/cwz080.
(338) Lee, T.-Y.; Chen, Y.-J.; Lu, C.-T.; Ching, W.-C.; Teng, Y.-C.; Huang, H.-D.; Chen, Y.-J. DbSNO: A Database of Cysteine S-Nitrosylation. *Bioinformatics* **2012**, *28* (17), 2293–2295. https://doi.org/10.1093/bioinformatics/bts436.
(339) Gräff, J.; Tsai, L.-H. Histone Acetylation: Molecular Mnemonics on the Chromatin. *Nat. Rev. Neurosci.* **2013**, *14* (2), 97–111. https://doi.org/10.1038/nrn3427.
(340) Lin, B.; Qing, X.; Liao, J.; Zhuo, K. Role of Protein Glycosylation in Host-Pathogen Interaction. *Cells* **2020**, *9* (4), 1022. https://doi.org/10.3390/cells9041022.
(341) Lee, D. Y.; Teyssier, C.; Strahl, B. D.; Stallcup, M. R. Role of Protein Methylation in Regulation of Transcription. *Endocr. Rev.* **2005**, *26* (2), 147–170. https://doi.org/10.1210/er.2004-0008.
(342) A, M.; Latario, C. J.; Pickrell, L. E.; Higgs, H. N. Lysine Acetylation of Cytoskeletal Proteins: Emergence of an Actin Code. *J. Cell Biol.* **2020**, *219* (12). https://doi.org/10.1083/jcb.202006151.
(343) Blackwell, E.; Ceman, S. Arginine Methylation of RNA-Binding Proteins Regulates Cell Function and Differentiation. *Mol. Reprod. Dev.* **2012**, *79* (3), 163–175. https://doi.org/10.1002/mrd.22024.
(344) Csizmok, V.; Forman-Kay, J. D. Complex Regulatory Mechanisms Mediated by the Interplay of Multiple Post-Translational Modifications. *Curr. Opin. Struct. Biol.* **2018**, *48*, 58–67. https://doi.org/10.1016/j.sbi.2017.10.013.
(345) Giansanti, P.; Tsiatsiani, L.; Low, T. Y.; Heck, A. J. R. Six Alternative Proteases for Mass Spectrometry–Based Proteomics beyond Trypsin. *Nat. Protoc.* **2016**, *11* (5), 993–1006. https://doi.org/10.1038/nprot.2016.057.
(346) Shliaha, P. V.; Gorshkov, V.; Kovalchuk, S. I.; Schwämmle, V.; Baird, M. A.; Shvartsburg, A. A.; Jensen, O. N. Middle-Down Proteomic Analyses with Ion Mobility Separations of Endogenous Isomeric Proteoforms. *Anal. Chem.* **2020**, *92* (3), 2364–2368. https://doi.org/10.1021/acs.analchem.9b05011.
(347) Yuan, Z.-F.; Sidoli, S.; Marchione, D. M.; Simithy, J.; Janssen, K. A.; Szurgot, M. R.; Garcia, B. A. EpiProfile 2.0: A Computational Platform for Processing Epi-Proteomics Mass Spectrometry Data. *J. Proteome Res.* **2018**, *17* (7), 2533–2541. https://doi.org/10.1021/acs.jproteome.8b00133.
(348) Ye, X.; Nalepa, G.; Welcker, M.; Kessler, B. M.; Spooner, E.; Qin, J.; Elledge, S. J.; Clurman, B. E.; Harper, J. W. Recognition of Phosphodegron Motifs in Human Cyclin

E by the SCFFbw7 Ubiquitin Ligase *. *J. Biol. Chem.* **2004**, *279* (48), 50110–50119. https://doi.org/10.1074/jbc.M409226200.
(349) Leutert, M.; Entwisle, S. W.; Villén, J. Decoding Post-Translational Modification Crosstalk With Proteomics. *Mol. Cell. Proteomics* **2021**, *20*. https://doi.org/10.1016/j.mcpro.2021.100129.
(350) The SUM Breast Cancer Cell Line Knowledge Base (SLKBase). *The SUM Breast Cancer Cell Line Knowledge Base (SLKBase)*. The SUM Breast Cancer Cell Line Knowledge Base (SLKBase). https://sumlineknowledgebase.com/?page_id=650 (accessed 2022-11-10).
(351) Trumpp-Kallmeyer, S.; Rubin, J. R.; Humblet, C.; Hamby, J. M.; Hollis Showalter, H. D. Development of a Binding Model to Protein Tyrosine Kinases for Substituted Pyrido[2,3-d]Pyrimidine Inhibitors. *J. Med. Chem.* **1998**, *41* (11), 1752–1763. https://doi.org/10.1021/jm970634p.
(352) Gupta, A. K.; Bhunia, S. S.; Balaramnavar, V. M.; Saxena, A. K. Pharmacophore Modelling, Molecular Docking and Virtual Screening for EGFR (HER 1) Tyrosine Kinase Inhibitors. *SAR QSAR Environ. Res.* **2011**, *22* (3–4), 239–263. https://doi.org/10.1080/1062936X.2010.548830.
(353) Hanahan, D.; Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* **2011**, *144* (5), 646–674. https://doi.org/10.1016/j.cell.2011.02.013.
(354) Mohammadi, M.; McMahon, G.; Sun, L.; Tang, C.; Hirth, P.; Yeh, B. K.; Hubbard, S. R.; Schlessinger, J. Structures of the Tyrosine Kinase Domain of Fibroblast Growth Factor Receptor in Complex with Inhibitors. *Science* **1997**, *276* (5314), 955–960. https://doi.org/10.1126/science.276.5314.955.
(355) Sun, L.; Tran, N.; Liang, C.; Tang, F.; Rice, A.; Schreck, R.; Waltz, K.; Shawver, L. K.; McMahon, G.; Tang, C. Design, Synthesis, and Evaluations of Substituted 3-[(3- or 4-Carboxyethylpyrrol-2-Yl)Methylidenyl]Indolin-2-Ones as Inhibitors of VEGF, FGF, and PDGF Receptor Tyrosine Kinases. *J. Med. Chem.* **1999**, *42* (25), 5120–5130. https://doi.org/10.1021/jm9904295.
(356) Tucker, J. A.; Klein, T.; Breed, J.; Breeze, A. L.; Overman, R.; Phillips, C.; Norman, R. A. Structural Insights into FGFR Kinase Isoform Selectivity: Diverse Binding Modes of AZD4547 and Ponatinib in Complex with FGFR1 and FGFR4. *Structure* **2014**, *22* (12), 1764–1774. https://doi.org/10.1016/j.str.2014.09.019.
(357) Chen, H.; Ma, J.; Li, W.; Eliseenkova, A. V.; Xu, C.; Neubert, T. A.; Miller, W. T.; Mohammadi, M. A Molecular Brake in the Kinase Hinge Region Regulates the Activity of Receptor Tyrosine Kinases. *Mol. Cell* **2007**, *27* (5), 717–730. https://doi.org/10.1016/j.molcel.2007.06.028.
(358) Chen, H.; Huang, Z.; Dutta, K.; Blais, S.; Neubert, T. A.; Li, X.; Cowburn, D.; Traaseth, N. J.; Mohammadi, M. Cracking the Molecular Origin of Intrinsic Tyrosine Kinase Activity through Analysis of Pathogenic Gain-of-Function Mutations. *Cell Rep.* **2013**, *4* (2), 376–384. https://doi.org/10.1016/j.celrep.2013.06.025.
(359) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera--a Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25* (13), 1605–1612. https://doi.org/10.1002/jcc.20084.
(360) Koziczak, M.; Holbro, T.; Hynes, N. E. Blocking of FGFR Signaling Inhibits Breast Cancer Cell Proliferation through Downregulation of D-Type Cyclins. *Oncogene* **2004**, *23* (20), 3501–3508. https://doi.org/10.1038/sj.onc.1207331.

(361) Goldfarb, M. Signaling By Fibroblast Growth Factors: The Inside Story. *Sci. STKE* **2001**, *2001* (106), pe37–pe37. https://doi.org/10.1126/stke.2001.106.pe37.
(362) Anderson, H. J.; Galileo, D. S. Small-Molecule Inhibitors of FGFR, Integrins and FAK Selectively Decrease L1CAM-Stimulated Glioblastoma Cell Motility and Proliferation. *Cell. Oncol.* **2016**, *39* (3), 229–242. https://doi.org/10.1007/s13402-016-0267-7.
(363) Nomura, S.; Yoshitomi, H.; Takano, S.; Shida, T.; Kobayashi, S.; Ohtsuka, M.; Kimura, F.; Shimizu, H.; Yoshidome, H.; Kato, A.; Miyazaki, M. FGF10/FGFR2 Signal Induces Cell Migration and Invasion in Pancreatic Cancer. *Br. J. Cancer* **2008**, *99* (2), 305–313. https://doi.org/10.1038/sj.bjc.6604473.
(364) Neben, C. L.; Tuzon, C. T.; Mao, X.; Lay, F. D.; Merrill, A. E. FGFR2 Mutations in Bent Bone Dysplasia Syndrome Activate Nucleolar Stress and Perturb Cell Fate Determination. *Hum. Mol. Genet.* **2017**, *26* (17), 3253–3270. https://doi.org/10.1093/hmg/ddx209.
(365) Neben, C. L.; Idoni, B.; Salva, J. E.; Tuzon, C. T.; Rice, J. C.; Krakow, D.; Merrill, A. E. Bent Bone Dysplasia Syndrome Reveals Nucleolar Activity for FGFR2 in Ribosomal DNA Transcription. *Hum. Mol. Genet.* **2014**, *23* (21), 5659–5671. https://doi.org/10.1093/hmg/ddu282.
(366) Ye, T.; Wei, X.; Yin, T.; Xia, Y.; Li, D.; Shao, B.; Song, X.; He, S.; Luo, M.; Gao, X.; He, Z.; Luo, C.; Xiong, Y.; Wang, N.; Zeng, J.; Zhao, L.; Shen, G.; Xie, Y.; Yu, L.; Wei, Y. Inhibition of FGFR Signaling by PD173074 Improves Antitumor Immunity and Impairs Breast Cancer Metastasis. *Breast Cancer Res. Treat.* **2014**, *143* (3), 435–446. https://doi.org/10.1007/s10549-013-2829-y.
(367) Liu, Y.; Zhang, L.; Chen, X.; Chen, D.; Shi, X.; Song, J.; Wu, J.; Huang, F.; Xia, Q.; Xiang, Y.; Zheng, X.; Cai, Y. The Novel FGFR Inhibitor F1-7 Induces DNA Damage and Cell Death in Colon Cells. *Br. J. Cancer* **2022**, *127* (6), 1014–1025. https://doi.org/10.1038/s41416-022-01878-4.
(368) Kang, X.; Lin, Z.; Xu, M.; Pan, J.; Wang, Z. Deciphering Role of FGFR Signalling Pathway in Pancreatic Cancer. *Cell Prolif.* **2019**, *52* (3), e12605. https://doi.org/10.1111/cpr.12605.
(369) Phanhthilath, N.; Hakim, S.; Su, C.; Liu, A.; Subramonian, D.; Lesperance, J.; Zage, P. E. Mechanisms of Efficacy of the FGFR1–3 Inhibitor AZD4547 in Pediatric Solid Tumor Models. *Invest. New Drugs* **2020**, *38* (6), 1677–1686. https://doi.org/10.1007/s10637-020-00933-2.
(370) Lo, A. K.-F.; Dawson, C. W.; Young, L. S.; Ko, C.-W.; Hau, P.-M.; Lo, K.-W. Activation of the FGFR1 Signalling Pathway by the Epstein–Barr Virus-Encoded LMP1 Promotes Aerobic Glycolysis and Transformation of Human Nasopharyngeal Epithelial Cells. *J. Pathol.* **2015**, *237* (2), 238–248. https://doi.org/10.1002/path.4575.
(371) Krause, M.; Gautreau, A. Steering Cell Migration: Lamellipodium Dynamics and the Regulation of Directional Persistence. *Nat. Rev. Mol. Cell Biol.* **2014**, *15* (9), 577–590. https://doi.org/10.1038/nrm3861.
(372) Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P. The Cytoskeleton and Cell Behavior. In *Molecular Biology of the Cell. 4th edition*; Garland Science, 2002.
(373) Goswami, S.; Philippar, U.; Sun, D.; Patsialou, A.; Avraham, J.; Wang, W.; Di Modugno, F.; Nistico, P.; Gertler, F. B.; Condeelis, J. S. Identification of Invasion Specific Splice Variants of the Cytoskeletal Protein Mena Present in Mammary

Tumor Cells during Invasion in Vivo. *Clin. Exp. Metastasis* **2009**, *26* (2), 153–159. https://doi.org/10.1007/s10585-008-9225-8.
(374) Varland, S.; Vandekerckhove, J.; Drazic, A. Actin Post-Translational Modifications: The Cinderella of Cytoskeletal Control. *Trends Biochem. Sci.* **2019**, *44* (6), 502–516. https://doi.org/10.1016/j.tibs.2018.11.010.
(375) Martineau, Y.; Müller, D.; Pyronnet, S. Targeting Protein Synthesis in Cancer Cells. *Oncoscience* **2014**, *1* (7), 484–485.
(376) Houles, T.; Roux, P. P. Defining the Role of the RSK Isoforms in Cancer. *Semin. Cancer Biol.* **2018**, *48*, 53–61. https://doi.org/10.1016/j.semcancer.2017.04.016.
(377) Chambers, R. W. On the Recognition of TRNA by Its Aminoacyl-TRNA Liqase121Work in the Author's Laboratory Referred to Here Was Supported by Grants from the U.S. Public Health Service (Grant 5 RO 1 GM7262) and the American Cancer Society (Grant P-476).2Enzymes Catalyzing the Linking Together of Two Molecules, Coupled with the Breaking of a Pyrophosphate Link in ATP, Etc., Had, before the Enzyme Commission Reports ["Comprehensive Biochemistry" (M. Florkin and E. H. Stotz, Eds.) Vol. 13, Amsterdam, Elsevier, 19651, Been Known as Synthetases. Since This Terminology Differs from All Other Systematic Enzyme Names in That It Is Based on the Product and Not on the Substrate, the Enzyme Commission Decided That a New Systematic Name Was Necessary and Chose the Naine "Ligase," "Synthetase" Was Retained as a Trivial Name Solely in Order to Avoid Extensive Changes in the Existing Nomenclatnre. [Eds.]. In *Progress in Nucleic Acid Research and Molecular Biology*; Davidson, J. N., Cohn, W. E., Eds.; Academic Press, 1971; Vol. 11, pp 489–525. https://doi.org/10.1016/S0079-6603(08)60336-0.
(378) Shi, Y.; Wang, Y.; Bingle, L.; Gong, L.; Heng, W.; Li, Y.; Fang, W. In Vitro Study of HIF-1 Activation and VEGF Release by BFGF in the T47D Breast Cancer Cell Line under Normoxic Conditions: Involvement of PI-3K/Akt and MEK1/ERK Pathways. *J. Pathol.* **2005**, *205* (4), 530–536. https://doi.org/10.1002/path.1734.
(379) Yamaguchi, H.; Condeelis, J. Regulation of the Actin Cytoskeleton in Cancer Cell Migration and Invasion. *Biochim. Biophys. Acta BBA - Mol. Cell Res.* **2007**, *1773* (5), 642–652. https://doi.org/10.1016/j.bbamcr.2006.07.001.
(380) Yan, J.; Mihaylov, V.; Xu, X.; Brzostowski, J. A.; Li, H.; Liu, L.; Veenstra, T. D.; Parent, C. A.; Jin, T. A Gβγ Effector, ElmoE, Transduces GPCR Signaling to the Actin Network during Chemotaxis. *Dev. Cell* **2012**, *22* (1), 92–103. https://doi.org/10.1016/j.devcel.2011.11.007.
(381) Aplin, A. E.; Juliano, R. L. Integrin and Cytoskeletal Regulation of Growth Factor Signaling to the MAP Kinase Pathway. *J. Cell Sci.* **1999**, *112 (Pt 5)*, 695–706. https://doi.org/10.1242/jcs.112.5.695.
(382) Lee, J. M. *The Actin Cytoskeleton and the Regulation of Cell Migration*; 2013; Vol. 1.
(383) Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P. The Self-Assembly and Dynamic Structure of Cytoskeletal Filaments. In *Molecular Biology of the Cell*; Garland Science, 2002.
(384) Swaney, K. F.; Li, R. Function and Regulation of the Arp2/3 Complex during Cell Migration in Diverse Environments. *Curr. Opin. Cell Biol.* **2016**, *42*, 63–72. https://doi.org/10.1016/j.ceb.2016.04.005.

(385) Yarmola, E. G.; Bubb, M. R. Profilin: Emerging Concepts and Lingering Misconceptions. *Trends Biochem. Sci.* **2006**, *31* (4), 197–205. https://doi.org/10.1016/j.tibs.2006.02.006.
(386) Small, J. V.; Rottner, K.; Kaverina, I. Functional Design in the Actin Cytoskeleton. *Curr. Opin. Cell Biol.* **1999**, *11* (1), 54–60. https://doi.org/10.1016/S0955-0674(99)80007-6.
(387) Witke, W.; Sharpe, A. H.; Hartwig, J. H.; Azuma, T.; Stossel, T. P.; Kwiatkowski, D. J. Hemostatic, Inflammatory, and Fibroblast Responses Are Blunted in Mice Lacking Gelsolin. *Cell* **1995**, *81* (1), 41–51. https://doi.org/10.1016/0092-8674(95)90369-0.
(388) Ouderkirk, J. L.; Krendel, M. Non-Muscle Myosins in Tumor Progression, Cancer Cell Invasion and Metastasis. *Cytoskelet. Hoboken NJ* **2014**, *71* (8), 447–463. https://doi.org/10.1002/cm.21187.
(389) Smillie, L. B. Structure and Functions of Tropomyosins from Muscle and Non-Muscle Sources. *Trends Biochem. Sci.* **1979**, *4* (7), 151–155. https://doi.org/10.1016/0968-0004(79)90003-3.
(390) Anderson, T. W.; Vaughan, A. N.; Cramer, L. P. Retrograde Flow and Myosin II Activity within the Leading Cell Edge Deliver F-Actin to the Lamella to Seed the Formation of Graded Polarity Actomyosin II Filament Bundles in Migrating Fibroblasts. *Mol. Biol. Cell* **2008**, *19* (11), 5006–5018. https://doi.org/10.1091/mbc.E08-01-0034.
(391) Sanborn, K. B.; Mace, E. M.; Rak, G. D.; Difeo, A.; Martignetti, J. A.; Pecci, A.; Bussel, J. B.; Favier, R.; Orange, J. S. Phosphorylation of the Myosin IIA Tailpiece Regulates Single Myosin IIA Molecule Association with Lytic Granules to Promote NK-Cell Cytotoxicity. *Blood* **2011**, *118* (22), 5862–5871. https://doi.org/10.1182/blood-2011-03-344846.
(392) Rigbolt, K. T. G.; Prokhorova, T. A.; Akimov, V.; Henningsen, J.; Johansen, P. T.; Kratchmarova, I.; Kassem, M.; Mann, M.; Olsen, J. V.; Blagoev, B. System-Wide Temporal Characterization of the Proteome and Phosphoproteome of Human Embryonic Stem Cell Differentiation. *Sci. Signal.* **2011**, *4* (164), rs3. https://doi.org/10.1126/scisignal.2001570.
(393) Hu, L.; Fang, L.; Zhang, Z.-P.; Yan, Z.-L. TPM1 Is a Novel Predictive Biomarker for Gastric Cancer Diagnosis and Prognosis. *Clin. Lab.* **2020**, *66* (4). https://doi.org/10.7754/Clin.Lab.2019.190235.
(394) Yang, R.; Zheng, G.; Ren, D.; Chen, C.; Zeng, C.; Lu, W.; Li, H. The Clinical Significance and Biological Function of Tropomyosin 4 in Colon Cancer. *Biomed. Pharmacother. Biomedecine Pharmacother.* **2018**, *101*, 1–7. https://doi.org/10.1016/j.biopha.2018.01.166.
(395) Nicholson-Dykstra, S. M.; Higgs, H. N. Arp2 Depletion Inhibits Sheet-like Protrusions but Not Linear Protrusions of Fibroblasts and Lymphocytes. *Cell Motil. Cytoskeleton* **2008**, *65* (11), 904–922. https://doi.org/10.1002/cm.20312.
(396) Rogers, S. L.; Wiedemann, U.; Stuurman, N.; Vale, R. D. Molecular Requirements for Actin-Based Lamella Formation in Drosophila S2 Cells. *J. Cell Biol.* **2003**, *162* (6), 1079–1088. https://doi.org/10.1083/jcb.200303023.
(397) Suraneni, P.; Fogelson, B.; Rubinstein, B.; Noguera, P.; Volkmann, N.; Hanein, D.; Mogilner, A.; Li, R. A Mechanism of Leading-Edge Protrusion in the Absence of Arp2/3 Complex. *Mol. Biol. Cell* **2015**, *26* (5), 901–912. https://doi.org/10.1091/mbc.E14-07-1250.

(398) Yamaguchi, H.; Lorenz, M.; Kempiak, S.; Sarmiento, C.; Coniglio, S.; Symons, M.; Segall, J.; Eddy, R.; Miki, H.; Takenawa, T.; Condeelis, J. Molecular Mechanisms of Invadopodium Formation. *J. Cell Biol.* **2005**, *168* (3), 441–452. https://doi.org/10.1083/jcb.200407076.
(399) Chan, A. Y.; Bailly, M.; Zebda, N.; Segall, J. E.; Condeelis, J. S. Role of Cofilin in Epidermal Growth Factor–Stimulated Actin Polymerization and Lamellipod Protrusion. *J. Cell Biol.* **2000**, *148* (3), 531–542.
(400) DesMarais, V.; Macaluso, F.; Condeelis, J.; Bailly, M. Synergistic Interaction between the Arp2/3 Complex and Cofilin Drives Stimulated Lamellipod Extension. *J. Cell Sci.* **2004**, *117* (Pt 16), 3499–3510. https://doi.org/10.1242/jcs.01211.
(401) Ostrowska, Z.; Moraczewska, J. Cofilin - a Protein Controlling Dynamics of Actin Filaments. *Postepy Hig. Med. Doswiadczalnej Online* **2017**, *71* (0), 339–351. https://doi.org/10.5604/01.3001.0010.3818.
(402) Bamburg, J. R. Proteins of the ADF/Cofilin Family: Essential Regulators of Actin Dynamics. *Annu. Rev. Cell Dev. Biol.* **1999**, *15* (1), 185–230. https://doi.org/10.1146/annurev.cellbio.15.1.185.
(403) Agnew, B. J.; Minamide, L. S.; Bamburg, J. R. Reactivation of Phosphorylated Actin Depolymerizing Factor and Identification of the Regulatory Site *. *J. Biol. Chem.* **1995**, *270* (29), 17582–17587. https://doi.org/10.1074/jbc.270.29.17582.
(404) Niwa, R.; Nagata-Ohashi, K.; Takeichi, M.; Mizuno, K.; Uemura, T. Control of Actin Reorganization by Slingshot, a Family of Phosphatases That Dephosphorylate ADF/Cofilin. *Cell* **2002**, *108* (2), 233–246. https://doi.org/10.1016/S0092-8674(01)00638-9.
(405) Pennington, K. L.; Chan, T. Y.; Torres, M. P.; Andersen, J. L. The Dynamic and Stress-Adaptive Signaling Hub of 14-3-3: Emerging Mechanisms of Regulation and Context-Dependent Protein–Protein Interactions. *Oncogene* **2018**, *37* (42), 5587–5604. https://doi.org/10.1038/s41388-018-0348-3.
(406) Gohla, A.; Bokoch, G. M. 14-3-3 Regulates Actin Dynamics by Stabilizing Phosphorylated Cofilin. *Curr. Biol. CB* **2002**, *12* (19), 1704–1710. https://doi.org/10.1016/s0960-9822(02)01184-3.
(407) Uhart, M.; Bustos, D. M. Human 14-3-3 Paralogs Differences Uncovered by Cross-Talk of Phosphorylation and Lysine Acetylation. *PLOS ONE* **2013**, *8* (2), e55703. https://doi.org/10.1371/journal.pone.0055703.
(408) Choudhary, C.; Kumar, C.; Gnad, F.; Nielsen, M. L.; Rehman, M.; Walther, T. C.; Olsen, J. V.; Mann, M. Lysine Acetylation Targets Protein Complexes and Co-Regulates Major Cellular Functions. *Science* **2009**, *325* (5942), 834–840. https://doi.org/10.1126/science.1175371.
(409) Tsai, C.-F.; Wang, Y.-T.; Yen, H.-Y.; Tsou, C.-C.; Ku, W.-C.; Lin, P.-Y.; Chen, H.-Y.; Nesvizhskii, A. I.; Ishihama, Y.; Chen, Y.-J. Large-Scale Determination of Absolute Phosphorylation Stoichiometries in Human Cells by Motif-Targeting Quantitative Proteomics. *Nat. Commun.* **2015**, *6*, 6622. https://doi.org/10.1038/ncomms7622.
(410) Bokoch, G. M.; Vlahos, C. J.; Wang, Y.; Knaus, U. G.; Traynor-Kaplan, A. E. Rac GTPase Interacts Specifically with Phosphatidylinositol 3-Kinase. *Biochem. J.* **1996**, *315* (Pt 3), 775–779.
(411) Head, J. A.; Jiang, D.; Li, M.; Zorn, L. J.; Schaefer, E. M.; Parsons, J. T.; Weed, S. A. Cortactin Tyrosine Phosphorylation Requires Rac1 Activity and Association with

the Cortical Actin Cytoskeleton. *Mol. Biol. Cell* **2003**, *14* (8), 3216–3229. https://doi.org/10.1091/mbc.E02-11-0753.
(412) Kelley, L. C.; Hayes, K. E.; Ammer, A. G.; Martin, K. H.; Weed, S. A. Cortactin Phosphorylated by ERK1/2 Localizes to Sites of Dynamic Actin Regulation and Is Required for Carcinoma Lamellipodia Persistence. *PloS One* **2010**, *5* (11), e13847. https://doi.org/10.1371/journal.pone.0013847.
(413) Manabe, R.; Kovalenko, M.; Webb, D. J.; Horwitz, A. R. GIT1 Functions in a Motile, Multi-Molecular Signaling Complex That Regulates Protrusive Activity and Cell Migration. *J. Cell Sci.* **2002**, *115* (Pt 7), 1497–1510. https://doi.org/10.1242/jcs.115.7.1497.
(414) Dephoure, N.; Zhou, C.; Villén, J.; Beausoleil, S. A.; Bakalarski, C. E.; Elledge, S. J.; Gygi, S. P. A Quantitative Atlas of Mitotic Phosphorylation. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (31), 10762–10767. https://doi.org/10.1073/pnas.0805139105.
(415) Zahedi, R. P.; Lewandrowski, U.; Wiesner, J.; Wortelkamp, S.; Moebius, J.; Schütz, C.; Walter, U.; Gambaryan, S.; Sickmann, A. Phosphoproteome of Resting Human Platelets. *J. Proteome Res.* **2008**, *7* (2), 526–534. https://doi.org/10.1021/pr0704130.
(416) Bryant, D. A.; Guglielmi, G.; de Marsac, N. T.; Castets, A.-M.; Cohen-Bazire, G. The Structure of Cyanobacterial Phycobilisomes: A Model. *Arch. Microbiol.* **1979**, *123* (2), 113–127. https://doi.org/10.1007/BF00446810.
(417) Gantt, E.; Lipschultz, C. A. Energy Transfer in Phycobilisomes from Phycoerythrin to Allophycocyanin. *Biochim. Biophys. Acta BBA - Bioenerg.* **1973**, *292* (3), 858–861. https://doi.org/10.1016/0005-2728(73)90036-4.
(418) Gantt, E.; Lipschultz, C. A.; Zilinskas, B. Further Evidence for a Phycobilisome Model from Selective Dissociation, Fluorescence Emission, Immunoprecipitation, and Electron Microscopy. *Biochim. Biophys. Acta BBA - Bioenerg.* **1976**, *430* (2), 375–388. https://doi.org/10.1016/0005-2728(76)90093-1.
(419) Zilinskas, B. A.; Glick, R. E. Noncovalent Intermolecular Forces in Phycobilisomes of Porphyridium Cruentum 1 2. *Plant Physiol.* **1981**, *68* (2), 447–452. https://doi.org/10.1104/pp.68.2.447.
(420) Merchant, S. S.; Helmann, J. D. Chapter 2 - Elemental Economy: Microbial Strategies for Optimizing Growth in the Face of Nutrient Limitation. In *Advances in Microbial Physiology*; Poole, R. K., Ed.; Academic Press, 2012; Vol. 60, pp 91–210. https://doi.org/10.1016/B978-0-12-398264-3.00002-4.
(421) Tambussi, E. A.; Bartoli, C. G.; Beltrano, J.; Guiamet, J. J.; Araus, J. L. Oxidative Damage to Thylakoid Proteins in Water-Stressed Leaves of Wheat (Triticum Aestivum). *Physiol. Plant.* **2000**, *108* (4), 398–404. https://doi.org/10.1034/j.1399-3054.2000.t01-1-100409.x.
(422) Padyana, A. K.; Bhat, V. B.; Madyastha, K. M.; Rajashankar, K. R.; Ramakumar, S. Crystal Structure of a Light-Harvesting Protein C-Phycocyanin from Spirulina Platensis. *Biochem. Biophys. Res. Commun.* **2001**, *282* (4), 893–898. https://doi.org/10.1006/bbrc.2001.4663.
(423) Altschul, S. F.; Madden, T. L.; Schäffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs. *Nucleic Acids Res.* **1997**, *25* (17), 3389–3402. https://doi.org/10.1093/nar/25.17.3389.
(424) Mailloux, R. J.; Willmore, W. G. S-Glutathionylation Reactions in Mitochondrial Function and Disease. *Front. Cell Dev. Biol.* **2014**, *2*.

(425) Rinalducci, S.; Pedersen, J. Z.; Zolla, L. Generation of Reactive Oxygen Species upon Strong Visible Light Irradiation of Isolated Phycobilisomes from Synechocystis PCC 6803. *Biochim. Biophys. Acta BBA - Bioenerg.* **2008**, *1777* (5), 417–424. https://doi.org/10.1016/j.bbabio.2008.02.005.
(426) Tamary, E.; Kiss, V.; Nevo, R.; Adam, Z.; Bernát, G.; Rexroth, S.; Rögner, M.; Reich, Z. Structural and Functional Alterations of Cyanobacterial Phycobilisomes Induced by High-Light Stress. *Biochim. Biophys. Acta BBA - Bioenerg.* **2012**, *1817* (2), 319–327. https://doi.org/10.1016/j.bbabio.2011.11.008.
(427) Neill, S.; Desikan, R.; Hancock, J. Hydrogen Peroxide Signalling. *Curr. Opin. Plant Biol.* **2002**, *5* (5), 388–395. https://doi.org/10.1016/s1369-5266(02)00282-0.
(428) Niu, L.; Liao, W. Hydrogen Peroxide Signaling in Plant Development and Abiotic Responses: Crosstalk with Nitric Oxide and Calcium. *Front. Plant Sci.* **2016**, *7*.
(429) Giorgio, M.; Trinei, M.; Migliaccio, E.; Pelicci, P. G. Hydrogen Peroxide: A Metabolic by-Product or a Common Mediator of Ageing Signals? *Nat. Rev. Mol. Cell Biol.* **2007**, *8* (9), 722–728. https://doi.org/10.1038/nrm2240.
(430) Belcastro, E.; Gaucher, C.; Corti, A.; Leroy, P.; Lartaud, I.; Pompella, A. Regulation of Protein Function by S-Nitrosation and S-Glutathionylation: Processes and Targets in Cardiovascular Pathophysiology. *Biol. Chem.* **2017**, *398* (12), 1267–1293. https://doi.org/10.1515/hsz-2017-0150.
(431) Jin, Y.; Liu, Z.; Li, Y.; Liu, W.; Tao, Y.; Wang, G. A Structural and Functional Study on the 2-C-Methyl-D-Erythritol-4-Phosphate Cytidyltransferase (IspD) from Bacillus Subtilis. *Sci. Rep.* **2016**, *6* (1), 36379. https://doi.org/10.1038/srep36379.
(432) August, J. T.; Ortiz, P. J.; Hurwitz, J. Ribonucleic Acid-Dependent Ribonucleotide Incorporation: I. PURIFICATION AND PROPERTIES OF THE ENZYME. *J. Biol. Chem.* **1962**, *237* (12), 3786–3793. https://doi.org/10.1016/S0021-9258(19)84523-4.
(433) Yu, J.; Loh, K.; Song, Z.; Yang, H.; Zhang, Y.; Lin, S. Update on Glycerol-3-Phosphate Acyltransferases: The Roles in the Development of Insulin Resistance. *Nutr. Diabetes* **2018**, *8* (1), 1–10. https://doi.org/10.1038/s41387-018-0045-x.
(434) Chen, Y.-J.; Lu, C.-T.; Huang, K.-Y.; Wu, H.-Y.; Chen, Y.-J.; Lee, T.-Y. GSHSite: Exploiting an Iteratively Statistical Method to Identify S-Glutathionylation Sites with Substrate Specificity. *PLOS ONE* **2015**, *10* (4), e0118752. https://doi.org/10.1371/journal.pone.0118752.
(435) Mikesh, L. M.; Ueberheide, B.; Chi, A.; Coon, J. J.; Syka, J. E. P.; Shabanowitz, J.; Hunt, D. F. The Utility of ETD Mass Spectrometry in Proteomic Analysis. *Biochim. Biophys. Acta* **2006**, *1764* (12), 1811–1822. https://doi.org/10.1016/j.bbapap.2006.10.003.
(436) McConnell, E. W.; Werth, E. G.; Hicks, L. M. The Phosphorylated Redox Proteome of Chlamydomonas Reinhardtii: Revealing Novel Means for Regulation of Protein Structure and Function. *Redox Biol.* **2018**, *17*, 35–46. https://doi.org/10.1016/j.redox.2018.04.003.
(437) Nagarkoti, S.; Dubey, M.; Awasthi, D.; Kumar, V.; Chandra, T.; Kumar, S.; Dikshit, M. S-Glutathionylation of P47phox Sustains Superoxide Generation in Activated Neutrophils. *Biochim. Biophys. Acta BBA - Mol. Cell Res.* **2018**, *1865* (2), 444–454. https://doi.org/10.1016/j.bbamcr.2017.11.014.
(438) Foyer, C. H. Reactive Oxygen Species, Oxidative Signaling and the Regulation of Photosynthesis. *Environ. Exp. Bot.* **2018**, *154*, 134–142. https://doi.org/10.1016/j.envexpbot.2018.05.003.

(439) Wang, Y.-C.; Peterson, S. E.; Loring, J. F. Protein Post-Translational Modifications and Regulation of Pluripotency in Human Stem Cells. *Cell Res.* **2014**, *24* (2), 143–160. https://doi.org/10.1038/cr.2013.151.
(440) Venne, A. S; Zahedi, R. P. *The potential of fractional diagonal chromatography strategies for the enrichment of post-translational modifications | Elsevier Enhanced Reader*. https://doi.org/10.1016/j.euprot.2014.07.001.
(441) Duong, V.-A.; Park, J.-M.; Lee, H. Review of Three-Dimensional Liquid Chromatography Platforms for Bottom-Up Proteomics. *Int. J. Mol. Sci.* **2020**, *21* (4), 1524. https://doi.org/10.3390/ijms21041524.
(442) Dit Fouque, K. J.; Kaplan, D.; Voinov, V. G.; Holck, F. H. V.; Jensen, O. N.; Fernandez-Lima, F. Proteoform Differentiation Using Tandem Trapped Ion Mobility, Electron Capture Dissociation, and ToF Mass Spectrometry. *Anal. Chem.* **2021**, *93* (27), 9575–9582. https://doi.org/10.1021/acs.analchem.1c01735.
(443) Doll, S.; Burlingame, A. L. Mass Spectrometry-Based Detection and Assignment of Protein Posttranslational Modifications. *ACS Chem. Biol.* **2015**, *10* (1), 63–71. https://doi.org/10.1021/cb500904b.
(444) Xue, J.; Domingo-Almenara, X.; Guijas, C.; Palermo, A.; Rinschen, M. M.; Isbell, J.; Benton, H. P.; Siuzdak, G. Enhanced In-Source Fragmentation Annotation Enables Novel Data Independent Acquisition and Autonomous METLIN Molecular Identification. *Anal. Chem.* **2020**, *92* (8), 6051–6059. https://doi.org/10.1021/acs.analchem.0c00409.
(445) Zhang, F.; Ge, W.; Ruan, G.; Cai, X.; Guo, T. Data-Independent Acquisition Mass Spectrometry-Based Proteomics and Software Tools: A Glimpse in 2020. *PROTEOMICS* **2020**, *20* (17–18), 1900276. https://doi.org/10.1002/pmic.201900276.
(446) Tsai, P. L.; Chen, S.-F.; Huang, S. Y. Mass Spectrometry-Based Strategies for Protein Disulfide Bond Identification. *Rev. Anal. Chem.* **2013**, *32* (4), 257–268. https://doi.org/10.1515/revac-2013-0011.
(447) Huang, K.; Huang, L.; van Breemen, R. B. Detection of Reactive Metabolites Using Isotope-Labeled Glutathione Trapping and Simultaneous Neutral Loss and Precursor Ion Scanning with Ultra-High-Pressure Liquid Chromatography Triple Quadrupole Mass Spectrometry. *Anal. Chem.* **2015**, *87* (7), 3646–3654. https://doi.org/10.1021/ac504737x.

8. Appendix

8.1. Proteome Discoverer Search Parameters for Pairwise PTM Search (Chapter 3)

Processing Workflow

The workflow tree:

|-(0) Spectrum Files
  |-(1) Spectrum Selector
    |-(2) Sequest HT
      |-(4) Percolator
        |-(5) IMP-ptmRS

Processing node 0: Spectrum Files

Input Data:
- File Name(s): K:\Kish\FAIMS experiment\Experiment 3\RAW files\HeLa_FAIMS_75.raw

---
Processing node 1: Spectrum Selector
---
1. General Settings:
- Precursor Selection: Use MS1 Precursor
- Use Isotope Pattern in Precursor Reevaluation: True
- Provide Profile Spectra: Automatic

2. Spectrum Properties Filter:
- Lower RT Limit: 5
- Upper RT Limit: 0
- First Scan: 0
- Last Scan: 0
- Lowest Charge State: 0
- Highest Charge State: 0
- Min. Precursor Mass: 700 Da
- Max. Precursor Mass: 7000 Da
- Total Intensity Threshold: 0
- Minimum Peak Count: 1

3. Scan Event Filters:
- MS Order: Is Not MS1
- Min. Collision Energy: 0
- Max. Collision Energy: 1000
- Scan Type: Is Full

4. Peak Filters:
- S/N Threshold (FT-only): 1.5

5. Replacements for Unrecognized Properties:
- Unrecognized Charge Replacements: Automatic
- Unrecognized Mass Analyzer Replacements: ITMS
- Unrecognized MS Order Replacements: MS2
- Unrecognized Activation Type Replacements: CID
- Unrecognized Polarity Replacements: +
- Unrecognized MS Resolution@200 Replacements: 60000
- Unrecognized MSn Resolution@200 Replacements: 30000

6. Precursor Pattern Extraction:
- Precursor Clipping Range Before: 2.5 Da
- Precursor Clipping Range After: 5.5 Da

Processing node 2: Sequest HT
---
1. Input Data:
- Protein Database: HUMAN_160919_reviewed_uniprot-20432_Sept19.fasta
- Enzyme Name: Trypsin (Full)
- Max. Missed Cleavage Sites: 4
- Min. Peptide Length: 6
- Max. Peptide Length: 50
- Max. Number of Peptides Reported: 10

2. Tolerances:
- Precursor Mass Tolerance: 10 ppm
- Fragment Mass Tolerance: 0.5 Da
- Use Average Precursor Mass: False
- Use Average Fragment Mass: False

3. Spectrum Matching:
- Use Neutral Loss a Ions: True
- Use Neutral Loss b Ions: True
- Use Neutral Loss y Ions: True
- Use Flanking Ions: True
- Weight of a Ions: 0
- Weight of b Ions: 1
- Weight of c Ions: 0
- Weight of x Ions: 0
- Weight of y Ions: 1
- Weight of z Ions: 0

4. Dynamic Modifications:
- Max. Equal Modifications Per Peptide: 5
- Max. Dynamic Modifications Per Peptide: 7

- 1. Dynamic Modification: Oxidation / +15.995 Da (M)
- 2. Dynamic Modification: Phospho / +79.966 Da (S, T, Y)
- 3. Dynamic Modification: Methyl / +14.016 Da (K, R)

6. Dynamic Modifications (protein terminus):
- 1. N-Terminal Modification: Met-loss / -131.040 Da (M)
- 2. N-Terminal Modification: Met-loss+Acetyl / -89.030 Da (M)

7. Static Modifications:
- 1. Static Modification: Carbamidomethyl / +57.021 Da (C)

Processing node 4: Percolator
---
1. Target/Decoy Strategy:
- Target/Decoy Selection: Concatenated
- Validation based on: q-Value

2. Input Data:
- Maximum Delta Cn: 0.05
- Maximum Rank: 0

3. FDR Targets:
- Target FDR (Strict): 0.01
- Target FDR (Relaxed): 0.05

Processing node 5: IMP-ptmRS
---
1. Scoring:
- PhosphoRS Mode: False
- Report Only PTMs: True
- Use Diagnostic Ions: True
- Use Fragment Mass Tolerance of Search Node: True
- Fragment Mass Tolerance: 0.5 Da
- Consider neutral loss peaks for CID, HCD and EThcD: Automatic
- Maximum Peak Depth: 8
- Use a mass accuracy correction: False

2. Performance:
- Maximum Number of Position Isoforms: 500
- Maximum PTMs per peptide: 10

Consensus Workflow

The workflow tree:

```
  |-(0) MSF Files
    |-(1) PSM Grouper
      |-(2) Peptide Validator
        |-(3) Peptide and Protein Filter
          |-(4) Protein Scorer
            |-(5) Protein Grouping
              |-(12) Peptide in Protein Annotation
              |-(13) Modification Sites
            |-(6) Protein FDR Validator
          |-(11) Peptide Isoform Grouper
          |-(10) Protein Annotation

Post-processing nodes:
------------------------------

  |-(7) Display Settings
  |-(8) Result Statistics
  |-(9) Data Distributions

---------------------------------------------------------------------
Processing node 0: MSF Files
---------------------------------------------------------------------
1. Storage Settings:
- Spectra to Store:  Identified or Quantified
- Feature Traces to Store:  All

2. Merging of Identified Peptide and Proteins:
- Merge Mode:  Globally by Search Engine Type

3. FASTA Title Line Display:
- Reported FASTA Title Lines:  Best match
- Title Line Rule:  standard

4. PSM Filters:
- Maximum Delta Cn:  0.05
- Maximum Rank:  0
- Maximum Delta Mass:  0 ppm

Hidden Parameters:
- MSF File(s) (Hidden):
        D:\Kish\PD stuff\Study File\HeLa_FAIMS_45.msf
        D:\Kish\PD stuff\Study File\HeLa_FAIMS_52.msf
        D:\Kish\PD stuff\Study File\HeLa_FAIMS_52_(2).msf
        D:\Kish\PD stuff\Study File\HeLa_FAIMS_52_(3).msf
        D:\Kish\PD stuff\Study File\HeLa_FAIMS_67_(1).msf
        D:\Kish\PD stuff\Study File\HeLa_FAIMS_67_(2).msf
        D:\Kish\PD stuff\Study File\HeLa_FAIMS_67_(3).msf
```

D:\Kish\PD stuff\Study File\HeLa_FAIMS_75.msf
D:\Kish\PD stuff\Study File\HeLa_FAIMS_90.msf
D:\Kish\PD stuff\Study File\HeLa_FAIMS_90_(2).msf
D:\Kish\PD stuff\Study File\HeLa_FAIMS_Sweep.msf
D:\Kish\PD stuff\Study File\NoFAIMS_HeLa.msf
D:\Kish\PD stuff\Study File\HeLa_FAIMS_0.msf
D:\Kish\PD stuff\Study File\HeLa_FAIMS_0_(2).msf
D:\Kish\PD stuff\Study File\2020-08-19_1ug_HeLa_FAIMS-45_replicate1.msf
D:\Kish\PD stuff\Study File\2020-08-19_1ug_HeLa_FAIMS-45_replicate2.msf
D:\Kish\PD stuff\Study File\2020-08-19_1ug_HeLa_FAIMS-60_replicate1.msf
D:\Kish\PD stuff\Study File\2020-08-19_1ug_HeLa_FAIMS-60_replicate2.msf
D:\Kish\PD stuff\Study File\2020-08-19_1ug_HeLa_FAIMS-60_replicate3.msf
D:\Kish\PD stuff\Study File\2020-08-19_1ug_HeLa_FAIMS-75_replicate1.msf
D:\Kish\PD stuff\Study File\2020-08-19_1ug_HeLa_FAIMS-75_replicate2-(2).msf
D:\Kish\PD stuff\Study File\2020-08-19_1ug_HeLa_FAIMS-90_replicate2-(2).msf
D:\Kish\PD stuff\Study File\2020-08-19_1ug_HeLa_FAIMSsweep_replicate1-(2).msf
D:\Kish\PD stuff\Study File\2020-08-19_1ug_HeLa_FAIMSsweep_replicate2-(2).msf
D:\Kish\PD stuff\Study File\2020-08-19_1ug_HeLa_FAIMSvoltoff_replicate2-(2).msf
D:\Kish\PD stuff\Study File\2020-08-19_1ug_HeLa_noFAIMS_replicate1-(2).msf
D:\Kish\PD stuff\Study File\2020-08-19_1ug_HeLa_noFAIMS_replicate2-(2).msf

---
Processing node 1: PSM Grouper
---
1. Peptide Group Modifications:
- Site Probability Threshold: 75

Processing node 2: Peptide Validator
---
1. General Validation Settings:
- Validation Mode: Automatic (Control peptide level error rate if possible)
- Target FDR (Strict) for PSMs: 0.01
- Target FDR (Relaxed) for PSMs: 0.05
- Target FDR (Strict) for Peptides: 0.01

- Target FDR (Relaxed) for Peptides: 0.05

2. Specific Validation Settings:
- Validation Based on: q-Value
- Target/Decoy Selection for PSM Level FDR Calculation Based on Score: Automatic
- Reset Confidences for Nodes without Decoy Search (Fixed score thresholds): False

---
Processing node 3: Peptide and Protein Filter
---
1. Peptide Filters:
- Peptide Confidence At Least: Medium
- Keep Lower Confident PSMs: False
- Minimum Peptide Length: 6
- Remove Peptides Without Protein Reference: False

2. Protein Filters:
- Minimum Number of Peptide Sequences: 1
- Count Only Rank 1 Peptides: False
- Count Peptides Only for Top Scored Protein: False

Processing node 4: Protein Scorer

No parameters

Processing node 5: Protein Grouping
---
1. Protein Grouping:
- Apply strict parsimony principle: True

Processing node 12: Peptide in Protein Annotation
---
1. Flanking Residues:
- Annotate Flanking Residues of the Peptide: True
- Number Flanking Residues in Connection Tables: 1

2. Modifications in Peptide:
- Protein Modifications Reported: Only for Master Proteins

3. Modifications in Protein:
- Modification Sites Reported: All And Specific
- Minimum PSM Confidence: High
- Report Only PTMs: True

4. Positions in Protein:
- Protein Positions for Peptides: Only for Master Proteins

Processing node 13: Modification Sites
---
1. General:
- Report Only PTMs: True
- Only Master Proteins: True
- Motif Radius: 6

Processing node 6: Protein FDR Validator
---
1. Confidence Thresholds:
- Target FDR (Strict): 0.01
- Target FDR (Relaxed): 0.05

Processing node 11: Peptide Isoform Grouper

No parameters

Processing node 10: Protein Annotation
---
1. Annotation Aspects:
- 1. Aspect: Biological Process
- 2. Aspect: Cellular Component
- 3. Aspect: Molecular Function
- 4. Aspect: None
- 5. Aspect: None
- 6. Aspect: None

Processing node 7: Display Settings
---
1. General:
- Filter Set:
	### Filter Set MasterProteinFilter.filterset contains the following filters:
	### Row Filter for TargetProtein:
	### Master is equal to Master
	###

'magellan filter set' 1 'MasterProteinFilter.filterset' FiltersetProperties 1 'LastFileName' 'C:\Users\frank.berg\Desktop\MasterProteinFilter.filterset' Filter 'TargetProtein' 1 NARY_AND 1 = FilterConditionProperties 1 'NamedComparableFilterCondition/DisplayPropertyHint' 'Master' property

'Thermo.PD.EntityDataFramework.MasterProteinAssessment, Thermo.Magellan.EntityDataFramework' 'IsMasterProtein' constant 'Thermo.PD.EntityDataFramework.MasterProteinAssessment, Thermo.Magellan.EntityDataFramework' 'IsMasterProtein'

Processing node 8: Result Statistics

No parameters

Processing node 9: Data Distributions

---

1. ID Distributions (Bottom-up):
- Peptides to Use: All PSMs

8.2. Proteome Discoverer Search Parameters for Broad PTM Search (Chapter 3)

Processing Workflow

The workflow tree:

|-(0) Spectrum Files
  |-(1) Spectrum Selector
    |-(5) PMI-Byonic
      |-(4) Percolator
        |-(6) IMP-ptmRS

Processing node 1: Spectrum Selector
---
1. General Settings:
- Precursor Selection: Use MS1 Precursor
- Use Isotope Pattern in Precursor Reevaluation: True
- Provide Profile Spectra: Automatic

2. Spectrum Properties Filter:
- Lower RT Limit: 5
- Upper RT Limit: 0
- First Scan: 0
- Last Scan: 0
- Lowest Charge State: 0
- Highest Charge State: 0
- Min. Precursor Mass: 700 Da
- Max. Precursor Mass: 7000 Da
- Total Intensity Threshold: 0
- Minimum Peak Count: 1

3. Scan Event Filters:
- Mass Analyzer: Is ITMS
- MS Order: Is Not MS1
- Min. Collision Energy: 0
- Max. Collision Energy: 1000
- Scan Type: Is Full

4. Peak Filters:
- S/N Threshold (FT-only): 1.5

5. Replacements for Unrecognized Properties:
- Unrecognized Charge Replacements: Automatic
- Unrecognized Mass Analyzer Replacements: ITMS
- Unrecognized MS Order Replacements: MS2
- Unrecognized Activation Type Replacements: CID
- Unrecognized Polarity Replacements: +
- Unrecognized MS Resolution@200 Replacements: 60000
- Unrecognized MSn Resolution@200 Replacements: 30000

6. Precursor Pattern Extraction:
- Precursor Clipping Range Before: 2.5 Da
- Precursor Clipping Range After: 5.5 Da

Processing node 5: PMI-Byonic
---
0. Help:
- Help: Click here (on ... button) for help

1. Protein Database:
- Protein database: acces fasta.fasta
- Add decoys: True
- Add common contaminants: False

2. Sample Digestion:
- Enzyme name: Trypsin (Full)
- Max missed cleavages: 2

3. Instrument Parameters:
- Precursor mass tolerance: 10 ppm
- Fragmentation type: HCD
- Fragment mass tolerance: 0.5 Da
- Fragment mass tolerance (2nd), if applicable: 0.5 Da
- Do NOT edit this parameter: No recalibration available
- Recalibration (from Preview): None

4. Static and Dynamic Modifications:
- Total common modification max: 4
- Total rare modification max: 0
- Modifications:
    Phospho / +79.966331 @ S, T, Y | common2
    Acetyl / +42.010565 @ K, R | common2
    Methyl / +14.01565 @ K, R | common2
    Carbamidomethyl / +57.021464 @ C | common1
    Oxidation / +15.994915 @ M | common1
    Met-loss+Acetyl / -89.02992 @ Protein NTerm M | common1
    Met-loss / -131.040485 @ Protein NTerm M | common1
    Nitrosyl / +28.990164 @ C, Y | common2

Deamidated / +0.984016 @ N | common2
GG / +114.042927 @ K | common2
HexNAc / +203.079373 @ N, S, T | common2
% Custom modification text below

5. Glycan Modifications:
- Show all N-glycopeptides: False

6. Wildcard Search:
- Enable wildcard search: False
- Wildcard minimum mass: -40
- Wildcard maximum mass: 100

7. Spectrum Input Options:
- Apply charges: Apply to unassigned spectra
- Skip bad spectra: True
- Precursor isotope off by x: Too high or low (narrow)
- Maximum precursor mass: 10000
- Precursor and charge assignments: Compute from MS1
- Maximum # of precursors per scan: 2

8. Protein Output Options:
- Protein FDR: 1% FDR (or 20 reverse counts)
- Create focused database: False
- Export mzIdentML: False

9. Peptide Output Options:
- Automatic score cut: True
- Manual score cut: 0

Processing node 4: Percolator
---
1. Target/Decoy Strategy:
- Target/Decoy Selection: Concatenated
- Validation based on: q-Value

2. Input Data:
- Maximum Delta Cn: 0.05
- Maximum Rank: 0

3. FDR Targets:
- Target FDR (Strict): 0.01
- Target FDR (Relaxed): 0.05

Processing node 6: IMP-ptmRS
---

1. Scoring:
- PhosphoRS Mode: False
- Report Only PTMs: True
- Use Diagnostic Ions: True
- Use Fragment Mass Tolerance of Search Node: True
- Fragment Mass Tolerance: 0.5 Da
- Consider neutral loss peaks for CID, HCD and EThcD: Automatic
- Maximum Peak Depth: 8
- Use a mass accuracy correction: False

2. Performance:
- Maximum Number of Position Isoforms: 500
- Maximum PTMs per peptide: 10

Consensus Workflow

The workflow tree:

|-(0) MSF Files
  |-(1) PSM Grouper
    |-(2) Peptide Validator
      |-(3) Peptide and Protein Filter
        |-(4) Protein Scorer
          |-(5) Protein Grouping
          |-(6) Protein FDR Validator

Post-processing nodes:

|-(7) Display Settings
|-(8) Result Statistics
|-(9) Data Distributions

Processing node 0: MSF Files
---
1. Storage Settings:
- Spectra to Store: Identified or Quantified
- Feature Traces to Store: All

2. Merging of Identified Peptide and Proteins:
- Merge Mode: Globally by Search Engine Type

3. FASTA Title Line Display:
- Reported FASTA Title Lines: Best match

- Title Line Rule: standard

4. PSM Filters:
- Maximum Delta Cn: 0.05
- Maximum Rank: 0
- Maximum Delta Mass: 0.5 Da

Hidden Parameters:
- MSF File(s) (Hidden):
    D:\Kish\PMi Byonic\Troub 2\msf2\NoFAIMS_HeLa-(2).msf
    D:\Kish\PMi Byonic\Troub 2\msf2\LC rep 2-(3).msf
    D:\Kish\PMi Byonic\Troub 2\msf2\HeLa_FAIMS_Sweep-(2).msf
    D:\Kish\PMi Byonic\Troub 2\msf2\HeLa_FAIMS_90_(2)-(2).msf
    D:\Kish\PMi Byonic\Troub 2\msf2\HeLa_FAIMS_90-(2).msf
    D:\Kish\PMi Byonic\Troub 2\msf2\HeLa_FAIMS_75-(2).msf
    D:\Kish\PMi Byonic\Troub 2\msf2\HeLa_FAIMS_67_(2)-(2).msf
    D:\Kish\PMi Byonic\Troub 2\msf2\HeLa_FAIMS_67_(1)-(2).msf
    D:\Kish\PMi Byonic\Troub 2\msf2\HeLa_FAIMS_52_(3)-(2).msf
    D:\Kish\PMi Byonic\Troub 2\msf2\HeLa_FAIMS_52_(2)-(2).msf
    D:\Kish\PMi Byonic\Troub 2\msf2\HeLa_FAIMS_52-(2).msf
    D:\Kish\PMi Byonic\Troub 2\msf2\HeLa_FAIMS_0_(2)-(2).msf
    D:\Kish\PMi Byonic\Troub 2\msf2\HeLa_FAIMS_0-(2).msf
    D:\Kish\PMi Byonic\Troub 2\msf2\2020-08-19_1ug_HeLa_noFAIMS_replicate1-(2).msf
    D:\Kish\PMi Byonic\Troub 2\msf2\2020-08-19_1ug_HeLa_FAIMSvoltoff_replicate2-(2).msf
    D:\Kish\PMi Byonic\Troub 2\msf2\2020-08-19_1ug_HeLa_FAIMSsweep_replicate2-(2).msf
    D:\Kish\PMi Byonic\Troub 2\msf2\2020-08-19_1ug_HeLa_FAIMSsweep_replicate1-(2).msf
    D:\Kish\PMi Byonic\Troub 2\msf2\2020-08-19_1ug_HeLa_FAIMS-90_replicate2-(2).msf
    D:\Kish\PMi Byonic\Troub 2\msf2\2020-08-19_1ug_HeLa_FAIMS-75_replicate2.msf
    D:\Kish\PMi Byonic\Troub 2\msf2\2020-08-19_1ug_HeLa_FAIMS-75_replicate2-(2).msf
    D:\Kish\PMi Byonic\Troub 2\msf2\2020-08-19_1ug_HeLa_FAIMS-75_replicate1-(2).msf
    D:\Kish\PMi Byonic\Troub 2\msf2\2020-08-19_1ug_HeLa_FAIMS-60_replicate2.msf
    D:\Kish\PMi Byonic\Troub 2\msf2\2020-08-19_1ug_HeLa_FAIMS-60_replicate2-(2).msf
    D:\Kish\PMi Byonic\Troub 2\msf2\2020-08-19_1ug_HeLa_FAIMS-60_replicate1.msf
    D:\Kish\PMi Byonic\Troub 2\msf2\2020-08-19_1ug_HeLa_FAIMS-60_replicate1-(2).msf
    D:\Kish\PMi Byonic\Troub 2\msf2\2020-08-19_1ug_HeLa_FAIMS-45_replicate2-(2).msf

D:\Kish\PMi Byonic\Troub 2\msf2\2020-08-19_1ug_HeLa_FAIMS-45_replicate1-(2).msf
D:\Kish\PMi Byonic\Troub 2\msf2\45 rep 2-(2).msf
D:\Kish\MSF files for byonic 140721\67V_replicate3.msf
D:\Kish\MSF files for byonic 140721\60V_replicate3.msf

Processing node 1: PSM Grouper
---
1. Peptide Group Modifications:
- Site Probability Threshold: 75

Processing node 2: Peptide Validator
---
1. General Validation Settings:
- Validation Mode: Automatic (Without control of peptide level error rate)
- Target FDR (Strict) for PSMs: 0.01
- Target FDR (Relaxed) for PSMs: 0.05
- Target FDR (Strict) for Peptides: 0.01
- Target FDR (Relaxed) for Peptides: 0.05

2. Specific Validation Settings:
- Validation Based on: q-Value
- Target/Decoy Selection for PSM Level FDR Calculation Based on Score: Automatic
- Reset Confidences for Nodes without Decoy Search (Fixed score thresholds): False

---
Processing node 3: Peptide and Protein Filter
---
1. Peptide Filters:
- Peptide Confidence At Least: Low
- Keep Lower Confident PSMs: True
- Minimum Peptide Length: 4
- Remove Peptides Without Protein Reference: False

2. Protein Filters:
- Minimum Number of Peptide Sequences: 1
- Count Only Rank 1 Peptides: False
- Count Peptides Only for Top Scored Protein: False

Processing node 4: Protein Scorer

No parameters

Processing node 5: Protein Grouping

---
1. Protein Grouping:
- Apply strict parsimony principle: True

Processing node 6: Protein FDR Validator
---
1. Confidence Thresholds:
- Target FDR (Strict): 0.01
- Target FDR (Relaxed): 0.05

Processing node 7: Display Settings
---
1. General:
- Filter Set:
    ### Filter Set MasterProteinFilter.filterset contains the following filters:
    ### Row Filter for TargetProtein:
    ### Master is equal to Master
    ###

    'magellan filter set' 1 'MasterProteinFilter.filterset' FiltersetProperties 1 'LastFileName' 'C:\Users\frank.berg\Desktop\MasterProteinFilter.filterset' Filter 'TargetProtein' 1 NARY_AND 1 = FilterConditionProperties 1 'NamedComparableFilterCondition/DisplayPropertyHint' 'Master' property 'Thermo.PD.EntityDataFramework.MasterProteinAssessment, Thermo.Magellan.EntityDataFramework' 'IsMasterProtein' constant 'Thermo.PD.EntityDataFramework.MasterProteinAssessment, Thermo.Magellan.EntityDataFramework' 'IsMasterProtein'

Processing node 8: Result Statistics

No parameters

Processing node 9: Data Distributions
---
1. ID Distributions (Bottom-up):
- Peptides to Use: All PSMs

8.3. Proteome Discoverer Search Parameters for SILAC LC-FAIMS-MS/MS (Chapter 4)

Processing Workflow

The workflow tree:

|-(7) Minora Feature Detector
|-(6) Spectrum Selector
|-(8) PMI-Byonic
|-(9) Percolator
|-(10) IMP-ptmRS

Processing node 7: Minora Feature Detector

---

1. Peak & Feature Detection:
- Min. Trace Length: 5
- S/N Threshold: 1
- Max. ΔRT of Isotope Pattern Multiplets [min]: 0.2

2. Feature to ID Linking:
- PSM Confidence At Least: High

Processing node 6: Spectrum Selector

---

1. General Settings:
- Precursor Selection: Use MS1 Precursor
- Use Isotope Pattern in Precursor Reevaluation: True
- Provide Profile Spectra: Automatic

2. Spectrum Properties Filter:
- Lower RT Limit: 0
- Upper RT Limit: 0
- First Scan: 0
- Last Scan: 0
- Lowest Charge State: 0
- Highest Charge State: 0
- Min. Precursor Mass: 350 Da
- Max. Precursor Mass: 5000 Da
- Total Intensity Threshold: 0
- Minimum Peak Count: 1

3. Scan Event Filters:
- MS Order: Is Not MS1

- Min. Collision Energy: 0
- Max. Collision Energy: 1000
- Scan Type: Is Full

4. Peak Filters:
- S/N Threshold (FT-only): 1.5

5. Replacements for Unrecognized Properties:
- Unrecognized Charge Replacements: Automatic
- Unrecognized Mass Analyzer Replacements: ITMS
- Unrecognized MS Order Replacements: MS2
- Unrecognized Activation Type Replacements: CID
- Unrecognized Polarity Replacements: +
- Unrecognized MS Resolution@200 Replacements: 60000
- Unrecognized MSn Resolution@200 Replacements: 30000

6. Precursor Pattern Extraction:
- Precursor Clipping Range Before: 2.5 Da
- Precursor Clipping Range After: 5.5 Da

Processing node 8: PMI-Byonic
---
0. Help:
- Help: Click here (on ... button) for help

1. Protein Database:
- Protein database: uniprot-Homo+sapiens+[9606]_review.fasta
- Add decoys: True
- Add common contaminants: False

2. Sample Digestion:
- Enzyme name: Trypsin (Full)
- Max missed cleavages: 4

3. Instrument Parameters:
- Precursor mass tolerance: 10 ppm
- Fragmentation type: HCD
- Fragment mass tolerance: 0.05 Da
- Fragment mass tolerance (2nd), if applicable: 0.05 Da
- Do NOT edit this parameter: No recalibration available
- Recalibration (from Preview): None

4. Static and Dynamic Modifications:
- Total common modification max: 4
- Total rare modification max: 1
- Modifications:
    Phospho / +79.966331 @ S, T, Y | common2

Acetyl / +42.010565 @ K, R | common2
Methyl / +14.01565 @ K, R | common2
Met-loss+Acetyl / -89.02992 @ Protein NTerm M | common1
Met-loss / -131.040485 @ Protein NTerm M | common1
HexNAc / +203.079373 @ N, S, T | common2
Nitrosyl / +28.990164 @ C, Y | common2
GG / +114.042927 @ K | common2
Carbamidomethyl / +57.021464 @ C | fixed
Oxidation / +15.994915 @ M | common2
% Custom modification text below

5. Glycan Modifications:
- Show all N-glycopeptides: False

6. Wildcard Search:
- Enable wildcard search: False
- Wildcard minimum mass: -40
- Wildcard maximum mass: 100

7. MS/MS Peak Filtering:
- Required number of MS/MS diagnostic peaks: 0
- M/z tolerance [Da]: 0.02

8. Spectrum Input Options:
- Apply charges: Apply to unassigned spectra
- Skip bad spectra: True
- Precursor isotope off by x: Too high or low (narrow)
- Maximum precursor mass: 10000
- Precursor and charge assignments: Compute from MS1
- Maximum # of precursors per scan: 2

9. Protein and Peptide Output Options:
- Protein FDR: 1% FDR (or 20 reverse counts)
- Create focused protein database: False
- Export mzIdentML: False
- Automatic peptide score cut: True
- Manual peptide score cut: 0

Processing node 9: Percolator
---
1. Target/Decoy Strategy:
- Target/Decoy Selection: Concatenated
- Validation based on: q-Value

2. Input Data:
- Maximum Delta Cn: 0.05
- Maximum Rank: 0

3. FDR Targets:
- Target FDR (Strict): 0.01
- Target FDR (Relaxed): 0.05

Processing node 10: IMP-ptmRS
---
1. Scoring:
- PhosphoRS Mode: False
- Report Only PTMs: True
- Use Diagnostic Ions: True
- Use Fragment Mass Tolerance of Search Node: True
- Fragment Mass Tolerance: 0.5 Da
- Consider neutral loss peaks for CID, HCD and EThcD: Automatic
- Maximum Peak Depth: 8
- Use a mass accuracy correction: False

2. Performance:
- Maximum Number of Position Isoforms: 500
- Maximum PTMs per peptide: 10

Consensus Workflow

The workflow tree:

```
 |-(0) MSF Files
  |-(10) Feature Mapper
   |-(11) Precursor Ions Quantifier
  |-(1) PSM Grouper
   |-(2) Peptide Validator
    |-(3) Peptide and Protein Filter
     |-(4) Protein Scorer
      |-(7) Protein FDR Validator
      |-(5) Protein Grouping
       |-(6) Peptide in Protein Annotation
       |-(15) Modification Sites
     |-(9) Protein Marker
     |-(8) Protein Annotation
```

Post-processing nodes:

```
 |-(12) Result Statistics
 |-(13) Display Settings
 |-(14) Data Distributions
```

Processing node 0: MSF Files
---
1. Storage Settings:
- Spectra to Store: Identified or Quantified
- Feature Traces to Store: All

2. Merging of Identified Peptide and Proteins:
- Merge Mode: Globally by Search Engine Type

3. FASTA Title Line Display:
- Reported FASTA Title Lines: Best match
- Title Line Rule: standard

4. PSM Filters:
- Maximum Delta Cn: 0.05
- Maximum Rank: 0
- Maximum Delta Mass: 0 ppm

Processing node 10: Feature Mapper
---
1. Chromatographic Alignment:
- Perform RT Alignment: True
- Maximum RT Shift [min]: 10
- Mass Tolerance: 10 ppm
- Parameter Tuning: Coarse

2. Feature Linking and Mapping:
- RT Tolerance [min]: 0
- Mass Tolerance: 0 ppm
- Min. S/N Threshold: 5

Processing node 11: Precursor Ions Quantifier
---
1. General Quantification Settings:
- Peptides to Use: Unique + Razor
- Consider Protein Groups for Peptide Uniqueness: True
- Use Shared Quan Results: True
- Reject Quan Results with Missing Channels: False

2. Precursor Quantification:
- Precursor Abundance Based On: Intensity
- Min. # Replicate Features [%]: 0

3. Normalization and Scaling:
- Normalization Mode: Total Peptide Amount
- Scaling Mode: On All Average

4. Exclude Peptides from Protein Quantification:
- For Normalization: Use All Peptides
- For Protein Roll-Up: Use All Peptides
- For Pairwise Ratios: Exclude Modified

5. Quan Rollup and Hypothesis Testing:
- Protein Abundance Calculation: Summed Abundances
- N for Top N: 3
- Protein Ratio Calculation: Pairwise Ratio Based
- Maximum Allowed Fold Change: 100
- Imputation Mode: None
- Hypothesis Test: t-test (Background Based)

6. Quan Ratio Distributions:
- 1st Fold Change Threshold: 2
- 2nd Fold Change Threshold: 4
- 3rd Fold Change Threshold: 6
- 4th Fold Change Threshold: 8
- 5th Fold Change Threshold: 10

Processing node 1: PSM Grouper
---
1. Peptide Group Modifications:
- Site Probability Threshold: 75

Processing node 2: Peptide Validator
---
1. General Validation Settings:
- Validation Mode: Automatic (Control peptide level error rate if possible)
- Target FDR (Strict) for PSMs: 0.01
- Target FDR (Relaxed) for PSMs: 0.05
- Target FDR (Strict) for Peptides: 0.01
- Target FDR (Relaxed) for Peptides: 0.05

2. Specific Validation Settings:
- Validation Based on: q-Value
- Target/Decoy Selection for PSM Level FDR Calculation Based on Score: Automatic
- Reset Confidences for Nodes without Decoy Search (Fixed score thresholds): False

---
Processing node 3: Peptide and Protein Filter
---

1. Peptide Filters:
- Peptide Confidence At Least: High
- Keep Lower Confident PSMs: False
- Minimum Peptide Length: 6
- Remove Peptides Without Protein Reference: False

2. Protein Filters:
- Minimum Number of Peptide Sequences: 1
- Count Only Rank 1 Peptides: False
- Count Peptides Only for Top Scored Protein: False

Processing node 4: Protein Scorer

No parameters

Processing node 7: Protein FDR Validator
---
1. Confidence Thresholds:
- Target FDR (Strict): 0.01
- Target FDR (Relaxed): 0.05

Processing node 5: Protein Grouping
---
1. Protein Grouping:
- Apply strict parsimony principle: True

Processing node 6: Peptide in Protein Annotation
---
1. Flanking Residues:
- Annotate Flanking Residues of the Peptide: True
- Number Flanking Residues in Connection Tables: 1

2. Modifications in Peptide:
- Protein Modifications Reported: Only for Master Proteins

3. Modifications in Protein:
- Modification Sites Reported: All And Specific
- Minimum PSM Confidence: High
- Report Only PTMs: True

4. Positions in Protein:
- Protein Positions for Peptides: Only for Master Proteins

Processing node 15: Modification Sites
---
1. General:
- Report Only PTMs: True
- Only Master Proteins: True
- Motif Radius: 6

Processing node 9: Protein Marker
---
5. Annotate Species:
- As Species Map: False
- As Species Names: False

6. Mark Additional Entities:
- Annotation Groups: False
- Pathway Groups: False
- Modification Sites: True
- Peptide Isoform Groups: True

Processing node 8: Protein Annotation
---
1. Annotation Aspects:
- 1. Aspect: Biological Process
- 2. Aspect: Cellular Component
- 3. Aspect: Molecular Function
- 4. Aspect: None
- 5. Aspect: None
- 6. Aspect: None

2. Annotation/Pathway Groups:
- Protein Database: uniprot-Homo+sapiens+[9606]_review.fasta

8.4. Western blotting (Chapter 4)

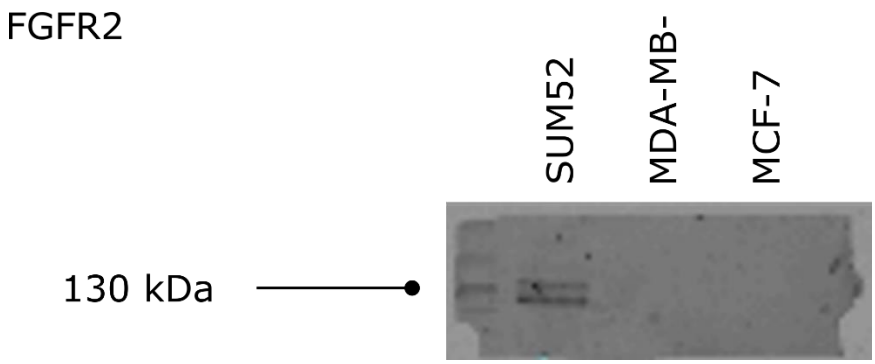


Figure 8.1 | Raw Western blot for FGFR2.

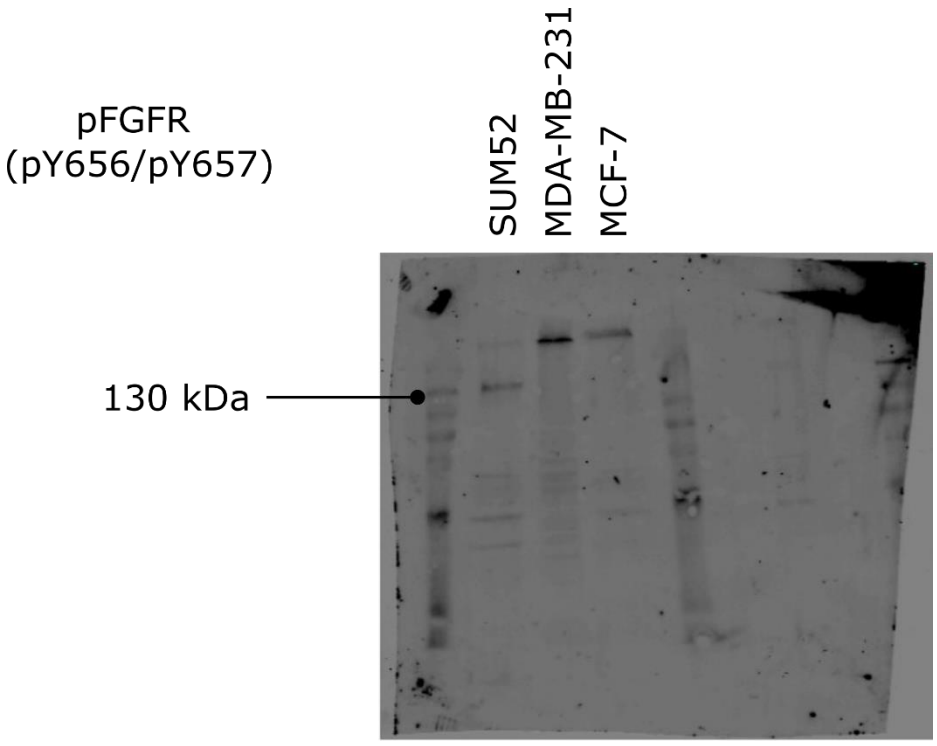


Figure 8.2 | Raw Western blot for phospho-FGFR2 (tyrosine-656/tyrosine-657).

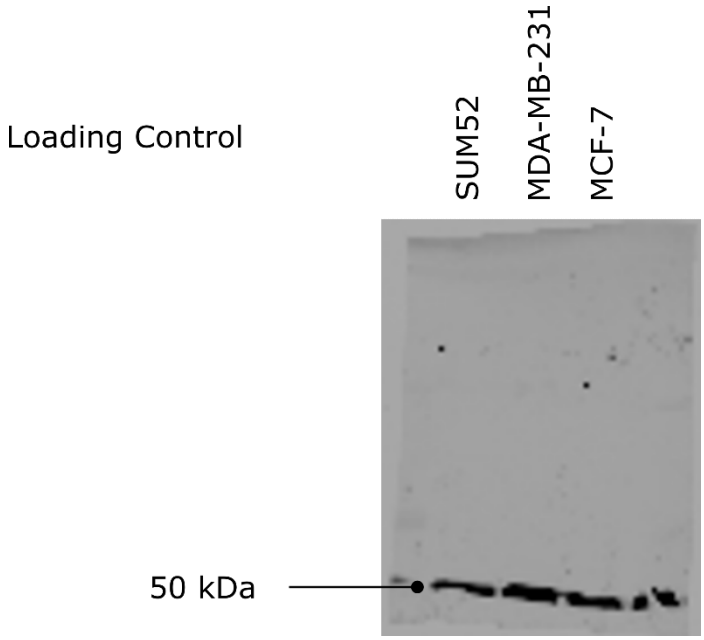


Figure 8.3 | RAW Western blot for α-tubulin loading control.

8.5. Gene Ontology Analysis of Downregulated Proteins with Both AZD4547 and SU5402

Table 8.1 | Proteins corresponding to enrichment for Molecular Function: Cytoskeletal Protein Binding

Accession	Gene Name
Q8WZ42	dynamin 1 like
P29966	gap junction protein alpha 1
P06396	gelsolin
P06703	kinesin family member 5B
P05109	myristoylated alanine rich protein kinase C substrate
Q9Y2K3	myosin heavy chain 13
Q9UKX2	myosin heavy chain 15
Q6ZT98	myosin heavy chain 2
P26447	S100 calcium binding protein A4
Q9UKX3	S100 calcium binding protein A6
Q8TF72	S100 calcium binding protein A8
Q5T5C0	shroom family member 3
P33176	syntaxin binding protein 5
P17302	tubulin tyrosine ligase like 7
O00429	titin

Table 8.2 | Proteins corresponding to enrichment for Molecular Function: Actin Filament Binding

Accession	Gene Name
Q8WZ42	gelsolin
P29966	myristoylated alanine rich protein kinase C substrate
P06396	myosin heavy chain 13
Q9Y2K3	myosin heavy chain 15
Q9UKX2	myosin heavy chain 2
Q9UKX3	shroom family member 3
Q8TF72	titin

8.6. Native Mass Spectra Peak Annotation (Chapter 5)

Table 8.3 | Protein species corresponding to native mass spectrum of *Arthrospira maxima* from sample culturing to MS analysis (Figure 5.3). PCB = phycocyanobilin chromophore, N4-methylAsp = methylation PTM at aspartic acid, experimental molecular weight reported with standard deviation.

Proteoform	**Theoretical Molecular Weight (Da)**	**PCB**	**N4-methylAsp**	**N-terminal methionine loss**	**Experimental Molecular Weight (Da)**	**Mass Deviation (%)**
C-Phycocyanin Dimer (αβ)	37,465	3	1	0	37,467 ± 0.5	0.007
Allophycocyanin Hexamer (α3β3)	107,327	6	3	3	107,359 ± 0.5	0.030

Table 8.4 | Protein species corresponding to native mass spectra of *Arthrospira maxima* (zoomed in 2800 – 3700 *m/z*) (Figure 5.4). GS = Glutathione, PCB = phycocyanobilin chromophore, N4-methylAsp = methylation PTM at aspartic acid, experimental molecular weight reported with standard deviation.

Proteoform	**Theoretical Molecular Weight (Da)**	**PCB**	**N4-methylAsp**	**N-terminal methionine loss**	**Experimental Molecular Weight (Da)**	**Mass Deviation (%)**
C-Phycocyanin Dimer (αβ)	37,465	3	1	0	37,467 ± 0.5	0.007
C-Phycocyanin Dimer (αβ) + GS	37,770	3	1	0	37,770 ± 0.04	0.001

Table 8.5 | | Protein species corresponding to native mass spectra of *Arthrospira platensis* (Figure 5.5). GS = Glutathione, PCB = phycocyanobilin chromophore, N4-methylAsp = methylation PTM at aspartic acid, experimental molecular weight reported with standard deviation.

Proteoform	**Theoretical Molecular Weight (Da)**	**PCB**	**N4-methylAsp**	**N-terminal methionine loss**	**Experimental Molecular Weight (Da)**	**Mass Deviation**
C-Phycocyanin Dimer (αβ)	37,465	3	1	0	37,467 ± 0.5	0.007
C-Phycocyanin Dimer (αβ) + GS	37,770	3	1	0	37,770 ± 0.04	0.001
C-Phycocyanin Hexamer (α3β3)	112,394	9	3	0	112,400 ± 1.3	0.005

C-Phycocyanin Hexamer (αβ) + GS	112,699	9	3	0	112,706 ± 1.0	0.006
C-Phycocyanin Hexamer (αβ) + 2xGS	113,004	9	3	0	113,014 ±1.0	0.009
allophycocyanin Hexamer (α3β3)	107,327	6	3	3	107,359 ± 0.5	0.030

Table 8.6 | Protein species corresponding to native and native top-down mass spectra of precursor ion peak 3434 *m/z* and HCD dissociation into MS^2 peaks (Figure 5.6). GS = Glutathione, PCB = phycocyanobilin chromophore, N4-methylAsp = methylation PTM at aspartic acid, experimental molecular weight reported with standard deviation.

Proteoform	**Theoretical Molecular Weight (Da)**	**PCB**	**N4-methylAsp**	**N-terminal methionine loss**	**Experimental Molecular Weight (Da)**	**Mass Deviation**
C-Phycocyanin Dimer (αβ) + GS	37,770	3	1	0	37,770 ± 1.5	0.001
C-Phycocyanin α	18,187	1	0	0	18,184 ± 1.1	0.016
C-Phycocyanin β + GS	19,586	2	1	0	19,585 ± 1.4	0.005

Table 8.7 | Protein species corresponding to native mass spectra of *Arthrospira platensis* for high light exposure experiment (Figure 5.9). GS = Glutathione, PCB = phycocyanobilin chromophore, N4-methylAsp = methylation PTM at aspartic acid, experimental molecular weight reported with standard deviation.

Proteoform	**Theoretical Molecular Weight (Da)**	**PCB**	**N4-methylAsp**	**N-terminal methionine loss**	**Experimental Molecular Weight (Da)**	**Mass Deviation**
C-Phycocyanin Dimer (αβ)	37,465	3	1	0	37,467 ± 0.5	0.007
C-Phycocyanin Dimer (αβ) + GS	37,770	3	1	0	37770 ± 1.2	0.001

Table 8.8 | Protein species corresponding to native mass spectra of *Arthrospira platensis* for hydrogen peroxide exposure experiment (Figure 5.10). GS = Glutathione, PCB = phycocyanobilin chromophore, N4-methylAsp = methylation PTM at aspartic acid, experimental molecular weight reported with standard deviation.

Proteoform	**Theoretical Molecular Weight (Da)**	**PCB**	**N4-methylAsp**	**N-terminal methionine loss**	**Experimental Molecular Weight (Da)**	**Mass Deviation**
C-Phycocyanin Dimer (αβ)	37,465	3	1	0	37,465 ±1.0	0.00
C-Phycocyanin Dimer (αβ) + GS	37,770	3	1	0	37770 ± 0.6	0.00